

1

Learning to Identify Extragalactic 2 Radio Sources

3

4

Matthew J. Alger

5 A thesis submitted for the degree of
6 Doctor of Philosophy
7 of The Australian National University



Australian
National
University

September 2021

¹⁰
¹¹
¹²REVISION DIFF

© Matthew J. Alger 2021
All rights reserved

13 I declare that the research presented in this thesis represents original work that I
14 carried out during my candidature at the Australian National University, except for
15 contributions to multi-author papers incorporated in the thesis where my contribu-
16 tions are specified in this Statement of Contribution.

- 17 • *Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification*,
18 by **M. J. Alger**, J. K. Banfield, C. S. Ong, L. Rudnick, O. I. Wong, C. Wolf, H.
19 Andernach, R. P. Norris, and S. S. Shabala. Published in 2018 in the *Monthly*
20 *Notices of the Royal Astronomical Society*. Chapter 4 in this thesis.
- 21 • *Radio Galaxy Zoo: Radio luminosity functions of extended sources*, by **M. J. Alger**,
22 O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. Andernach, L. Rudnick, S. S.
23 Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, R. P. Norris, and A. J. M.
24 Thomson. Not yet submitted. Chapter 5 in this thesis.
- 25 • *Interpretable Faraday Complexity Classification*, by **M. J. Alger**, J. D. Livingston, N.
26 M. McClure-Griffiths, J. L. Nabaglo., O. I. Wong, and C. S. Ong. Accepted; to be
27 published in *Publications of the Astronomical Society of Australia*. Chapter 6 in this
28 thesis.

29 For all three papers I wrote the entirety of the content, produced all figures ex-
30 cept where noted, and conducted all experiments. Other authors reviewed the papers,
31 made suggestions, discussed ideas, and vitally contributed to the Radio Galaxy Zoo
32 project.

33 Matthew J. Alger
34 September 8, 2021

REVISION DIFF

REVISION AND DIFF

To Shirley and Bob.

REVISION DIFF

³⁶ Acknowledgments

³⁷ I could not have completed this thesis without my incredible supervisors: Naomi
³⁸ McClure-Griffiths, who stepped into my panel when circumstances required it and
³⁹ made me feel just as welcome in her group as any student who had been there from
⁴⁰ the beginning; Ivy Wong, who was always available to help answer my many inane
⁴¹ questions on the depths of radio astronomy; Julie Banfield, who helped me learn the
⁴² ropes of astronomy and pointed out all of the important websites to me; and Cheng
⁴³ Soon Ong, who advised when I needed advising, pushed when I needed pushing, and
⁴⁴ was always there when I needed support.

⁴⁵ Much of this work was done in my office at Mount Stromlo while I bothered my
⁴⁶ office-mates: Eloise Birchall, Adam Rains, Lara Cullinane, and for a time Rajika Ku-
⁴⁷ ruwita and Alex Wallace. Without their welcome distractions I would not have been
⁴⁸ able to concentrate. Of course, Mount Stromlo was more than just my office, and I
⁴⁹ also must thank Henry Zovaro, Phil Taylor, Karlie Noon, Adam Thomas, Jack Liv-
⁵⁰ ington, Fiona Panther, and Alec Thomson for their contributions to the atmosphere
⁵¹ of the observatory while I was there, as well as Howard Coyle and Michelle Cicolini for
⁵² keeping the observatory running smoothly. I would also like to thank my colleagues
⁵³ at Data61: Cody Christopher, Steve Edwards, and Angus Gruen, who provided an
⁵⁴ excellent sounding board to my many astronomical problems.

⁵⁵ Chapters 4 and 5 rely heavily on results from the Radio Galaxy Zoo citizen science
⁵⁶ project. I would like to thank all 10 000+ volunteers who contributed to the project,
⁵⁷ as well as the members of the Radio Galaxy Zoo collaboration, notably Stas Shabala,
⁵⁸ Larry Rudnick, Heinz Andernach, Avery Garon, Anna Kapinska, and Ray Norris, all of
⁵⁹ whom contributed to discussion of my work for these chapters. I must also thank those
⁶⁰ who provided me funding over this PhD: the Australian Government, the Research
⁶¹ School of Astronomy and Astrophysics, Data61/CSIRO, and the Astronomical Society
⁶² of Australia.

⁶³ A special thank you to those who proof-read this thesis: Rachel Alger, Jakub Nabaglo,
⁶⁴ Henry Zovaro, Jack Livingston, Sarah Lawson, Eloise Birchall, and Adam Rains.

⁶⁵ Finally, I would like to thank my family, Rachel, Geoff, and Kelly Alger; as well as
⁶⁶ my partner Sarah Lawson; and my friends Nicholas Hogan, Mitchell Busby, Bronnagh
⁶⁷ Norris, and Jakub Nabaglo; for their emotional support through my PhD.

68 The work and writing behind this thesis was completed in Australia on the land
69 of the Ngunnawal and Ngambri people. I would like to acknowledge and celebrate
70 the traditional custodians of this land, and pay my respects to their Elders past and
71 present.

REVISION DIFF

72 Abstract

73 Radio observations of actively accreting supermassive black holes outside of the galaxy
74 can provide insight into the history of galaxies and their evolution. With the construc-
75 tion of fast new radio telescopes and the undertaking of large new radio surveys in
76 the lead-up to the Square Kilometre Array (SKA), radio astronomy faces a ‘data del-
77 uge’ where traditional methods of data analysis cannot keep up with the scale of the
78 data. Astronomers are increasingly looking to machine learning to provide ways of
79 handling large-scale data like these. This thesis introduces machine learning methods
80 for use in wide-area radio surveys and demonstrate their application to radio astron-
81 omy data. To help understand the issues facing large-scale wide-area radio surveys,
82 and contribute toward their solutions, we consider the problems of automated radio-
83 infrared cross-identification and Faraday complexity classification.

84 We developed an automated machine learning method for cross-identifying radio
85 objects with their infrared counterparts, training the algorithm with data from the cit-
86 izen science project Radio Galaxy Zoo. The trained result performed comparably to
87 an algorithm trained on expert cross-identifications, demonstrating the benefit of non-
88 expert labelling in radio astronomy. By examining the theoretical maximum accuracy
89 of this algorithm we showed that existing pilot studies for future surveys were not
90 sufficiently large enough to train machine learning methods. We showed the utility of
91 our cross-identification algorithm by applying it instead to a large survey, Faint Images
92 of the Radio Sky at Twenty Centimeters (FIRST), producing the largest catalogue of
93 cross-identified extended sources available at the time of writing. From this catalogue,
94 we calculated a mid-infrared-divided fractional radio luminosity function as well as
95 an estimate of energy injected into the intergalactic medium by active galactic nuclei
96 jets—one of the first applications of machine learning to radio astronomy to obtain a
97 physics result. A key result from this work was that the limitation in our sample size
98 was not due to the number of radio objects cross-identified but rather by the number
99 of available redshift measurements. Finally, we developed interpretable features for
100 spectropolarimetric measurements of radio sources and used these features to design
101 a machine learning algorithm that can identify Faraday complexity, while the features
102 themselves may be used for other tasks. The methods in this thesis will be applicable
103 to future radio surveys such as the Evolutionary Map of the Universe (EMU) contin-
104 uum survey and the Polarised Sky Survey of the Universe’s Magnetism (POSSUM),
105 as well as surveys produced with the SKA, allowing the development of higher reso-
106 lution radio luminosity functions, better estimates of the impact of radio galaxies on
107 their environments, faster analysis of polarised surveys, and better quality rotation
108 measure grids.

REVISION DIFF

Contents

109	Acknowledgments	vii
111	Abstract	ix
112	List of Figures	xv
113	List of Tables	xvii
114	List of Constants	xix
115	List of Abbreviations	xxi
116	1 Introduction	1
117	1.1 Problems in extragalactic radio astronomy	2
118	1.2 Big data in astronomy	3
119	1.3 Machine learning in astronomy	4
120	1.4 How this thesis fits in	5
121	1.5 Thesis outline	6
122	1.6 Contributions	6
123	1.7 Works produced during this PhD	8
124	2 Radio Sources	11
125	2.1 The Extragalactic Radio Sky	11
126	2.2 Radio emission	13
127	2.2.1 Synchrotron radiation	14
128	2.2.2 Polarisation	14
129	2.3 Radio galaxies and active galactic nuclei	18
130	2.3.1 What we see when we look at AGN	18
131	2.3.2 Extended structure	20
132	2.3.3 The unified model	21
133	2.3.4 Polarised structure	22
134	2.3.5 AGN luminosity	23
135	2.3.6 The role of AGN	23
136	2.4 Classifying AGN	24
137	2.4.1 Statistical and manual classification of AGN	24
138	2.4.2 Machine learning classification of AGN	25
139	2.5 Cross-identification	27
140	2.5.1 Why do we need to cross-identify?	27

141	2.5.2	Methods for cross-identification	27
142	2.5.2.1	Positional matching	27
143	2.5.2.2	Other automated methods	28
144	2.6	Aggregating Radio Components	28
145	2.6.1	Missing emission in radio observations	29
146	2.6.2	Methods of aggregation	31
147	2.7	Summary: radio sources	31
148	3	Machine Learning for Astroinformatics	33
149	3.1	Prediction	33
150	3.1.1	Predictors	33
151	3.1.2	Classification	34
152	3.1.3	Regression	35
153	3.2	Data and representation	35
154	3.3	Loss functions	36
155	3.3.1	Loss function for regression	37
156	3.3.2	Loss function for binary classification	38
157	3.3.3	Gradient descent	39
158	3.4	Models	39
159	3.4.1	Logistic regression	40
160	3.4.2	Decision tree ensembles	40
161	3.4.3	Convolutional neural networks	41
162	3.5	Labels	42
163	3.5.1	Where do labels come from?	42
164	3.5.2	Label noise	43
165	3.6	Summary: machine learning	43
166	4	Radio Cross-identification	45
167	4.1	Introduction to cross-identification	45
168	4.2	Data	48
169	4.2.1	ATLAS	48
170	4.2.2	SWIRE	49
171	4.2.3	Radio Galaxy Zoo	49
172	4.3	Method	49
173	4.3.1	Cross-identification as binary classification	49
174	4.3.2	Limitations of our approach	53
175	4.3.3	Feature vector representation of infrared sources	56
176	4.3.4	Binary Classifiers	56
177	4.3.5	Labels	57
178	4.3.6	Experimental Setup	59
179	4.4	Results	60
180	4.4.1	Application to ATLAS-CDFS	60
181	4.4.2	Application to ATLAS-ELAIS-S1	65
182	4.5	Discussion	68

183	4.6	Summary	70
184	4.7	Acknowledgements	71
185	A	<u>Classification models</u>	71
186		A.1 <u>Logistic Regression</u>	72
187		A.2 <u>Convolutional neural networks</u>	72
188		A.3 <u>Random Forests</u>	72
189	B	<u>Accuracy tables</u>	74
190	C	<u>SWIRE object scores</u>	74
191	D	<u>ATLAS component cross-identifications</u>	76
192	E	<u>Cross-identification figures</u>	80
193	5	Radio Luminosity Functions	85
194	1	Introduction	86
195	2	Data	87
196	2.1	RGZ	88
197	2.2	FIRST	88
198	2.3	AllWISE	88
199	2.4	SDSS	89
200	3	Method	89
201	3.1	Visual verification	90
202	4	Radio luminosity functions	91
203	5	Discussion	96
204	5.1	Biases and uncertainties	96
205	5.2	Extended radio galaxies in the low- z Universe	97
206	5.3	Future work	99
207	6	Summary	99
208	7	Acknowledgements	100
209	F	<u>Sankey diagrams</u>	100
210	G	<u>Visual verification results</u>	100
211	H	<u>Radio luminosity function</u>	104
212	I	<u>Redshift completeness estimate</u>	104
213	J	<u>Giant radio galaxies</u>	105
214	6	Faraday Complexity	107
215	1	Introduction	108
216	2	Faraday Complexity	109
217	2.1	Prior work	109
218	2.2	Assumptions on Faraday dispersion functions	111
219	3	Classification approach	112
220	3.1	Features	112
221	3.2	Interpreting distances	114
222	3.3	Classifiers	115
223	4	Experimental method and results	115
224	4.1	Data	115

225	4.1.1	Simulated training and validation data	115
226	4.1.2	Observational data	116
227	4.2	Results on ‘ASKAP’ dataset	117
228	4.3	Results on ‘ATCA’ dataset	117
229	4.4	Results on observed FDFs	119
230	5	Discussion	121
231	5.1	Complexity and seeming ‘not simple’	121
232	5.2	Limitations	122
233	6	Conclusion	122
234	7	Acknowledgements	123
235	K	<u>Simulating observed FDFs</u>	123
236	L	<u>2-Wasserstein begets Faraday moments</u>	124
237	M	<u>Euclidean distance in the no-RMSF case</u>	125
238	N	<u>Hyperparameters for LR and XGB</u>	126
239	O	<u>Predictions on real data</u>	128
240	7	Conclusion	131
241	1	Future Work	133
242	2	Implications for radio citizen science	135
243	3	Implications for wide-area radio surveys	135
244	4	Final remarks	136
245	Appendix		136
246	5	Classification models	136
247	4.1	Logistic Regression	136
248	4.1	Convolutional neural networks	136
249	4.1	Random Forests	137
250	5	Accuracy tables	137
251	5	SWIRE object scores	138
252	5	ATLAS component cross-identifications	140
253	5	Cross-identification figures	145
254	5	Sankey diagrams	145
255	5	Radio luminosity function	145
256	5	Redshift completeness estimate	146
257	5	Giant radio galaxies	146
258	5	Visual verification results	148
259	5	2-Wasserstein begets Faraday moments	150
260	5	Euclidean distance in the no-RMSF case	151
261	5	Hyperparameters for LR and XGB	152
262	5	Predictions on real data	153
263	5	Simulating observed FDFs	153
264	Bibliography		155

List of Figures

266	2.1	False-colour image of the radio sky from the GLEAM survey.	12
267	2.2	A Faraday screen.	16
268	2.3	Examples of a FRI and a FRII radio galaxy.	19
269	2.4	Three radio galaxies with interesting structure.	20
270	2.5	The unified model of AGN.	21
271	2.6	RLF for star-forming galaxies and AGN, from Mauch and Sadler (2007).	23
272	2.7	A single dish telescope and a radio array.	29
273	2.8	An example of a ‘resolved out’ radio galaxy.	30
274	4.1	Examples showing key definitions of radio emission regions used throughout this chapter.	46
275	4.2	An example of finding the host galaxy of a radio source using our sliding-window method.	50
276	4.3	Our cross-identification method once a binary classifier has been trained.	50
277	4.4	A radio source breaking our assumption that there are no other radio sources with 1 arcmin of the source.	53
278	4.5	A radio source where the window centred on the host galaxy does not contain enough radio information to correctly identify the galaxy as a host.	54
279	4.6	A radio source breaking our assumption that the whole radio source is visible in the chosen radius.	54
280	4.7	Cumulative number of radio components in the expert and Radio Galaxy Zoo training sets with different signal-to-noise ratios.	57
281	4.8	CDFS field training and testing quadrants.	58
282	4.9	Balanced accuracy on the candidate classification task plotted against accuracy on the cross-identification task.	61
283	4.10	An example of predicted host galaxies in the candidate classification task.	62
284	4.11	Performance of our method with different binary classifiers on the binary classification task.	63
285	4.12	Performance of our approach using different binary classifiers on the cross-identification task.	64
286	4.13	Performance of different classification models on the binary classification task, tested on ELAIS-S1.	65
287	4.14	Performance of different classification models on the cross-identification task, tested on ELAIS-S1.	66
288	4.15	Balanced accuracies of classifiers with different SNR cutoffs.	67

301	4.16	Architecture of our CNN.	73
302	4.17	Examples of resolved sources with high disagreement between cross- identifiers.	83
304	5.1	RGZ-Ex radio luminosity function compared with the RLFs of Mauch and Sadler (2007).	91
305	5.2	<i>WISE</i> colour-colour distributions.	93
306	5.3	RLFs split by <i>WISE</i> colour regions.	94
308	5.4	Bivariate radio luminosity function showing radio luminosity against projected physical extent.	95
310	5.5	<u>Number of components removed from FIRST by each filter.</u>	101
311	5.6	<u>Number of sources removed by each filter.</u>	101
312	5.7	<u>Estimated completeness as a function of mid-infrared colour and magnitude.</u>	105
313	6.1	A simple FDF and its corresponding polarised spectra.	110
314	6.2	A complex FDF and its corresponding polarised spectra.	110
315	6.3	An illustration of our FDF features.	114
316	6.4	Mean prediction as a function of component depth separation and min- imum component amplitude for XGB and LR.	118
317	6.5	Principal component analysis for simulated data with observations over- laid.	119
319	6.6	Estimated rates of Faraday complexity for the Livingston and O'Sullivan datasets as functions of threshold.	120
320	6.7	The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex.	128
322	6.8	The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex.	129
324			
325			

List of Tables

327 4.1	Catalogues of ATLAS/SWIRE cross-identifications for the CDFS and 328 ELAIS-S1 fields.	48
329 4.2	Number of compact and resolved radio objects in each CDFS quadrant. .	57
330 4.3	Balanced accuracies for different binary classification models on CDFS. .	74
331 4.4	Balanced accuracies for different binary classification models on ELAIS- 332 S1.	75
333 4.5	Cross-identification accuracies for different classification models on CDFS. .	75
334 4.6	Cross-identification accuracies for different classification models on ELAIS- 335 S1.	76
336 4.7	Scores output by our trained classifiers for SWIRE CDFS candidate host 337 galaxies.	77
338 4.8	Scores output by our trained classifiers for SWIRE ELAIS-S1 candidate 339 host galaxies.	78
340 4.9	Cross-identifications for ATLAS CDFS components.	81
341 4.10	<u>Cross-identifications for ATLAS ELAIS-S1 components. Columns are 342 defined in Appendix D. Full table electronic.</u>	82
343 5.1	Medians and standard deviations used to normalise input features <u>for 344 our classifiers.</u>	89
345 5.2	Validation objects.	102
346 5.3	Giant radio galaxies found in RGZ-Ex.	106
347 6.1	Confusion matrix entries for LR and XGB on 'ASKAP' and 'ATCA' sim- 348 ulated datasets, and the CNN confusion matrix entries adapted from 349 Brown et al. (2018).	117
350 6.2	<u>XGB hyperparameters for the 'ATCA' dataset.</u>	127
351 6.3	<u>LR hyperparameters for the 'ATCA' dataset.</u>	127
352 6.4	<u>XGB hyperparameters for the 'ASKAP' dataset.</u>	127
353 6.5	<u>LR hyperparameters for the 'ASKAP' dataset.</u>	127

REVISION DIFF

³⁵⁴

List of Constants

³⁵⁵ The values of the following constants, except where otherwise noted, are drawn from
³⁵⁶ the NIST Reference on Constants, Units, and Uncertainty (Mohr et al., 2019) which
³⁵⁷ itself draws from the 2018 CODATA recommended values.

Symbol	Unit	Name	Value
ϵ_0	F m^{-1}	Vacuum permittivity	$8.8541878128(13) \times 10^{-12}$
G	$\text{m}^3 \text{ kg}^{-1} \text{ s}^{-2}$	Gravitational constant	$6.67430(15) \times 10^{-11}$
m_p	kg	Proton mass	$1.67262192369(51) \times 10^{-27}$
m_e	kg	Electron mass	$9.1093837015(28) \times 10^{-31}$
c	m s^{-1}	Speed of light	2.99792458×10^8
σ_T	m^2	Thomson cross section	$6.6524587321(60) \times 10^{-29}$

REVISION DIFF

List of Abbreviations

- ³⁵⁸ The following list summarises abbreviations that are commonly used in this thesis.
- ³⁵⁹ • AGN: active galactic nuclei, energetic objects at the centre of galaxies
 - ³⁶⁰ • ASKAP: Australian Square Kilometre Array Pathfinder, a next-generation radio telescope in Murchison
 - ³⁶¹ • ATCA: Australia Telescope Compact Array, a radio telescope in Narrabri
 - ³⁶² • CNN: convolutional neural network, a classifier which works on images and spectra
 - ³⁶³ • EMU: Evolutionary Map of the Universe, an upcoming large radio survey
 - ³⁶⁴ • FDF: Faraday dispersion function, a representation of a polarisation spectrum
 - ³⁶⁵ • FIRST: Faint Images of the Radio Sky at Twenty Centimeters, a large radio survey
 - ³⁶⁶ • FRI: Fanaroff-Riley type I, an edge-darkened radio galaxy
 - ³⁶⁷ • FRII: Fanaroff-Riley type II, an edge-brightned radio galaxy
 - ³⁶⁸ • ISM: interstellar medium, the stuff between stars
 - ³⁶⁹ • LR: logistic regression, a classification model
 - ³⁷⁰ • MWA: Murchison Widefield Array, a next-generation radio telescope in Murchison
 - ³⁷¹ • NVSS: NRAO VLA Sky Survey, a large radio survey
 - ³⁷² • POSSUM: Polarisation Sky Survey of the Universe's Magnetism, an upcoming large radio polarisation survey
 - ³⁷³ • RACS: Rapid ASKAP Continuum Survey, a new large radio survey
 - ³⁷⁴ • RF: random forests, a classification model
 - ³⁷⁵ • RGZ: Radio Galaxy Zoo, a citizen science project to cross-identify and aggregate radio sources
 - ³⁷⁶ • RLF: radio luminosity function, a description of how common radio galaxies of different energies are

- 384 • RM: rotation measure, the amount of Faraday rotation between a polarised source
385 and an observer
- 386 • RMSF: rotation measure spread function, the kernel convolving a FDF
- 387 • SDSS: Sloan Digital Sky Survey, a large optical spectroscopic and photometric
388 survey
- 389 • SFR: star formation rate, the recent rate of star formation
- 390 • SKA: Square Kilometre Array, a next-generation radio telescope yet to be built
- 391 • SNR: signal-to-noise ratio, the ratio of total intensity to noise level
- 392 • SWIRE: *Spitzer* Wide-area Infrared Extragalactic Survey, a deep infrared survey
- 393 • VLA: Very Large Array, a radio telescope in New Mexico
- 394 • WISE: *Wide-field Infrared Survey Explorer*, a space-based infrared telescope
- 395 • XGB: extreme gradient boosted trees, a classification model

397 Introduction

398 Many great results come from study at the intersection of two fields, and the combina-
399 tion of astronomy and informatics is no exception. The resulting interdisciplinary field
400 is called *astroinformatics*, and concerns the application of statistical and machine learn-
401 ing techniques to problems in astronomy and astrophysics. Machine learning, a col-
402 lection of methods for formalising and solving data-driven problems at scale, is a nat-
403 ural fit for radio astronomy: Radio astronomy faces a ‘data deluge’ thanks to new and
404 upgraded telescopes and associated wide-area surveys to be undertaken with them.
405 The goal in the near future is to be able to process data—and conduct science on that
406 data—on the fly as the volume of data grows too large to store. This is a requirement
407 for the Square Kilometre Array (SKA), a grand international undertaking to build a
408 distributed radio array across Western Australia and South Africa with an intended
409 collecting area of one square kilometre. There exists no comparable telescope today.
410 The SKA will be able to perform new tests of general relativity, help pin down the large-
411 scale structure of the Universe, investigate the mysteries of the epoch of reionisation,
412 probe the history of galaxy evolution to new extents, and perhaps make unexpected
413 new discoveries (Diamond, 2017).

414 The SKA will have technical challenges unlike any so far in radio astronomy. Raw
415 data will stream from the telescope antennae at 2 petabytes per second, and up to 300
416 petabytes per year of science data is expected to be generated (Diamond, 2017). This
417 is a phenomenal amount of data, much of which won’t be stored, and the community
418 expectation is that machine learning will provide avenues to conduct science with such
419 a large dataset. Precursor projects to the SKA have begun to investigate these pathways
420 (e.g. Bonaldi et al., 2020; Kapinska, 2020; Mostert et al., 2021).

421 Even without the SKA, the data deluge has already begun. Three precursor tele-
422 scopes have been constructed: the Australian Square Kilometre Array Pathfinder
423 (ASKAP) and Murchison Widefield Array (MWA) in Western Australia, and MeerKAT
424 in South Africa. While MeerKAT will eventually form part of the SKA itself, all three
425 are already online and generating science data at astonishing rates. ASKAP will soon
426 conduct six surveys of the entire southern radio sky: the Evolutionary Map of the Uni-
427 verse (EMU; Kapinska, 2020; Norris et al., 2011), the Widefield ASKAP L-Band Legacy
428 All-Sky Blind Survey (Koribalski et al., 2020, WALLABY), the First Large Absorption
429 Survey in HI (FLASH), an ASKAP Survey for Variables and Slow Transients (VAST;
430 Murphy et al., 2013), the Galactic ASKAP Spectral Line Survey (GASKAP), and the

Polarisation Sky Survey of the Universe’s Magnetism (POSSUM); as well as the deeper but smaller Deep Investigation of Neutral Gas Origins (DINGO) and the Commensal Real-Time ASKAP Fast-Transients survey (CRAFT; Macquart et al., 2010). Atop these future surveys, the recent Rapid ASKAP Continuum Survey (RACS; McConnell et al., 2020) has redefined our knowledge of the southern radio sky with shallow observations at 15 arcsecond resolution—compare to the previous largest radio survey, the NRAO VLA Sky Survey (NVSS; Condon et al., 1998), with 45 arcsecond resolution over the northern sky.

Machine learning methods for radio astronomy will be developed for and tested upon surveys like RACS and EMU. The path between an astronomical problem and a machine learning problem, however, is not a straightforward one. The goal when casting an astronomy question as something mathematical or computational is to convert the question into one with a known method of solution, such as classification or regression. Along the way, astronomical concepts and assumptions need to be turned into something a computer can deal with. Despite the wide availability of machine learning software and tools, there is no automatic or easy way to make this transformation. This necessitates research in astroinformatics.

This thesis concerns applications of machine learning to radio astronomy for the identification of extended extragalactic radio sources in wide-area surveys. We will present new methods of cross-identifying radio objects with their corresponding infrared and optical observations, demonstrate the applicability of these methods to existing wide-area radio surveys (shedding light on radio source population astronomy along the way), and develop a new way to identify complexity in polarised radio sources.

1.1 Problems in extragalactic radio astronomy

Galaxies produce radio emission through a variety of methods. The main emission mechanisms are star formation and active galactic nuclei (AGN, Section 2.3) and only the latter show extended structure well beyond the galaxy itself. AGN are the central focus of this thesis. They are intensely energetic objects at the centre of galaxies, which actively accrete matter and eject huge jets of plasma that develop into extended lobes over huge distances. Radio astronomy has many uses for AGN: their energy scales provide a test-bed for high energy physics, and the extremely bright lobes and jets can be seen throughout the Universe making AGN an accessible probe of the distant and old Universe.

AGN are thought to be critical to galaxy evolution and perhaps the early reionisation of the Universe (Bosch-Ramon, 2018), but their exact role in their host galaxies is an open question. The radiative and mechanical energy released by AGN impacts the interstellar medium (ISM) and is a key component of contemporary galaxy simulations and models (Morganti, 2017). The quenching of star formation due to AGN activity is called *AGN feedback*, the idea being that the energy expelled from an AGN is returned to the ISM of the galaxy, heating the gas so it cannot condense into stars. The

472 different impacts of radiative and mechanical energy, the location and scales within
473 the galaxy for which star formation is quenched, and whether star formation material
474 is ejected from the galaxy by the AGN are all open questions (Husemann & Harrison,
475 2018). One of our key results in Chapter 5 is an estimation of the mechanical energy
476 contributed to the intergalactic medium by AGN. Solving these questions requires an
477 understanding of AGN at both small and large scales, as well as connecting AGN to
478 their host galaxies at other wavelengths so that redshift, emission lines, star formation
479 rates, etc. can be determined. Chapter 4 describes our new method for performing
480 such cross-identifications.

481 The large scales of AGN also provide insight into the larger-scale structure of the
482 Universe. Giant radio galaxies for example (Section 2.3.1) are difficult to identify due
483 to their size and disconnected appearance (Section 2.6), but are so large that they can
484 be used to probe galaxy clusters (Banfield et al., 2016) and even the large-scale struc-
485 ture of the Universe (Reiprich et al., 2020). Other large-scale effects seem to exist, such
486 as the apparent alignment of radio galaxies ([Contigiani et al., 2017; Panwar et al., 2020](#))
487 ([Contigiani et al., 2017; Panwar et al., 2020; Taylor & Jagannathan, 2016](#)), though inves-
488 tigation continues as to whether this effect is real or due to some unknown systematic
489 bias. New radio surveys will reveal more radio sources than ever before, and if they can
490 be identified, radio structures in these surveys will allow us to investigate the structure
491 of the Universe.

492 The magnetic structure of AGN and their extended lobes may be probed by ra-
493 dio polarimetry observations (Anderson et al., 2015; Grant, 2011). Through polarisation,
494 though, extragalactic AGN can provide insight into our own Galaxy: the Far-
495 day depth and complexity (Section 2.2.2) of extragalactic radio sources can be used
496 to quantify local magnetic fields. With more polarised radio sources to be revealed
497 through upcoming wide-area polarisation surveys, the magnetic field of the Milky Way
498 and its surrounding intergalactic medium can be better resolved. Polarimetry also al-
499 lows us to determine some aspects of the structure of unresolved extragalactic sources,
500 even though we cannot spatially see that structure: spatially extended, polarised radio
501 sources may have different polarisation spectra to those which are spatially compact
502 when projected onto the sky.

503 Other problems in radio astronomy relate to the new level of data that we are about
504 to obtain from large telescopes like the SKA, which we discuss in Section 1.2.

505 1.2 Big data in astronomy

506 The scale of radio data underpins many of the methodology problems facing radio as-
507 tronomy. There are two main scientific benefits that come from large-scale data: better
508 statistics and more unusual objects. However, methods for dealing with radio data at
509 scale are still very much in their infancy, and need to be developed before instruments
510 like ASKAP and MeerKAT can be used to their full potential.

511 Many results in astronomy are statistical, from measuring the expansion of the
512 Universe to understanding the distribution of galaxy properties. With more observa-

tions we can not just narrow the uncertainty of these results, but also diversify them. When the number of objects under study is large, we can subdivide the population into subpopulations based on their physical properties and determine a statistic on each subpopulation. This can help understand the physical basis behind the statistic, or remove unwanted subpopulations from analysis. Even with less data it is still possible to subdivide or filter populations, but this will dramatically raise the uncertainty in the results due to the low sample size of each bin. An example of such a statistic is the radio luminosity function (RLF), which describes the density of radio sources throughout the Universe. It can be divided into a fractional RLF to examine the distribution of subpopulations, or to remove the effect of star-forming galaxies. We use large datasets to improve the uncertainty of RLFs and subdivide RLFs by the infrared properties of the population in Chapter 5.

With large datasets, highly unusual or rare objects are more likely to be included. Much of astronomy has been pushed forward by serendipitous discoveries, and (provided we have some way of combing through the dataset) large datasets should provide a wealth of such discoveries to be found. These may be found either through identifying objects where statistical methods seem to fail, or perhaps through direct searches (Norris, 2017a). By applying machine learning techniques, we found a number of new, rare giant radio galaxies in Chapter 5.

But with the benefits of big data come new challenges. At these new scales, the ability to store all of our science data is no longer a given. Many methods that previously had the luxury to run over a whole dataset at their own pace will now need to process data on-the-fly. This is in contrast to how most science observations are currently performed, with new discoveries coming from legacy surveys many years after they were conducted. Many discoveries are unplanned, and losing the ability to make serendipitous discoveries would be a major blow to astronomy (Norris, 2017a): how can we deal with so much data but still retain the ability to discover the unknown? Even with storage, the scale is tremendous: FIRST, for example, contains around 900 000 sources, of which very few were manually labelled. Over 10 000 volunteers labelled interesting objects in FIRST over four years of the Radio Galaxy Zoo project, with 75 000 aggregated labels passing quality assurance testing. While a phenomenal and, in radio, unparalleled labelling effort, it pales in comparison to the estimated 70 000 000 sources that EMU will find (Banfield et al., 2015).

546 1.3 Machine learning in astronomy

Machine learning is the process and practice of designing algorithms that automatically find and exploit useful patterns in data (Deisenroth et al., 2020). It is best suited for scenarios where these patterns are hard to encode by hand. Typically hard-to-describe problems might include data exploration, data visualisation, or interpolation. We face a deluge of data, and our existing methods for data analysis and astronomical discovery are hard to automate due to the often complex, [visual-imagery-based](#) nature of much of astronomy. This is where machine learning comes in: if we can find

554 a way to encode the processes underpinning astronomy as algorithms, we could apply
555 them at scale. Machine learning has found much success in other fields with similar
556 difficult-to-describe problems like image classification and outlier identification.

557 From the other direction, machine learning finds interesting and unique applica-
558 tions in radio astronomy. Radio images are usually quite different from images in the
559 typical image analysis problems faced in computer science. Standard assumptions are
560 frequently broken: noise is correlated across the image, objects can be different shapes
561 when viewed at different wavelengths, and the number of colours in an image may
562 be in the hundreds. Data points are spatially located and can be close together or far
563 apart in ways that are not solely based on how they look or behave. We often do not
564 have the option of obtaining more data, sometimes because it is prohibitively expen-
565 sive (e.g. to undertake a new survey) and sometimes because it simply does not exist
566 (as we only have one Universe to look at or because the event is a transient one-off).
567 The burgeoning field of astroinformatics promises to prove interesting.

568 1.4 How this thesis fits in

569 Current methods of identifying radio sources in wide-area radio surveys work best
570 for compact, isolated objects. As resolution and depth of radio surveys increases in
571 the lead-up to the SKA, these limitations are becoming more apparent: resolution in-
572 creases apparent complexity and depth increases the number of radio sources visible
573 in any given patch of sky.

574 Source identification is critical for analysis of radio sources in large-scale surveys
575 whether one is interested in individual sources or their bulk properties. Without solv-
576 ing issues like cross-identification and Faraday complexity we will not be able to make
577 full use of the SKA when it arrives. As we will discuss in Section 2.5, cross-identification
578 of radio sources with their counterparts in infrared or optical is key to understanding
579 these sources. Many [radio](#) galaxy properties can only be derived from [non-radio](#) ob-
580 servations of the galaxy, the stellar component of which can be traced by emission in
581 the infrared and optical. This can yield properties including distance, star formation
582 rate, and even mass of the central black hole powering the radio source itself. We de-
583 velop an automated machine learning method for cross-identification in Chapter 4 and
584 apply it to obtain physical results in Chapter 5.

585 Bulk analysis of polarised sources greatly benefits from being able to assess their
586 Faraday complexity. The most important application is perhaps in the development of
587 so-called *rotation measure grids*, which help characterise the magnetic field of the Milky
588 Way and surrounding intergalactic medium. Without knowing which sources are com-
589 plex and which are not, we cannot estimate the grid reliably. A fast way to estimate
590 Faraday complexity also allows us to quickly determine which sources may need fol-
591 lowing up, either with more expensive algorithms or further observations. Chapter 6
592 details our development of an automated Faraday complexity classifier.

593 **1.5 Thesis outline**

594 We begin by introducing key concepts from radio astronomy in Chapter 2, including
 595 radio active galactic nuclei (AGN), as well as motivations for and difficulties in cross-
 596 identifying observed radio emission from AGN. Chapter 3 introduces machine learn-
 597 ing and describes the machine learning background required for the remainder of the
 598 thesis. These chapters together comprise the background knowledge.

599 Chapter 4 is my paper *Radio Galaxy Zoo: Machine learning for radio source host galaxy*
 600 *cross-identification* (Alger et al., 2018). This chapter describes a new, machine learn-
 601 ing method for cross-identifying extended radio emission with host galaxies in the
 602 infrared. We apply this approach to cross-identifying all extended radio sources in
 603 the Faint Images of the Radio Sky at Twenty Centimeters radio survey (FIRST; Becker
 604 et al., 1995) with their infrared counterparts in AllWISE (Cutri et al., 2013) and use the
 605 resulting catalogue of cross-identifications to create a fractional radio luminosity func-
 606 tion in Chapter 5, which itself is my paper *Radio Galaxy Zoo: Radio luminosity functions*
 607 *of extended sources* (Alger et al., in prep.). Chapter 6 is my paper *Interpretable Faraday*
 608 *Complexity Classification* (Alger et al., 2021), which introduces an interpretable method
 609 for classifying radio emission as Faraday complex or simple, which could be used to
 610 identify whether it is an extended source that is below the resolution limit. Chapter 7
 611 provides a discussion of the insights gained from the findings of this thesis and sug-
 612 gests future directions for research building on our results.

613 **1.6 Contributions**

614 My main contributions to radio astroinformatics in this thesis are:

- 615 • I introduce a new method for cross-identifying radio emission which can learn
 616 from existing catalogues, the first application of machine learning to radio cross-
 617 identification (Chapter 4);
- 618 • I demonstrate an application of this new method to the creation of fractional ra-
 619 dio luminosity functions, which require considerably more cross-identifications
 620 than non-fractional radio luminosity functions, in the process creating the largest
 621 available catalogue of extended, cross-identified radio sources (Chapter 5);
- 622 • I produce a fractional radio luminosity function with divisions based on mid-
 623 infrared colours associated with the host galaxies of the radio emission, helping
 624 to understand how radio galaxies evolve throughout the Universe (Chapter 5);
- 625 • I highlight the requirement of considerably more redshifts in understanding fu-
 626 ture wide-area radio surveys (Chapter 5);
- 627 • I introduce a new method to identify Faraday complexity using an interpretable
 628 classifier, as well as features for Faraday dispersion functions which can be used
 629 for other machine learning tasks (Chapter 6); and

-
- ⁶³⁰ • I apply machine learning to real radio spectropolarimetric data for the first time
⁶³¹ (Chapter 6).

REVISION DIFF

632 1.7 Works produced during this PhD

633 During my PhD, I had the good fortune to collaborate with many talented authors
 634 and contribute to a number of related works. The following is a list of papers which I
 635 co-authored in this time:

- 636 • *Radio Galaxy Zoo: Machine learning for radio source host galaxy cross-identification*
 637 (2018). **M. J. Alger**, J. K. Banfield, C. S. Ong, L. Rudnick, O. I. Wong, C. Wolf,
 638 H. Andernach, R. P. Norris, and S. S. Shabala. *Monthly Notices of the Royal Astro-*
 639 *nomical Society* 478, pp. 5547–5563. [This is the content of Chapter 4 with minor](#)
 640 [modifications.](#)
- 641 • *Radio Galaxy Zoo: Radio luminosity functions of extended sources* (in prep.). **M. J.**
 642 **Alger**, O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. Andernach, L. Rudnick,
 643 S. S. Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, R. P. Norris, and A. J. M.
 644 Thomson. [This is the content of Chapter 5 with minor modifications.](#)
- 645 • *Interpretable Faraday Complexity Classification* ([in press 2021](#)). **M. J. Alger**, J. D. Liv-
 646 ington, N. M. McClure-Griffiths, J. L. Nabaglo., O. I. Wong, and C. S. Ong. *Pub-*
 647 *lications of the Astronomical Society of Australia* [-38, E022](#). [This is the content of](#)
 648 [Chapter 6 with minor modifications.](#)
- 649 • *Radio Galaxy Zoo Data Release 1: visual identification of 75,641 radio morphologies from*
 650 *the FIRST and ATLAS surveys* (in prep.). O. I. Wong, A. F. Garon, **M. J. Alger**, K.
 651 W. Willett, L. Rudnick, J. K. Banfield, J. Swan, S. S. Shabala, H. Andernach, R. P.
 652 Norris, B. D. Simmons, A. D. Kapińska, N. Seymour, et al.
- 653 • *Radio Galaxy Zoo: CLARAN - a deep learning classifier for radio morphologies* (2019).
 654 C. Wu, O. I. Wong, L. Rudnick, S. S. Shabala, **M. J. Alger**, J. K. Banfield, C. S. Ong,
 655 S. V. White, A. F. Garon, R. P. Norris, H. Andernach, J. Tate, V. Lukic, H. Tang,
 656 K. Schawinski, and F. I. Diakogiannis. *Monthly Notices of the Royal Astronomical*
 657 *Society* 482, pp. 1211–1230.
- 658 • *Heightened Faraday Complexity in the inner 1 kpc of the Galactic Centre* (2021), by J.
 659 D. Livingston, N. M. McClure-Griffiths, B. M. Gaensler, A. Seta, and **M. J. Alger**.
 660 *Monthly Notices of the Royal Astronomical Society* 502, pp. 3814–3828.
- 661 • *Radio Galaxy Zoo: Unsupervised Clustering of Convolutionally Auto-encoded Radio-*
 662 *astronomical Images* (2019). N. O. Ralph, R. P. Norris, G. Fang, L. A. F. Park, T. J.
 663 Galvin, **M. J. Alger**, H. Andernach, C. Lintott, L. Rudnick, S. S. Shabala, and O. I.
 664 Wong. *Publications of the Astronomical Society of the Pacific* 131, 108011.
- 665 • *Radio Galaxy Zoo: Knowledge Transfer Using Rotationally Invariant Self-organizing*
 666 *Maps* (2019). T. J. Galvin, M. Huynh, R. P. Norris, X. R. Wang, E. Hopkins, O. I.
 667 Wong, S. S. Shabala, L. Rudnick, **M. J. Alger**, and K. L. Polsterer. *Publications of*
 668 *the Astronomical Society of the Pacific* 131, 108009.

669 I also gave a number of conference talks:

- 670 • *Machine Learning Methods for Radio Host Cross-Identification with Crowdsourced La-*
671 *bels*, presented in 2017 at the seventh SKA Pathfinder Radio Continuum Survey
672 meeting, Perth, Australia.
- 673 • *Learning to Cross-identify Wide-area Radio Surveys with Radio Galaxy Zoo: Data chal-*
674 *lenges in citizen science*, presented in 2018 at the Collaborative Conference on Com-
675 putational and Data Intensive Science, Melbourne, Australia.
- 676 • *Radio luminosity functions with Radio Galaxy Zoo and machine learning*, presented
677 in 2019 at the ninth SKA Pathfinder Radio Continuum Survey meeting, Lisbon,
678 Portugal.
- 679 • *Radio luminosity functions with machine learning and Radio Galaxy Zoo*, presented in
680 2019 at the Annual Scientific Meeting of the Astronomical Society of Australia.
- 681 • *Extracting Meaningful Features from Early-Science Radio Data*, presented in 2019 at
682 the Collaborative Conference on Computational and Data Intensive Science, Can-
683 berra, Australia.
- 684 • *Extracting Meaningful Features from Early-Science Radio Data*, presented in 2019 at
685 Artificial Intelligence in Astronomy, Garching, Germany.

REVISION DIFF

687 Radio Sources

688 As its title suggests, this thesis is focused on the identification of extended radio sources.
 689 This chapter introduces extended radio sources, describing what we see when we look
 690 at the sky with radio eyes and radio telescopes. We will discuss the different kinds of
 691 radio sources that we can observe, how they are distributed throughout the Universe,
 692 and key issues surrounding their identification. We will start by looking at the extra-
 693 galactic radio sky, introducing what it is we see through radio telescopes (Section 2.1).
 694 Then, we will describe the physics behind radio emission, including the emission mech-
 695 anisms that allow us to observe active galactic nuclei in radio, and how radio polarisa-
 696 tion can help us determine aspects of distant magnetic fields (Section 2.2). From there
 697 we will turn to active galactic nuclei as they will be the objects of interest in this the-
 698 sis (Section 2.3): what do they look like, how are they structured, and what do they
 699 do? Finally, we will cover three different tasks we may face when presented with a
 700 collection of observed AGN, including classification (Section 2.4), cross-identification
 701 (Section 2.5), and source aggregation (Section 2.6).

702 **2.1 The Extragalactic Radio Sky**

703 The extragalactic sky appears quite different at different wavelengths. While an op-
 704 tical observer may look at a distant galaxy and see spirals and halos, an infrared ob-
 705 server will see discs and dust. What does the radio astronomer see? Figure 2.1 shows
 706 a false-colour image of the radio sky from 72–231 MHz. The plane of the Milky Way is
 707 clearly visible through the centre, but nearly every other object in this image is a galaxy.
 708 These galaxies fall into two main categories: those that emit radio due to star forma-
 709 tion (called *star-forming galaxies*), and those that emit radio due to *active galactic nuclei*
 710 (AGN; called *radio galaxies* in this thesis). AGN can be observed at many frequencies:
 711 for the remainder of this thesis, except where otherwise noted, ‘AGN’ refers to *radio*
 712 ‘AGN’.

713 Non-AGN emission from distant galaxies traces the recent star-formation rate (SFR).
 714 Besides low-power thermal emission, stellar radio emission from galaxies mainly comes
 715 from massive ($\gtrsim 8 M_{\odot}$) stars, through two emission mechanisms. The first is through
 716 H II regions, which are ionised by such stars. The ionised electrons emit bremsstrahlung
 717 radiation at radio wavelengths. The second emission mechanism is supernovae. Mas-

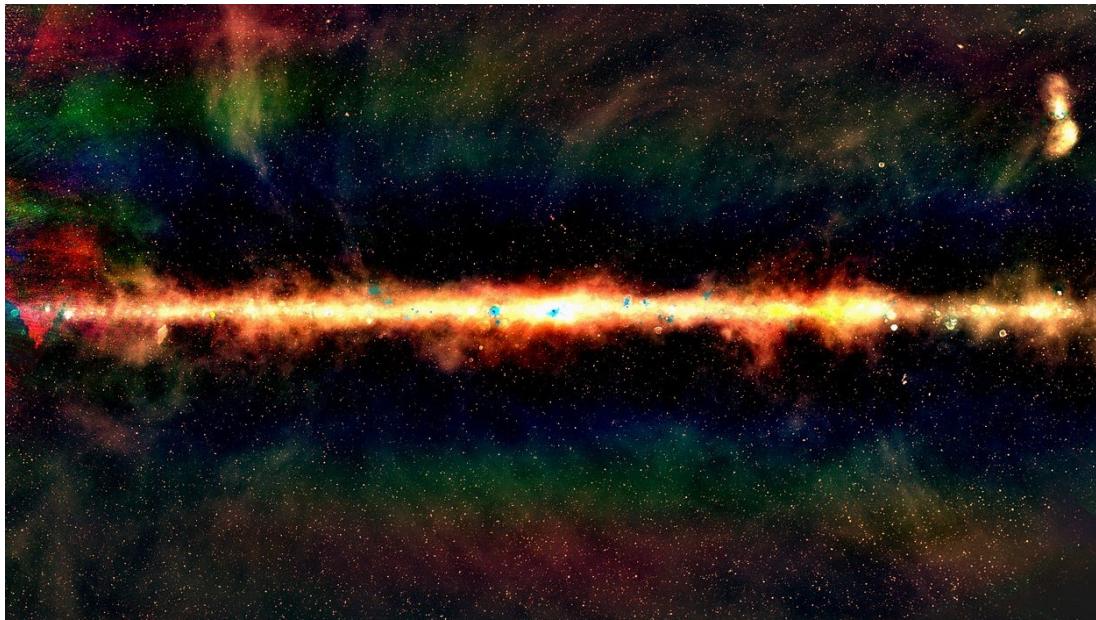


Figure 2.1: False-colour image of the radio sky from the GLEAM survey. (Image: Natasha Hurley-Walker, Curtin University/ICRAR; Hurley-Walker et al., 2017)

718 sive stars may end their lives in Type II and Type Ib supernovae, which can result in
 719 supernova remnants. **These remnants emit** **Interaction of the supernova remnant with**
 720 **the interstellar medium (ISM) causes the emission of** synchrotron radiation. Massive
 721 stars like these are short-lived (a few 10^6 yr), and the corresponding emitting electrons
 722 have similarly short lifetimes ($\lesssim 10^8$ yr). The radio effects of these stars are therefore
 723 also short-lived, which is why radio emission traces the recent SFR (Condon, 1992).
 724 Star formation-associated emission is mainly found in the disc of spiral galaxies, as
 725 this is where the star formation rate is highest. In particular, there is no star-forming
 726 radio emission extending outside of the galaxy proper. The radio power emitted by
 727 these galaxies at 1.4 GHz is on the order of 10^{18} – 10^{23} W Hz $^{-1}$ (Condon, 1992). For a
 728 radio survey like the NRAO VLA Sky Survey (NVSS; Condon et al., 1998), with a detec-
 729 tion limit of 2.3 mJy, this luminosity range corresponds to a maximum redshift range
 730 of 0.0004–0.1272 (corresponding to 6×10^6 – 1.646×10^9 yr lookback time¹). Upcoming
 731 surveys such as the Evolutionary Map of the Universe (EMU; Norris et al., 2011), with
 732 5 σ detection thresholds of 50 μ Jy (Norris et al., 2011), will push this redshift range to
 733 0.0030–0.6684 (corresponding to 4.2×10^7 – 6.261×10^9 yr lookback time).

734 AGN are energetic objects at the centre of galaxies, powered by accretion into su-
 735 permassive black holes. The extended, strongly-magnetised plasma they eject emits
 736 synchrotron radiation from accelerating relativistic electrons, which is what we see
 737 when we observe a radio galaxy. The radio luminosity of a radio galaxy can range
 738 from 10^{20} – 10^{28} W Hz $^{-1}$ (Pracy et al., 2016) at 1.4 GHz, making them some of the most

¹Assuming cosmological parameters from Chapter 5. Calculated using “A Cosmology Calculator for the World Wide Web” (Wright, 2006).

739 luminous objects in the Universe. They are therefore visible throughout the Universe,
740 with the most distant AGN detected at a redshift of 7.5 (Bañados et al., 2018). De-
741 pending on the orientation and type of AGN, as well as its interaction with its host
742 galaxy, the radio emission may extend far beyond the galaxy itself—up to megaparsec
743 scales—and this emission may have complex structure. Perhaps the most impressive
744 local example is Centaurus A (Cen A), the prominent double-lobed cloud in the upper-
745 right of Figure 2.1 extending over 8 degrees across the sky. Section 2.3 discusses AGN
746 in more detail.

747 Most ~~radio galaxies are~~ AGN emission is compact and unresolved in any given
748 radio survey due to the distance at which ~~they~~it can be detected and ~~their~~the ori-
749 entation or type of the galaxy. This means that ~~their~~observed structure does not al-
750 ways help to distinguish AGN radio emission from star-forming radio emission. How
751 can we tell these apart? Synchrotron emission has a considerably steeper spectral in-
752 dex than bremsstrahlung, but synchrotron emission dominates the bremsstrahlung in
753 star-forming galaxies at 1.4 GHz (Condon, 1992). Truly star-forming galaxies can be
754 distinguished from AGN host galaxies by using optical spectroscopy (e.g. Groves &
755 Kewley, 2007; Mauch & Sadler, 2007), but radio emission is detectable at much greater
756 distances than good quality optical spectra can be obtained at, making this solution im-
757 practical for many galaxies. Separating star-forming galaxies from AGN host galaxies
758 at radio wavelengths remains a difficult problem in radio astronomy.

759 Polarised radio surveys can provide extra information. While radio emission due
760 to star formation tends to not have detectable polarisation, AGN may be very strongly
761 polarised. This makes polarisation an excellent indicator of whether a source is an
762 AGN, though very incomplete: many AGN will also not have detectable polarisation,
763 and the polarised intensity is usually less than ten per cent of the total radio inten-
764 sity, meaning we detect far fewer polarised radio sources than we do radio sources in
765 general.

766 From the size scales described above, it should be clear that a survey of extended
767 radio sources will be dominated by AGN. Nevertheless, star-forming galaxies present a
768 significant part of the radio population, and the fraction of the radio sky they comprise
769 varies significantly with survey parameters.

770 2.2 Radio emission

771 Electromagnetic radiation in radio frequencies—about 10 MHz–1 THz (Condon & Ran-
772 som, 2016)—is called *radio emission*. This is a very broad range of frequencies and so
773 radio astronomy covers a very broad range of astrophysical phenomena, from cosmo-
774 logical background radiation to neutron stars. The focus of this thesis is the excit-
775 ing, dynamic, and so-called ‘violent universe’ of radio galaxies. These galaxies are ob-
776 served through their emission of synchrotron radiation and are studied through their
777 observed physical structure, the intensity and spectroscopic properties of their radia-
778 tion, and the polarisation and spectropolarimetric properties that are uniquely visible
779 in radio. This section introduces synchrotron radiation and radio polarisation.

780 **2.2.1 Synchrotron radiation**

Most radio emission from radio galaxies is *synchrotron radiation*, produced by relativistic charged particles accelerating in a magnetic field. A non-relativistic charged particle will spiral with a fixed angular frequency when it moves in a magnetic field in a process called *gyro radiation*. Synchrotron radiation is a relativistic effect: it can be thought of as gyro radiation which has been Lorentz transformed to energies much greater than mc^2 . The spectrum of *optically thin* synchrotron radiation follows a power law (Condon & Ransom, 2016):

$$S(\nu) \propto \nu^\alpha. \quad (2.1)$$

where ν is the frequency of radiation and α is called the *spectral index*². It is related to the energy distribution of the emitting electrons: assuming that the electron energy distribution follows a power law (which it generally does, Rybicki & Lightman, 2008), where the number density of electrons at a given energy E is given by

$$n(E) \propto E^\Gamma, \quad (2.2)$$

then

$$\alpha = \frac{\Gamma - 1}{2}. \quad (2.3)$$

781 The spectral index for synchrotron radiation tends to range from -2 to 0 (Condon &
782 Ransom, 2016) with spectral indices greater than 0 called ‘inverted’ spectra.

783 **2.2.2 Polarisation**

784 Electromagnetic radiation consists of waves of self-propagating, orthogonal electric
785 and magnetic fields. The orthogonality of these two waves allows us to characterise
786 the radiation just by the electric field. As a transverse wave, the electric field travels at
787 an angle in the plane perpendicular to the line-of-sight. This angle and its behaviour
788 is called the *polarisation* of the wave.

The polarisation can be characterised by decomposing the electric field into orthogonal components E_x and E_y , letting \hat{z} denote the axis of propagation:

$$\vec{E} = (\hat{x}E_x \exp(i\varphi_x) + \hat{y}E_y \exp(i\varphi_y)) \exp(i(\vec{k} \cdot \hat{z} - \omega t)). \quad (2.4)$$

In an astronomical context, \hat{z} is the line-of-sight from the source of the radiation to the observer. \vec{k} is the *wave vector* which points in the direction of travel and has magnitude $2\pi/\lambda$, and $\omega = 2\pi\nu$ is the *angular frequency*. φ_x and φ_y are the phase offsets of each component. As this wave propagates along the line-of-sight toward an observer, the electric field oscillates in an ellipse across the x - y plane. When the two components are in phase, this ellipse is degenerate and the radiation is called *linearly polarised*. When the two components are perfectly out of phase, the ellipse is a circle, and the radiation is called *circularly polarised*. Of course, any ellipse in between these extremes is also possible. For this reason, we decompose the polarisation into linearly polarised

²Note that the sign of α varies by convention, and both $S \propto \nu^\alpha$ and $S \propto \nu^{-\alpha}$ exist in the literature.

components and a circularly polarised component, called *Stokes parameters* (Condon & Ransom, 2016; Stokes, 1851). These are:

$$I = \frac{1}{R_0} \mathbb{E}_t [E_x^2 + E_y^2], \quad (2.5)$$

$$Q = \frac{1}{R_0} \mathbb{E}_t [E_x^2 - E_y^2], \quad (2.6)$$

$$U = \frac{1}{R_0} \mathbb{E}_t [2E_x E_y \cos(\varphi_x - \varphi_y)], \quad (2.7)$$

$$V = \frac{1}{R_0} \mathbb{E}_t [2E_x E_y \sin(\varphi_x - \varphi_y)]. \quad (2.8)$$

\mathbb{E}_t denotes the expectation value over time. I is the *total intensity* of the radiation. Q and U together describe the linear polarisation and together can be used to define the *polarisation angle* χ :

$$\tan(2\chi) = \frac{U}{Q}. \quad (2.9)$$

V is the circular polarisation and describes the eccentricity of the ellipse. For most extragalactic sources, the contribution of circular polarisation is tremendously small, especially compared to that linear polarisation, and can be assumed to be zero (Rayner et al., 2000; Saikia & Salter, 1988). Incoherent radiation may be composed of radiation with many different polarisations, and these polarisations may fully or partially cancel out: this is called *unpolarised* or *partially-polarised* radiation respectively. The total intensity of polarised radiation is called the *polarised intensity* P and is given by

$$P^2 = Q^2 + U^2 + V^2. \quad (2.10)$$

Note that $P^2 \leq I^2$. The *fractional polarisation* is the ratio between these two intensities:

$$p = \frac{P}{I}. \quad (2.11)$$

789 The synchrotron radiation from radio galaxies is polarised, though this polarisation
 790 is not always detectable as the polarised signal tends to be much weaker than the total
 791 intensity, on the order of ten per cent (O'Sullivan et al., 2015). Additionally, the most
 792 common non-AGN cause for radio emission is star formation, which does not gener-
 793 ally have detectable polarisation in extragalactic surveys. Polarisation is therefore an
 794 excellent way to confirm that a radio source is an AGN.

Polarisation can also be used to describe the magnetic structure of both the radio galaxy jets and lobes as well as the intervening medium. As polarised light from distant galaxies makes its way to us, magnetised plasma along the way can cause the polarisation angle to rotate due to the Faraday effect. The amount of rotation is called the *Faraday depth* ϕ , and is related to the electron density n_e and the line-of-sight magnetic

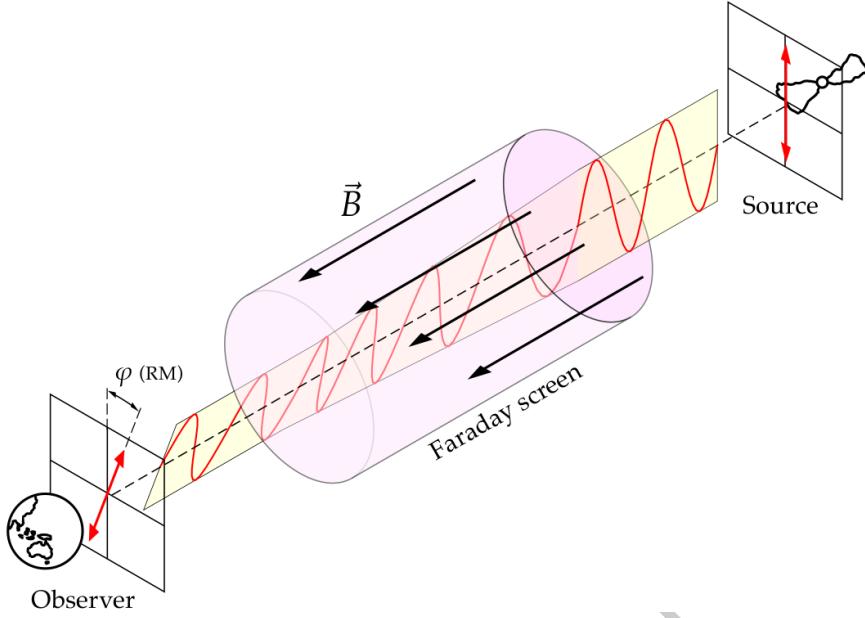


Figure 2.2: The Faraday effect for a single Faraday screen with magnetic field \vec{B} inducing a Faraday rotation of ϕ . Adapted from a figure by Bob Mellish.

field strength $\vec{B} \cdot \hat{z}$ of the intervening medium:

$$\phi(x, y) = \frac{e^3}{8\pi^2 \epsilon_0 m_e^2 c^3} \int_{\text{there}}^{\text{here}} n_e(x, y, z) \vec{B}(x, y, z) \cdot d\vec{z} \text{ rad m}^{-2}. \quad (2.12)$$

Here $d\vec{r}$ is the infinitesimal path length in pc (Brentjens & de Bruyn, 2005). Within the synthesised beam of a radio telescope there may be multiple lines-of-sight that go through different media and hence have different Faraday depths. An example of this is a radio galaxy that is sufficiently far away that its structure is unresolved by the telescope, and yet has different polarisation properties across its breadth. The leading constant of Equation 2.12 is around $2.62 \times 10^{-13} \text{ T}^{-1}$, more commonly written as $0.812 \text{ pc } \mu\text{G}^{-1} \text{ cm}^{-1}$ in CGS units with B in μG and z in pc. The amount of polarised radiation at each Faraday depth can be characterised by the *Faraday dispersion function* (FDF) or *Faraday spectrum* of the source, usually denoted $F(\phi) \in \mathbb{C}$. F is defined implicitly by its relationship with the polarised radiation P observed at wavelength λ :

$$P(\lambda^2) = \int_{-\infty}^{\infty} F(\phi) e^{2i\lambda^2\phi} d\phi. \quad (2.13)$$

⁷⁹⁵ One useful way of thinking about this equation is that F is the decomposition of $P(\lambda^2)$ into complex sinusoids of the form $e^{2i\lambda^2\phi}$.

⁷⁹⁶ If observed radiation has precisely one Faraday depth ϕ , then the polarised structure is called a *Faraday screen* and the source is called *Faraday simple*. In this degenerate case, the relationship between the polarisation angle χ and the squared wavelength λ^2

is linear:

$$\chi = \chi_0 + \phi\lambda^2, \quad (2.14)$$

and the FDF is a delta distribution:

$$F(\phi) = \delta(\phi - \phi). \quad (2.15)$$

797 ϕ is then called the *rotation measure* (RM). If the source is not Faraday simple, then it
798 is called *Faraday complex*, and the question of whether a source is Faraday simple or
799 Faraday complex is called *Faraday complexity*. A diagram of a Faraday screen is shown
800 in Figure 2.2. Until very recently, the frequency resolution of polarised surveys was
801 insufficient to meaningfully separate most complex arrangements of Faraday depths,
802 and so most sources were assumed to be simple and characterised entirely in terms of
803 their rotation measure (e.g. Taylor et al., 2009). Advancing telescope technology and
804 emphasis on polarisation science has opened new frontiers in spectropolarimetry and
805 upcoming and ongoing surveys (e.g. RACS and POSSUM) will likely report Faraday
806 complexity and produce Faraday depth catalogues instead of rotation measures.

If the polarised spectrum of a Faraday complex source is observed at multiple frequencies, then the multiple Faraday depths comprising it can be disentangled even though they spatially overlap in the radio image. This can provide insight into the polarised structure of the source as well as the intervening medium. This disentanglement is accomplished by inverting Equation 2.13, a process called *RM synthesis* (Brentjens & de Bruyn, 2005):

$$F(\phi) = \int_{-\infty}^{\infty} P(\lambda^2) e^{-2i\lambda^2\phi} d\lambda^2. \quad (2.16)$$

In reality we do not observe $P(\lambda^2)$ at all wavelengths nor with infinite resolution. In RM synthesis this is accounted for by the introduction of a *weighting function* (or *windowing function*, e.g. Heald, 2008) $W(\lambda^2)$. $W(\lambda^2)$ is nonzero if and only if an observation was taken with wavelength λ . Substituting $P(\lambda^2) \rightarrow P(\lambda^2)W(\lambda^2)$ into Equation 2.16 results in a sum which can be numerically evaluated:

$$F(\phi) \approx \int_{-\infty}^{\infty} P(\lambda^2)W(\lambda^2) e^{-2i\lambda^2\phi} d\lambda^2 = \sum_{j=1}^J P(\lambda_j^2)W(\lambda_j^2) e^{-2i\lambda_j^2\phi}. \quad (2.17)$$

807 $P(\lambda_j^2)$ is the observed polarisation at the j th value of wavelength, $W(\lambda_j^2)$ is the corre-
808 sponding j th weight, and J is the total number of wavelengths for which measurements
809 were taken. The weighting function W is analogous to the weighting function in radio
810 synthesis imaging. The most common choices of W are 1) uniform weighting³ with
811 $W(\lambda_j^2) = 1$ for all nonzero values, and 2) weighting by the inverse variance at each
812 wavelength.

Of course, no physical source has a precise Faraday depth, as there is always intrin-

³The analogous weighting scheme in radio synthesis imaging would be natural weighting, rather than uniform—an unfortunate overlap in terminology.

sic scatter. Along the line of sight, if we assume **Gaussian noise in an otherwise-constant that n_e is observed with random Gaussian noise** i.e. $n_e(z) \sim \mathcal{N}(\bar{n}_e, \sigma_{n_e}^2)$, and **a constant that B is constant** for simplicity, then we find

$$\phi \sim \mathcal{N}\left(\frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} B \bar{n}_e, \frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} B \sigma_{n_e}^2\right), \quad (2.18)$$

that is, the depth has an uncertainty proportional to the magnetic field strength and the noise in n_e . A similar result follows for noise in B only. There is no analytic solution for noise in both B and n_e , but if we approximate the integrand as a Gaussian by calculating the mean and variance, we find

$$\phi \sim \mathcal{N}\left(\frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} \frac{\bar{n}_e \sigma_B^2 + \bar{B} \sigma_{n_e}^2}{\sigma_B^2 + \sigma_{n_e}^2}, \frac{e^3}{8\pi^2\epsilon_0 m_e^2 c^3} \frac{\sigma_{n_e}^2 \sigma_B^2}{\sigma_{n_e}^2 + \sigma_B^2}\right). \quad (2.19)$$

813 We observe multiple lines-of-sight that are coalesced into one within the beam. Due
 814 to this noise, even with constant n_e and B across a source, we can see multiple Faraday
 815 depths as each line-of-sight is a sample from the above distribution.

816 2.3 Radio galaxies and active galactic nuclei

817 AGN are some of the most energetic objects in the Universe. They both provide a labo-
 818 ratory for extreme physics and are a key part of the life cycle of a galaxy (Heckman &
 819 Best, 2014). Powered by a supermassive black hole, they convert gravitational poten-
 820 tial energy into intense electromagnetic radiation at a broad range of frequencies. AGN
 821 that produce strong radio emission are called radio AGN, and methods of observing
 822 the complex structures that these radio AGN form as radio galaxies are the focus of
 823 this thesis.

824 2.3.1 What we see when we look at AGN

825 Observations are the crux of astronomy. While there are many models of how AGN
 826 evolve and how they interact with their surroundings—and indeed, the actual struc-
 827 ture of an AGN is very much an open question in astronomy—the evidence presented
 828 by observations is reliable and a good place to start discussing the structure, behaviour,
 829 and importance of AGN throughout the Universe.

830 As powerful sources of radio emission, radio AGN and their associated extended
 831 structure can be seen throughout the Universe. Sufficiently close or large radio galax-
 832 ies can be resolved by telescopes and their structure examined, while more distant or
 833 smaller radio galaxies may be unresolved and point-like. A well-resolved radio galaxy
 834 can be a striking thing: from the central AGN extend two opposing, tightly-collimated
 835 jets, which widen into huge lobes of radio-bright plasma. These lobes may have fur-
 836 ther structure, particularly bright regions called *hot-spots*, and the jets and lobes may
 837 be bent and distorted as they travel away from their host galaxy. For any given radio

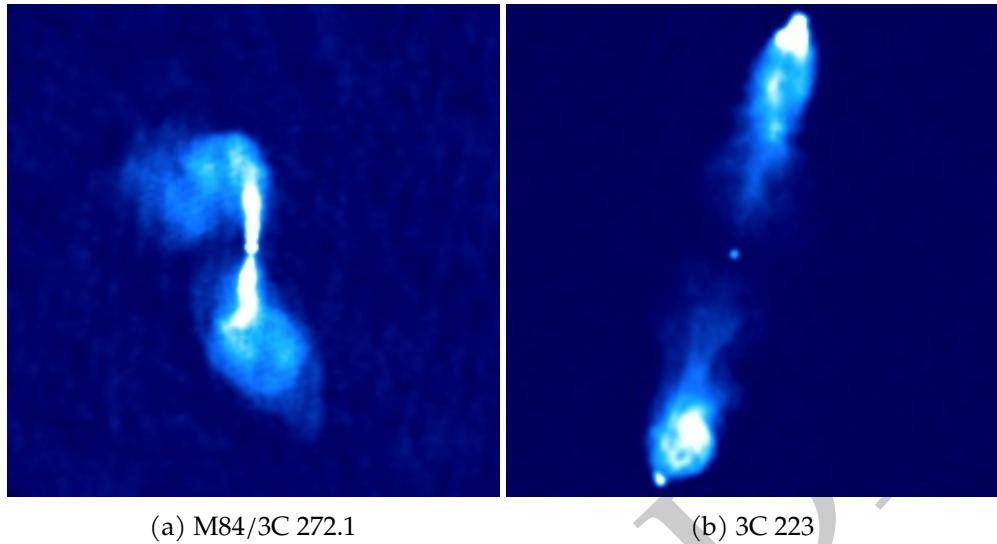


Figure 2.3: Examples of (a) a FRI (Laing & Bridle, 1987) and (b) a FRII radio galaxy (Leahy & Perley, 1991). Both are shown with an arcsinh stretch and were observed with the VLA.

838 galaxy, some of these features may or may not be present. In particular, radio galaxies
 839 are often divided into two classes based on the kinds of extended structure that are vis-
 840 ible, called Fanaroff-Riley type I (FRI) and Fanaroff-Riley type II (FRII) radio galaxies.
 841 FRI have wavy, diffuse lobes, appearing brighter toward the host galaxy and dimming
 842 further out (e.g. Figure 2.3a). FRII, on the other hand, have long, tightly-collimated jets
 843 and sharp-edged lobes with bright hot-spots (Urry & Padovani, 1995) at the very end
 844 of the lobes, and are brighter further away from the host galaxy (e.g. Figure 2.3b). FRII
 845 are also generally higher-luminosity (Fanaroff & Riley, 1974) than FRI, and therefore
 846 make up the majority of observed extended radio sources throughout the Universe.
 847 However, this is by no means the clear-cut divide it was once thought to be (Mingo
 848 et al., 2019) with the difference now being attributed largely to environmental effects
 849 rather than jet power. The current understanding is that FRII jets remain at relativistic
 850 energies up until the edge of the lobe, where they terminate in a shock that appears
 851 as a hot-spot, while FRI jets decelerate within the galaxy itself (Hardcastle & Croston,
 852 2020). This sharp difference in extended structure begins with environmental interac-
 853 tions at the very centre of the galaxy.

854 A radio galaxy can be tremendously extended, with increasingly many radio galax-
 855 ies being found with a length of over one megaparsec. Such large galaxies are called
 856 *giant radio galaxies*, but even non-giants are still quite big, regularly extending well out-
 857 side the stellar component of the host galaxy. We will discuss the extended structure
 858 in Section 2.3.2. Appendix J lists some giants discovered during the work of this thesis.

859 An AGN interacts with its host galaxy, and so the host galaxy of an AGN can also
 860 provide interesting insights into the structure and behaviour of the AGN. Early re-
 861 search indicated that the split between FRI and FRII radio galaxies was dependent
 862 on the mid-infrared and optical brightness (and therefore density) of the host galaxy
 863 (Bicknell, 1995; Ledlow & Owen, 1996) though more recent work suggests this may

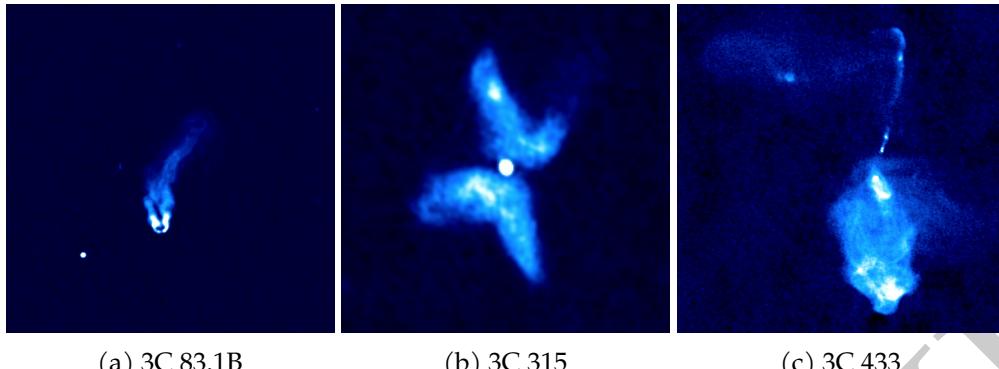


Figure 2.4: Radio galaxies, displayed with an arcsinh colour scale. All images were taken with the VLA. (a) is a narrow-angled tail radio galaxy (Leahy et al., n.d.), (b) is an X-shaped radio galaxy (Leahy et al., 1986), and (c) is a very unusually-shaped radio galaxy (Black et al., 1992).

not be a strong effect if it exists at all (Hardcastle & Croston, 2020). Chapter 5 investigates the distribution of radio luminosities conditioned on the mid-infrared colour of the host galaxy.

2.3.2 Extended structure

The jets and lobes of AGN can be very extended, with the largest known radio galaxies measuring over 4 Mpc end-to-end (Machalski et al., 2011). This is a much larger size than the radii of the host galaxies, and so the jets and lobes of AGN are uniquely poised to interact with the local environment. Environmental interactions both within and outside the host galaxy warp and distort the jets and lobes. Within the galaxy, the jets drive a bubble of energy in the [interstellar medium \(ISM; Mukherjee et al., 2016\)](#) [ISM \(Mukherjee et al., 2016\)](#), transferring energy into the ISM with different effects depending on the jet power (Mukherjee et al., 2018); the ISM on the other hand suppresses the jets and distorts them to varying amounts depending on the degree of interaction (Mukherjee et al., 2018). Outside the galaxy, the jets and lobes are bent by the intra-cluster medium and neighbouring galaxies (ICM; Garon et al., 2019; Rodman et al., 2019) and this structure may even be used as a probe for cluster environments (Banfield et al., 2016; Sakelliou et al., 2008).

The strong interaction of AGN with their environment leads to a great variety of exotic-shaped radio galaxies. Some morphological classes of this ‘radio galaxy zoo’ include X-shaped galaxies, which have two sets of lobes roughly perpendicular to each other; wide- and narrow-angled tail galaxies, which are bent about the core with large and small angles respectively; head-tail galaxies, which are so bent that the two lobes seem to be the same or nearly the same; double-doubles, which have two sets of lobes on each side; and many, many more. Some examples of radio galaxies with interesting structure are shown in Figure 2.4. Large-scale automated identification of these galaxies can be tricky owing to their variety, extent, and often disconnected structure.

AGN cores tend to have flat or inverted spectral indices around -0.5–1 (Condon

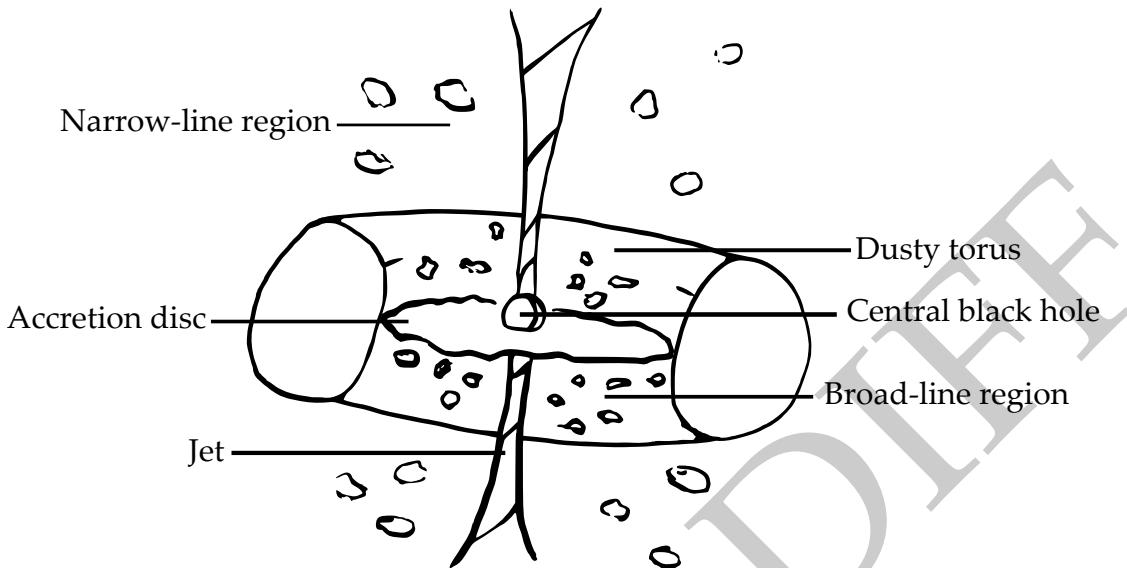


Figure 2.5: The unified model of AGN.

& Ransom, 2016; Randall et al., 2012). Moving out from the host galaxy, the spectral index steepens as the electrons are older and less energetic, with the spectral index of the lobes usually at about -0.7 (Condon & Ransom, 2016). The hot-spots of FRII galaxies have spectral indices between -0.5 – -0.7, becoming shallower as the electrons reaccelerate. These are only general trends: the spectral structure within lobes can be very complex (Treichel et al., 2001). The jets do not strongly emit and are only detectable for particularly deep observations or nearby radio galaxies.

2.3.3 The unified model

At their core, AGN are an accreting *supermassive black hole*: a body so dense that even light cannot escape its gravitational pull, with mass on the order of 10^7 – $10^9 M_{\odot}$ (Marconi & Hunt, 2003). Such black holes seem to exist at the centres of galaxies and these galaxies are called *host galaxies*. The current understanding of the structure of an AGN is as follows (Urry & Padovani, 1995). The black hole is surrounded by an accretion disc emitting in ultraviolet and X-ray. Beyond this is the broad-line region, named for the Doppler-broadened emission lines emitted by the energetic clouds of material surrounding the accretion disc. The broad-line region and accretion disc are themselves surrounded by a dusty torus (or some other disc-like structure) which prevents light from the centre of the AGN being observed from the sides. Further still from the accretion disc is the narrow-line region, where lower-energy gas produces narrow emission lines. From either side of the disc, an AGN produces two collimated outflows of relativistic plasma called jets, and these jets may interact with gas in the host galaxy to produce bright radio emission. The jets are not always visible. As the jets disperse further out from the centre of the AGN they widen into plumes of plasma known as *lobes*. This model of AGN unifies different observed classes of AGN by their orientation and

915 luminosity, and is hence known as the *unified model* (Antonucci, 1993). Recent work
 916 suggests that the unified model of AGN is not the full story (e.g. Zhuang & Ho, 2020).

There are many different ways to divide the set of radio AGN into classes. By morphology, radio AGN are often divided by the structure of the jets and lobes, with FRI and FRII the most striking examples. AGN can also be divided into *radiative-mode* and *jet-mode* by how they expel their energy (Heckman & Best, 2014). Radiative-mode AGN produce radiative energy in amounts higher than 1 per cent of their Eddington limit, while jet-mode AGN mainly output energy through their jets. The Eddington limit describes the maximum luminosity that a compact object can emit, and is given in Equation 2.20 (Rybicki & Lightman, 2008):

$$L_{\text{Eddington}}(M) = \frac{4\pi GMm_p c}{\sigma_T} \quad (2.20)$$

917 where M is the mass of the compact object.

918 Optical emission observed near the centre of the AGN can be used to divide radio
 919 AGN into broad-line and narrow-line galaxies. The former have broad spectral lines
 920 while the latter have narrow spectral lines, with broader spectral lines indicative of
 921 higher thermal energies. The most common interpretation, under the unified model,
 922 is that broad-line AGN are those seen end-on and narrow-line are those seen edge-on
 923 with the dusty torus obscuring the broad-line region. These narrow-line galaxies are
 924 usually the only ones for which we see significant extended structure.

925 2.3.4 Polarised structure

926 The magnetic field of AGN is thought to be critical to their structure (Sikora & Begelman, 2013). A strong magnetic field is required to eject and collimate the jets (Lovelace,
 927 1976) and the magnetic environment influences the structure of the jets (O'Sullivan et
 928 al., 2015). Polarisation provides a probe for measuring this magnetic field. Radiative-
 929 and jet-mode AGN have different fractional polarisations, with jet-mode AGN hav-
 930 ing a much wider range of fractional polarisations ($p \sim [0, 30]$ per cent) compared to
 931 radiative-mode AGN (limited to $p \lesssim 15$ per cent), with this difference attributable to
 932 the magnetic environment (O'Sullivan et al., 2015). Steep-spectrum ($\alpha > 0.5$) and flat-
 933 spectrum ($\alpha < 0.5$) AGN have differing fractional polarisations, with steep-spectrum
 934 sources having much higher fractional polarisation for frequencies > 5 GHz and flat-
 935 spectrum sources having higher fractional polarisation for frequencies < 1 GHz due
 936 to frequency-dependent depolarisation of the steep-spectrum sources (Saikia & Salter,
 937 1988). Hotspots of FRII radio galaxies have low polarisation (< 10 per cent) while
 938 the more diffuse sections may have much greater polarisation (> 20 per cent) (Saikia
 939 & Salter, 1988). The direction of the magnetic field is correlated with the direction of
 940 patterns in the total intensity of the source (Saikia & Salter, 1988).

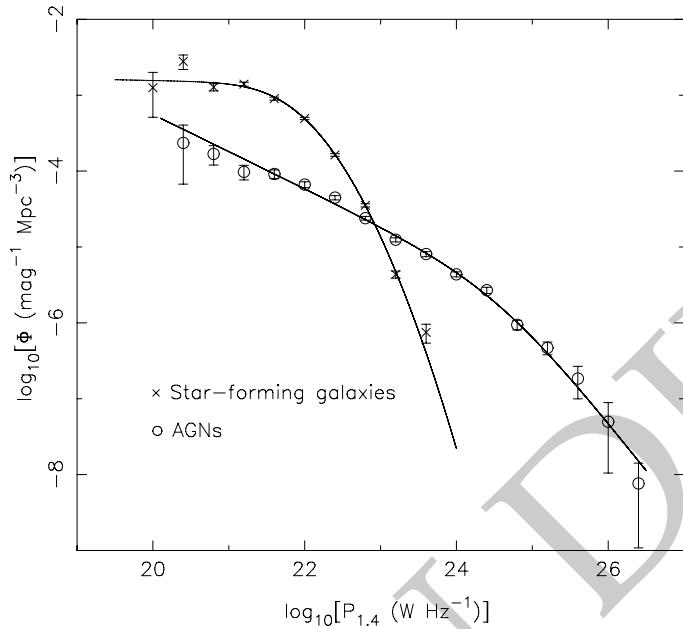


Figure 2.6: RLF for star-forming galaxies and AGN, from Mauch and Sadler (2007).

2.3.5 AGN luminosity

The distribution of AGN luminosities throughout the Universe can be characterised by the *radio luminosity function* (RLF). This describes the density of AGN within luminosity bins at a given frequency. The RLF at 1.4 GHz from Mauch and Sadler (2007) is shown in Figure 2.6. 1.4 GHz is arguably the most common large-scale survey frequency as it detects the easily observed 21 cm hydrogen line, but this RLF can be scaled to different frequencies by assuming a constant spectral index ($\alpha = -0.7$ for a typical AGN). In Chapter 5, we will use a machine learning-derived radio source catalogue to estimate a RLF for extended radio objects.

2.3.6 The role of AGN

AGN are an important part of galaxy evolution, though the exact mechanisms of this are not currently known. Understanding galaxy evolution requires understanding the links between it and AGN, and thus requires the study of AGN.

There are established relationships between AGN parameters and galaxy parameters: the mass of the central black hole is correlated with the stellar mass and velocity dispersion near the core in elliptical galaxies (Cattaneo et al., 2009) as well as the luminosity of the bulge (Kormendy & Richstone, 1995). Massive galaxies seem to be more likely to contain an AGN rather than an inactive galactic nucleus (Hardcastle & Croston, 2020). The large-scale behaviour of AGN (e.g. the Fanaroff-Riley divide) may be caused by interactions within the host galaxy (Hardcastle & Croston, 2020) and this interaction would be a significant energy contribution to the [interstellar medium ISM](#).

963 Perhaps the biggest topic in this field is that of *AGN feedback*, the role of AGN in the
964 enhancement or quenching of star formation within their host galaxies. Star formation
965 requires cold gas; AGN can push gas out of the galaxy (Zubovas & King, 2012) or heat
966 it so that it can no longer form stars. But the rapid expansion of AGN jets can produce
967 shocks in the local environment, compressing gas, triggering collapse of gas clouds,
968 and potentially increasing the star formation rate (Zubovas et al., 2014).

969 For a review of the role of AGN within galaxy evolution, see Kormendy and Ho
970 (2013). For a detailed review of the current understanding of AGN, see Hardcastle
971 and Croston (2020).

972 2.4 Classifying AGN

973 As discussed in Section 2.3.3, radio galaxies fall into many classes. Understanding the
974 mechanisms underlying these class distinctions is critical to understanding AGN. As
975 we have no way to directly see the core of an AGN (it's far too small to resolve at the
976 distances AGN occur and may also be occluded), our only method to investigate AGN
977 is to look at their large-scale behaviour. Some classes may relate to the fundamental
978 AGN core, some may be environmental, and some may be due to observation effects.
979 Much of our knowledge about AGN (such as the unified model) come from analysing
980 these classes and their differences. To investigate classes of AGN a large sample of
981 each class is required, and source classification approaches can divide a large dataset
982 from a radio survey into useful subsets. Knowing what class a source is may also help
983 analyse its properties as we can estimate its expected behaviour, perhaps with the aid
984 of models and simulations. Some classes may have interesting structure or properties
985 that can only be observed with additional detailed observations, so identifying which
986 sources require follow-up is a tightly related problem in radio astronomy. An excellent,
987 though now somewhat dated, summary of radio source classification is the review
988 paper by Urry and Padovani (1995), which we recommend for further reading.

989 Deciding which class a given radio galaxy falls into may be challenging, and doing
990 this automatically even more so. This section discusses approaches to classifying radio
991 galaxies.

992 2.4.1 Statistical and manual classification of AGN

993 Manual and statistical approaches to classifying AGN have dominated the radio as-
994 tronomy literature until very recently, due to the comparative lack of computational
995 power as well as a lack of good automated methods. Manual methods amount to ex-
996 amining the structure of a resolved source and determining its class: this is how we
997 usually identify bent radio galaxies, head-tail radio galaxies, X-shaped radio galax-
998 ies, and those radio galaxies with more unusual morphologies. Statistical approaches
999 identify properties of the source that can be combined and thresholded to separate
1000 the sources into categories en masse. Modern machine learning techniques for classi-
1001 fication of radio sources can be thought of as an extension of these statistical methods,

1002 where the properties and their combinations are identified automatically, but we will
1003 discuss these separately in Section 2.4.2.

1004 Arguably the most well-known radio classification scheme, FRI and FRII, was orig-
1005 inally defined on well-resolved radio galaxies by computing the ratio of the distance
1006 between the regions of highest brightness on opposite lobes and the total extent of the
1007 radio emission (Fanaroff & Riley, 1974). Sources with a ratio under 0.5 were called FRI
1008 and those with a ratio greater than 0.5 were called FRII. This classification has over time
1009 evolved into a less precise divide, with classification generally now morphological and
1010 based on the structure (diffuse, wavy plumes versus hot-spots and lobes for FRI and
1011 FRII respectively). The FRI and FRII divide has been further complicated by other re-
1012 lated categorisations such as the so-called “Fanaroff-Riley type 0” sources which seem
1013 to be the lower end of a continuum of radio sources with diffuse plumes (Capetti et al.,
1014 2020; Garofalo & Singh, 2019) as well as hybrid morphology radio sources (HyMoRS)
1015 which appear to be FRI on one side and FRII on the other (Gopal-Krishna & Wiita, 2000;
1016 Kapińska et al., 2017). Many classes are defined by explicitly statistical means; for ex-
1017 ample, steep- and flat-spectrum sources are divided by spectral index at $\alpha = 0.5$ (Urry
1018 & Padovani, 1995). For convenient analysis, radio sources are often also grouped into
1019 “observational” classes that don’t have a physical analogue based on their apparent
1020 structure, e.g. the GLEAM survey classifies radio sources into the number of apparent
1021 components, which is highly dependent on the observational parameters (White et al.,
1022 2020).

1023 More unusual or more loosely defined classes, such as X-shaped radio galaxies
1024 and giants, have often been identified by manual searches through large datasets, e.g.
1025 Cheung (2007), Dabhade et al. (2020) and notably the recent ROGUE I catalogue of
1026 32 616 morphologically classified radio galaxies (Żywucka et al., 2020). These searches
1027 are often aided by computer algorithms (e.g. Dabhade et al., 2020; Proctor, 2011).

1028 Radio sources are also more generally classified, such as into AGN or non-AGN
1029 emission (Kozieł-Wierzbowska et al., 2020), often using optical emission lines or opti-
1030 cal/infrared magnitude.

1031 2.4.2 Machine learning classification of AGN

1032 Machine learning based approaches for radio source classification are rapidly evolv-
1033 ing as the amount of radio data available through big surveys increases. Advances
1034 in tooling, such as the wide availability of hardware-accelerated automatic differentia-
1035 tion software, have also contributed to an explosion in machine learning applications
1036 in astronomy by making machine learning techniques more available to astronomy
1037 researchers.

1038 Morphological classification of galaxies with machine learning began in optical as-
1039 tronomy, probably due to the large sample sizes of well-resolved galaxies previously
1040 available. The earliest such paper is likely the application of neural networks to the
1041 task by Storrie-Lombardi et al. in 1992. From here, the field applied other classifica-
1042 tion algorithms such as decision trees (e.g. Owens et al., 1996). The Sloan Digital Sky
1043 Survey (SDSS) brought an explosion of new data in 2003, and new experiments in clas-

1044 sification soon followed (e.g. Ball et al., 2006; Ball et al., 2004). The Galaxy Zoo project
1045 leveraged hundreds of thousands of volunteers to produce an astonishingly large set
1046 of labelled optical galaxies from SDSS and subsequent papers used this as a training
1047 set for machine learning methods (Banerji et al., 2010; Dieleman et al., 2015; Zhu et al.,
1048 2019).

1049 While machine learning has been used in radio astronomy for some time (~~Condon et al., 1998, e.g. the NVSS~~
1050 ~~(e.g. the NVSS used neural networks to detect sidelobes; Condon et al., 1998)~~ its first ap-
1051 plication to radio source classification was most likely to identifying quasar candi-
1052 dates (Carballo et al., 2004) in Faint Images of the Radio Sky at Twenty Centimeters
1053 (FIRST; Becker et al., 1995). Proctor (2006) applied decision tree ensembles to identify
1054 bent double morphologies in FIRST, manually selecting features to characterise radio
1055 sources, while Bastien et al. (2017) used shapelet analysis to obtain features to feed into
1056 their decision tree ensembles. 2011–12 marked a revolution in computer vision with
1057 the discovery that deep convolutional neural networks (known as early as 1989, see
1058 LeCun et al., 1989), boosted dramatically by widely available training data generated
1059 by the internet and a huge increase in computational power from GPUs, could achieve
1060 greater-than-human performance on image classification tasks. Deep neural networks
1061 have since found use for morphological classification of radio sources, such as FRI vs.
1062 FRII (Aniyan & Thorat, 2017; Bowles et al., 2020; Lukic et al., 2019; Ma et al., 2019b;
1063 Samudre et al., 2020; Tang et al., 2019; see also Ma et al., 2018), compact vs. extended
1064 sources (Alhassan et al., 2018; Lukic et al., 2018; Lukic et al., 2019), and observational
1065 classes (Galvin et al., 2019; Ralph et al., 2019).

1066 There are also many works on classification of radio sources besides morphology.
1067 Machine learning has been applied to AGN classification tasks including blazar classi-
1068 fication (Arsioli & Dedin, 2020) and radio loudness (Beaklini et al., 2020). Deep learn-
1069 ing is also prevalent on this topic, with deep learning finding applications in Faraday
1070 complexity classification (Brown et al., 2018) and notably in transient detection (Agar-
1071 wal et al., 2020; Balakrishnan et al., 2020; Connor & van Leeuwen, 2018; Guo et al.,
1072 2019; Lin et al., 2020; Wang et al., 2019; Zhang et al., 2020).

1073 It is worth contrasting these machine learning approaches with non-machine learn-
1074 ing automated approaches, as the two are often conflated in the literature. Mingo et
1075 al. (2019), for example, use an automated version of detecting the brightness gradi-
1076 ent of extended radio sources to determine whether they are FRI or FRII en masse
1077 and apply this approach to the LoTSS survey. Segal et al. (2019) apply an information
1078 theoretic approach to estimating morphological complexity of a source. The key dif-
1079 ference between a machine learning automated approach and a non-machine learning
1080 automated approach is that the former has the capacity to change its behaviour based
1081 on available data, while the latter does not—though note that this is not necessarily a
1082 bad thing.

1083 2.5 Cross-identification

1084 *Cross-identification* is the problem of matching an observed astronomical object at one
1085 wavelength with its counterpart in a survey at some other wavelength. Sometimes the
1086 counterpart in question can be obvious, such as when the object being matched is com-
1087 pact and well-aligned in both wavelengths. This is not the case for radio, though: radio
1088 galaxies can be very extended and, due to environmental interactions, this extended
1089 structure may be complicated and have no obvious relationship to the galaxy at other
1090 wavelengths. The galaxy itself in radio cross-identification is called the *host galaxy* as
1091 it hosts the central AGN.

1092 2.5.1 Why do we need to cross-identify?

1093 Radio spectra do not generally contain much information besides the spectral index: as
1094 Jim Condon purportedly said, "There's nothing as useless as a radio source". Most in-
1095 formation about extragalactic sources is gleaned from other wavelengths, with optical
1096 and infrared showing physically meaningful differences due to emission and absorp-
1097 tion at these wavelengths. For this reason radio sources are typically cross-identified
1098 with their counterparts at other wavelengths to get information. Key features of a host
1099 galaxy that may be obtained from optical or infrared observations include its redshift,
1100 mass, star formation rate, and composition. From these we may also determine prop-
1101 erties of the radio plasma, most notably its physical extent and radio luminosity, both
1102 of which require knowing the redshift.

1103 2.5.2 Methods for cross-identification

1104 Most small radio surveys are cross-identified by astronomers, visually comparing the
1105 radio image to the optical or infrared image (e.g. Middelberg et al., 2008; Norris et
1106 al., 2006). The largest such catalogue is the ROGUE I catalogue with over 32 000
1107 cross-identifications (Żywucka et al., 2020). The Radio Galaxy Zoo project (Banfield
1108 et al., 2015) extended manual cross-identification to larger scales by crowdsourcing,
1109 asking volunteers online to cross-identify FIRST and ATLAS resulting in 75 000 cross-
1110 identified extended sources. This is the largest manually cross-identified source cat-
1111 alogue ever, by citizen scientists rather than professional astronomers—though this
1112 distinction does not seem to matter as the accuracy of cross-identifications in Radio
1113 Galaxy Zoo is still quite high, and there is an upper limit to how well even astronomers
1114 can cross-identify large surveys (Banfield et al., 2015; Wong et al., in prep.).

1115 2.5.2.1 Positional matching

1116 The simplest automated method for cross-identification is simply to look for the closest
1117 optical or infrared galaxy to some radio emission, a strategy called *positional matching* or
1118 *nearest neighbours*⁴. The distance to the nearest potential host galaxy is usually limited

⁴Not to be confused with the machine learning algorithm also called 'nearest neighbours', which does not make an appearance in this thesis.

to some small value, so that distant spurious relations are excluded. This distance is usually about $5''$ for infrared and $1''$ for optical but may be much further. Some authors modify this approach to selecting the nearest *bright* galaxy or some other constraints (e.g. Kimball & Ivezić, 2008). This approach generally works quite well for compact objects but extended radio structures may overlap with unrelated galaxies on the sky, leading to spurious cross-identifications.

The rate of spurious identification with this approach can be estimated by a simple model, assuming that the true host galaxy is not visible (either too faint or not nearby). Assume that a small circular region of the sky with radius a contains K potential host galaxies. The probability that a randomly selected potential host galaxy is within r rad of a given point is r^2/a^2 , with $r < a$. The probability of no potential host galaxy falling within r rad of a given point is therefore $(1 - r^2/a^2)^K$ and the probability of coincidental cross-identification is $1 - (1 - r^2/a^2)^K$. The AllWISE survey contains 747 634 026 infrared sources, over 4π sr, so it has an average source density of approximately $6 \times 10^7 \text{ sr}^{-1}$. With $r = 5''$, and $a = 1^\circ$ (corresponding to an area of 9.57×10^{-4} sr and thus 57 420 sources), we have a coincidental cross-identification rate of 10.5 per cent when the host galaxy is not visible.

2.5.2.2 Other automated methods

With very few non-trivial cross-identification methods in existence, the field is still very much in its infancy. Other automated methods for cross-identification can be grouped into machine learning and non-machine learning approaches. The two non-machine learning approaches are a Bayesian geometric model fitting approach (Fan et al., 2020; Fan et al., 2015) and a likelihood ratio approach (Weston et al., 2018; Weston, 2020). In Chapter 4 we introduce the very first machine learning approach for radio-infrared cross-identification (Alger et al., 2018). Galvin et al. (2020) later developed a hybrid machine learning and manual labelling approach to cross-identification.

2.6 Aggregating Radio Components

Unlike emission from galaxies observed at other wavelengths, the extended emission from radio galaxies can be disconnected when observed: a single radio galaxy may appear in observations as multiple discrete components. This is partly due to inhomogeneous emission over the radio galaxy structure—e.g. FRII hot-spots can be much brighter than the rest of the galaxy, so they might be visible while the rest of the galaxy is too faint to see—and partly due to the technique through which many radio observations are made, *interferometry*, which may screen out diffuse emission.

This potential of a radio galaxy to be split into multiple discrete blobs of emission also leads to a linguistic ambiguity not present at other wavelengths. A *radio source* may be either a single physical object that emits radio, or a single blob of disconnected radio emission. The same is true for *radio object*. There is no agreed-upon terminology. We will therefore adopt the following terminology from this point on, adapted from our paper (Alger et al., 2018) based on the Radio Galaxy Zoo terminology: a *radio source*



(a) The 64m telescope (Murriyang) at Parkes Observatory

(b) ATCA near Narrabri

Figure 2.7: (a) A single-dish telescope and (b) an array. Images: CSIRO.

or *radio galaxy* refers to all radio emission observed that comes from a single physical object, and a *radio component* refers to a single, contiguous region of radio emission above the noise level of an observation. We will avoid using *radio object*. Note that in the literature a *radio component* may be either a Gaussian fit to an observation or it may be a region of connected emission; in works where the former is the case (particularly in source-finding literature) the latter is often called a *radio island*.

Since radio galaxies can appear disconnected, aggregating observed radio components into physical sources is integral to understanding radio galaxies. Aggregating disparate components into a single source is important for two reasons. Firstly, without all components, the total luminosity of a source is dramatically underestimated. Secondly, we need all components to accurately examine the morphology of the source.

2.6.1 Missing emission in radio observations

Radio observations may be made either with *single-dish telescopes*, like the famous Parkes Radio Telescope (Murriyang), or *radio arrays*, like the Australia Telescope Compact Array (ATCA), both shown in Figure 2.7. Both have their advantages. Single-dish telescopes are able to measure absolute brightnesses (while arrays can only measure relative brightnesses, and must therefore be calibrated to a source of known brightness). Interferometric arrays can achieve incredibly high resolution, as the resolution is inversely proportional to the distance between the most distant array elements (while the resolution of single dish telescopes is inversely proportional to the diameter of the dish).

Radio telescopes can be thought of as sampling the *u-v plane*, the Fourier transform of the sky. The *u-v* plane is perpendicular to the line-of-sight. Each pair of antennae in an array samples two points on this plane, each corresponding to the vector between the antennae projected onto the *u-v* plane, called a *baseline*. Longer baselines therefore correspond to higher (spatial) frequencies, which is why long baselines provide

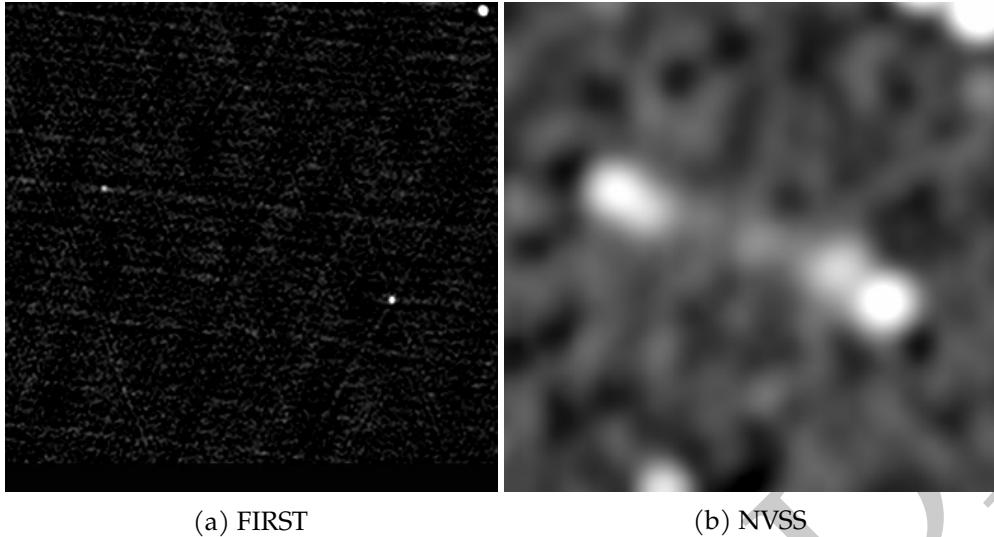


Figure 2.8: A fairly diffuse FRII, J0016+0420, observed with the VLA in the (a) FIRST (Becker et al., 1995) and (b) NVSS (Condon et al., 1998) surveys. (GRG1 from Dabhade et al., 2017)

high resolution. Diffuse emission is characterised mainly by low (spatial) frequency components, while compact emission is characterised by a broad range of frequency components, *so*. This means that large angular scale, diffuse emission both a) takes up less space on the u - v plane than compact sources and b) occupies spaces much closer to the origin on the u - v plane. Some intuition on this can be obtained by examining the Fourier transform of a 2D Gaussian:

$$\mathcal{F}_{x,y} \left[\frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \right] = e^{-2\sigma^2\pi^2(u^2+v^2)}. \quad (2.21)$$

From this equation we can see that the Fourier transform of a fairly *compact* Gaussian (*small σ*) would be quite broad, taking up many frequencies in the u - v plane, while a very *diffuse* Gaussian (*large σ*) would have a very narrow Fourier transform. The upshot of this is that long baselines sacrifice sensitivity to diffuse emission for high resolution. Single-dish radio telescopes are unable to make the same tradeoff, as they are only able to sample a disc centred on the origin⁵. This loss of larger scale diffuse emission is often called *resolving out*. An example of this is shown in Figure 2.8, where (a) and (b) are the same radio galaxy observed with the same telescope, the Very Large Array, with the only difference being that the VLA was in B configuration for (a) and the D configuration for (b). The B configuration moves the antennae of the VLA far apart, while the D configuration keeps them close together.

⁵This is, incidentally, why single-dish telescopes can measure the absolute brightness while arrays cannot: there is no way to measure the origin in the u - v plane as there is no way for two array antennae to be infinitely close together (forming a zero-length baseline), and the origin contains the absolute brightness information, much like how the centre of a Fourier transform contains the mean.

1192 **2.6.2 Methods of aggregation**

1193 Most large radio catalogues are of *components* rather than sources, most likely because
1194 manual methods remain the best way to aggregate components into sources. Many
1195 component catalogues have source catalogues of some interesting subset later derived
1196 from them by manual inspection (e.g. Dabhadé et al., 2017). Such manual searches
1197 are often assisted by automated methods (e.g. Proctor, 2011). The Radio Galaxy Zoo
1198 citizen science project has also, with the help of over 10 000 volunteers, aggregated over
1199 75 000 sources from FIRST (Wong et al. in prep) and ATLAS (Banfield et al., 2015),
1200 which is one of the largest (if not the largest) extant catalogues of manually aggregated
1201 sources.

1202 Automated methods have been developing steadily as data volumes have grown. A
1203 Bayesian approach, fitting models of sources to component locations, was introduced
1204 by Fan et al. (2015) and later expanded upon (Fan et al., 2020). Machine learning
1205 methods have also become increasingly popular. Proctor (2016) aggregated radio com-
1206 ponents using a decision tree classifier and used the results to search for giant radio
1207 galaxies (see also Proctor, 2006, 2011). Wu et al. (2019) applied a deep learning model
1208 to identify radio sources from an image of radio components, training this model on
1209 the Radio Galaxy Zoo FIRST dataset. Most recently Galvin et al. (2020) used a nested
1210 self-organising map, a kind of neural network, to cluster similar sources and then ag-
1211 gregated them into sources using manual labelling of the clusters and some heuristics.

1212 Any method for cross-identification of components can also be used to aggregate
1213 components. Cross-identify all components with their host galaxies, then components
1214 that share a host are almost always part of the same radio source. It is this approach
1215 we will later take to aggregating sources in Chapter 5. There is no obvious inverse
1216 algorithm (to go from aggregate sources to cross-identifications) although some al-
1217 gorithms attempt to solve both problems simultaneously (Fan et al., 2020; Fan et al.,
1218 2015).

1219 **2.7 Summary: radio sources**

1220 The radio sky is vibrant and exciting, and current and new wide-area radio surveys are
1221 opening up new avenues of exploring this space. In this chapter we introduced radio
1222 sources that you might see in the radio sky, including the often-extended AGN. We dis-
1223 cussed their radio emission and emission mechanisms, and described some aspects of
1224 their morphology and structure. We also looked at ways that radio sources can be clas-
1225 sified, aggregated, and cross-identified with their counterparts at other wavelengths.
1226 In the next chapter, we will introduce machine learning concepts that we will use to
1227 develop methods for exploiting the vast amount of data radio telescopes provide on
1228 radio sources.

REVISION DIFF

1230 **Machine Learning for**
 1231 **Astroinformatics**

1232 Machine learning was once described to me by an anonymous supervisor as “the statistics
 1233 kept at the back of the textbook”. But even accepting its grounding in statistics,
 1234 is this really an accurate description of the field? I think of machine learning as a
 1235 combination of three pieces: a data-driven way of formalising predictive problemsand
 1236 a way of converting between different kinds of statistical problems, as well as an
 1237 accompanyinganda set of methods and practices for handling data and uncertainty.
 1238 The eventual goal is to design some method or algorithm that automatically discovers
 1239 useful patterns in (potentially very large) data sets. There are three core components
 1240 of machine learning: the data, the model, and learning (Deisenroth et al., 2020). Before
 1241 discussing these, we will look at the kinds of problems that machine learning solves.

1242 **3.1 Prediction**

1243 Machine learning aims to solve *prediction tasks*: problems where we have some data
 1244 and we seek some kind of output based on that data. Central to prediction tasks are
 1245 predictors, the objects we train based on data.

1246 **3.1.1 Predictors**

1247 A *predictor* is an object that makes predictions based on an input. A predictor can be a
 1248 function or a probabilistic model, depending on the machine learning approach being
 1249 undertaken.

As a function, a predictor maps from some input domain \mathcal{X} into some output domain \mathcal{Y} , and is usually written as

$$f : \mathcal{X} \rightarrow \mathcal{Y}. \quad (3.1)$$

1250 \mathcal{X} and \mathcal{Y} are commonly (but certainly not always) a real vector space \mathbb{R}^n . Because the
 1251 goal of machine learning involves *finding* a suitable function f for the task at hand, the
 1252 set of functions is usually constrained. For example, if $\mathcal{X} = \mathbb{R}^n$, we might require that
 1253 f is a linear function $\mathbb{R}^n \rightarrow \mathbb{R}$, easily parametrised by $n + 1$ constants. This constraint is

1254 called a *model*. Models are often (but not always) parametrised and training the model
 1255 refers to finding a good set of parameters. The parameters are sometimes called *weights*.
 1256 To avoid ambiguity, parameters of the model that are not changed during training are
 1257 called *hyperparameters* and may include regularisation terms, constraints, or variations
 1258 of the model.

1259 Some predictors can be described as a probabilistic model. In this case a predictor is
 1260 a joint probability distribution between observations and hidden parameters (Deisen-
 1261 roth et al., 2020). Using a probabilistic predictor allows us to formally describe and
 1262 work with uncertainty both in the input space and output space. Such a predictor is
 1263 usually parametrised by a finite set of parameters, which already includes most com-
 1264 mon probability distributions.

1265 We will generally assume that our data are generated from some unobserved, true
 1266 function called the *groundtruth*. This might be a physical process, or a complicated
 1267 sampling function from some unknown vector space. The assumptions we make on
 1268 this generative function can greatly change the way we approach machine learning
 1269 problems.

1270 In some sense, the goal of machine learning is to identify a good predictor from
 1271 within the space of all possible predictors. Of course, this begs the question: what is a
 1272 ‘good’ predictor? We will return to this when we discuss learning, but for now, a good
 1273 predictor is one that approximates the groundtruth well.

1274 3.1.2 Classification

1275 *Classification* is the machine learning task of predicting discrete, unstructured values
 1276 (Deisenroth et al., 2020). These values are called *classes*. Classification is arguably the
 1277 most important prediction task, as many other problems can be formalised as classi-
 1278 fication. Astronomy has its fair share of classification tasks, from classical astronomy
 1279 tasks like galaxy morphology classification (appearing in machine learning literature
 1280 as e.g. Dieleman et al., 2015) to transient detection (e.g. Scalzo et al., 2017); see Sec-
 1281 tion 2.4 for more examples.

1282 A classification problem seeks a predictor where \mathcal{Y} represents a finite, discrete set
 1283 of classes. Classification tasks are usually delineated by the number of classes: there
 1284 are either two classes or more than two classes. The former are called *binary classifica-*
 1285 *tion* tasks and the latter are *multiclass classification* tasks. The reason for this split is that
 1286 binary classes are dramatically easier to reason about and analyse, and many special
 1287 cases exist for binary where they do not for multiclass.

1288 \mathcal{Y} for a binary task is usually represented as $\mathcal{Y} = \{0, 1\}$. 1 is called the *positive class*;
 1289 0 is called the *negative class*.

1290 An easy way to see why many tasks can be formalised as classification can be found
 1291 by taking any prediction problem $\mathcal{X} \rightarrow \mathcal{Y}$ and reinterpreting it as the binary classifica-
 1292 tion problem $\mathcal{X} \times \mathcal{Y} \rightarrow \{0, 1\}$, i.e. instead of taking an input and predicting an output,
 1293 take an input and a potential output and determine if they should be related. Of course
 1294 this is not always the most efficient way to solve a prediction problem but the many
 1295 known properties of classification make it an appealing framework to cast problems

1296 into. In Chapter 4, we will cast the radio astronomy problem of cross-matching galaxies
1297 seen in different wavelengths into a binary classification problem, and in Chapter 6
1298 we will classify radio observations as Faraday complex or Faraday simple.

1299 There are three different formalisms for the output of a binary classifier. The out-
1300 put may be a *score*, a *class probability estimate*, or a . A predicted class is
1301 what first comes to mind as a representation of binary classes: the classifier outputs
1302 either 0 or 1, with no way of representing ambiguity. A class probability estimate is a
1303 number from 0 to 1 which represents the probability that an example is in the positive
1304 class, i.e. if f is a classifier then $f(x) = p(y = 1 | x)$. Class probability estimates may
1305 be output by probabilistic classifiers, or may be estimated from non-probabilistic classi-
1306 fiers. Finally, a score is a value that tends to be larger for elements of the positive class,
1307 with less ambiguous examples being given a higher score. Scores are often converted
1308 into class probability estimates by applying a sigmoid (see Section 3.4.1). There are
1309 multiple ways to generalise these concepts to multiclass classification. Having a score
1310 or a class probability estimate can be more useful than only having a predicted class.
1311 For example, you could sort your examples by how likely they are to be a member
1312 of the positive class. This underpins our approach to casting cross-identification as
1313 binary classification in Chapter 4.

1314 3.1.3 Regression

1315 The other main kind of supervised prediction task is *regression*, which is the machine
1316 learning task of predicting ordered (and usually continuous) values. In a regression
1317 problem, we seek a predictor where \mathcal{Y} is a set of ordered values, usually a subset of \mathbb{R}^k
1318 for some positive natural k . Regression is ubiquitous in astronomy, from simple linear
1319 relationships like the famous Tully-Fisher relation (Tully & Fisher, 1977) to estimation
1320 of redshifts from photometric observations (called *photometric redshifts*; first introduced
1321 by Baum, 1962). We will not directly address any regression problems in this thesis,
1322 but we will make use of their results.

1323 3.2 Data and representation

1324 Machine learning is centred on data and the extraction of useful information from that
1325 data. Data can include anything from numeric information, documents, or images,
1326 to spectra or galaxies. A collection of data is called a *dataset* and an element of this
1327 dataset is (interchangably) called an *example* or *instance*. Generally, data are not easy to
1328 work with in their original form and must be converted into a numerical representation
1329 before use. We usually convert our data into real vectors in \mathbb{R}^n as it is relatively easy
1330 to work with both numerically and analytically. Each axis of this vector space is called
1331 a *feature* and the space as a whole is called the *feature space*. Features are non-trivial
1332 to choose, and finding good features often requires the expertise of a human who is
1333 well-versed in the original dataset (a *domain expert*). The process of finding features is
1334 called *feature selection*, *feature design*, or *feature engineering*.

What makes a feature good? Intuitively, we want to transform our data into a space where it is easy to work with: a space where properties we care about are obvious, easy to extract, behave nicely, and work well with our model. For this reason, features will vary tremendously depending on the problem being faced, and the same data may be represented in many different ways. Much of early machine learning literature focused on finding good methods to automatically develop features (generally called *feature extraction*), and much early applied machine learning focused on identifying these features manually so that good predictors could be easily found. An astronomical example is Proctor (2006), who developed features for representing radio galaxies for the purpose of sorting them. State-of-the-art models like deep neural networks (e.g. Dieleman et al., 2015) can be viewed as developing their own task-specific features as part of their training (see Section 3.4.3). These deep learning features can be useful (e.g. when it is hard or slow to define good features manually), but are usually hard to interpret. A good feature space will have a structure that reflects the components of the intrinsic structure of the input data which are useful for the prediction task at hand. Good features may also be useful in other related tasks, such as dataset exploration, dataset visualisation, or other prediction tasks. Chapter 6 largely focuses on finding good features for identifying Faraday complexity in polarised sources.

Labels comprise another very important piece of the machine learning puzzle. Training a predictor with supervised learning requires some known pairs of inputs and outputs, and the known outputs are called labels. Like features, labels also need to be encoded in some way, and this depends on the specific task. Much like features, we want to embed the labels into a space which is easy to work with and has a meaningful structure. For problems where we know the outputs we wish to obtain, this can be a lot simpler than feature selection. For example, a binary classification problem will have only two possible outputs. Common representations for this would be $\{0, 1\}$ as described in Section 3.1.2, but we could also represent the labels as $\{[1, 0]^T, [0, 1]^T\}$, called a *one-hot encoding*. The advantage of the former is its simplicity and ease of integration into binary classification equations, but the advantage of the latter is that it is easily extended into multiclass classification without imposing order on the classes. Despite being simpler to encode, labels can carry a lot more difficulty than features due to their comparative rarity: in essence, features are cheap and labels are expensive. We will discuss labels in more detail in Section 3.5.

3.3 Loss functions

Training a model is the process of using data to find a good predictor that fits the model's constraints. This is generally achieved by minimising a *loss* (also called *error* or *cost*) function over the model.

Put simply, a loss function is a function of a predictor and a dataset which is chosen to be a proxy for how good the predictor is at predicting that dataset. We try to choose loss functions that are high-valued for a predictor that poorly describes the dataset, and are low-valued for a predictor that well-describes the dataset. Sometimes (and

1376 in both cases listed in this section) the loss is minimised at zero, when the predictor
 1377 perfectly captures the dataset (though whether this is possible, or whether this is even
 1378 a desired result, is another question).

1379 What should the loss function be for a given problem? The answer is not always
 1380 obvious. Take for example a binary classification problem. The “obvious” loss would
 1381 be the complement of the accuracy: the rate at which the predictor incorrectly guesses
 1382 the label. This is easy to compute and we would like our predictor to have a high
 1383 accuracy. But this is not a good choice: it is tremendously hard to work with as it takes
 1384 on discrete values, because the accuracy is the number of correct predictions divided
 1385 by the total number of examples. It is hard to motivate with probabilistic arguments.
 1386 Finally, it is unclear how the accuracy should work in the case of a probabilistic model.

1387 Instead, the loss function is usually derived by making assumptions on the struc-
 1388 ture of the data and task. The main assumption we usually make is that data are drawn
 1389 *independently and identically distributed* (IID), that is, each example is drawn from the
 1390 same distribution and is not dependent on any other examples. We also assume a
 1391 structure of the noise in the observed labels: training data are almost never completely
 1392 accurate, and so there will be intrinsic noise in the distribution of labels about their
 1393 unobserved “true” value. To demonstrate these assumptions, we will now derive loss
 1394 functions for regression and binary classification. The loss function we derive depends
 1395 on the model we assume for the noise; for example, assuming Gaussian noise gives us
 1396 the common least-squares loss.

1397 3.3.1 Loss function for regression

To derive a loss function for regression, let us assume that our labels are a random variable y modelled by a predictor $y = f(x)$. Further, let us assume that a predicted y is normally distributed about its true value, i.e.

$$y \sim \mathcal{N}(\mu, \sigma^2) \quad (3.2)$$

for the true mean μ and standard deviation σ where \mathcal{N} is the normal distribution:

$$\mathcal{N}(a | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(a-\mu)^2}{2\sigma^2}}. \quad (3.3)$$

Under this assumption the probability that y is equal to a target t given an example x is

$$p(y = t | x) = \mathcal{N}(t | f(x), \sigma^2). \quad (3.4)$$

What would the probability be of observing a set of targets $T = \{t_1, \dots, t_n\}$ given corresponding examples $X = \{x_1, \dots, x_n\}$? Letting $Y = \{y_1, \dots, y_n\}$ be random variables like y , the joint probability distribution $p(Y = T | X)$ is

$$p(Y = T | X) = \prod_{i=1}^n p(y_i = t_i | x_i) \quad (3.5)$$

by using our independence assumption. We can then substitute the normal distribution:

$$p(Y = T | X) = \prod_{i=1}^n \mathcal{N}(t_i | f(x_i), \sigma^2). \quad (3.6)$$

$p(Y | X)$ is called the *likelihood*. We would like to maximise this likelihood over f , which is called a *maximum likelihood* approach to finding a predictor. It is, however, not very easy to work with in this current form. Maximising the likelihood is equivalent to minimising its negative logarithm, so we write:

$$\mathcal{L}(f; T, X) = - \sum_{i=1}^n \log \mathcal{N}(t_i | f(x_i), \sigma^2) \quad (3.7)$$

where \mathcal{L} is the *negative log-likelihood*, a loss function. We can simplify this dramatically by cancelling the logarithm and the exponential within the normal distribution:

$$\mathcal{L}(f; T, X) = \sum_{i=1}^n \frac{(t_i - f(x_i))^2}{2\sigma^2} \quad (3.8)$$

and by noting that arbitrary scaling of \mathcal{L} does not change the minimising f we can scale \mathcal{L} by σ^2 and arrive at the *sum-of-squares error*, also known as the *least-squares error*, the most common loss function for regression:

$$\mathcal{L}(f; T, X) = \frac{1}{2} \sum_{i=1}^n (t_i - f(x_i))^2. \quad (3.9)$$

The factor of half helps keep the derivative tidy:

$$\frac{d\mathcal{L}}{d\theta}(f; T, X) = \sum_{i=1}^n (t_i - f(x_i)) \frac{df}{d\theta}(x_i). \quad (3.10)$$

¹³⁹⁸ 3.3.2 Loss function for binary classification

Now we will calculate a loss function for binary classification. As for regression, we first assume a form for the noise. Assume that our labels are a random variable $y \in \{0, 1\}$ and that the prediction y is drawn from a Bernoulli distribution based on a predictor $f(x)$:

$$p(y = t | x) = \mathcal{B}(t; f(x)). \quad (3.11)$$

The Bernoulli distribution is parametrised by one parameter, usually called $p \in (0, 1)$, and in this case set to $f(x)$. It is:

$$\mathcal{B}(a; p) = p^a (1 - p)^{1-a}. \quad (3.12)$$

It can be thought of as a biased coin toss with a probability p of tossing heads. To gain some intuition into how this expression works, imagine setting a to 0 and then to 1. Continuing to derive the loss function, we once again determine the likelihood

making the IID assumption:

$$p(Y = T \mid X) = \prod_{i=1}^n p(y_i = t_i \mid x_i) = \prod_{i=1}^n f(x_i)^{t_i} (1 - f(x_i))^{1-t_i}. \quad (3.13)$$

Then we find the negative log-likelihood and hence what is known as the *binary cross-entropy loss* for binary classification:

$$\mathcal{L}(f; T, X) = - \sum_{i=1}^n \log \left(f(x_i)^{t_i} (1 - f(x_i))^{1-t_i} \right) \quad (3.14)$$

$$= - \sum_{i=1}^n t_i \log f(x_i) + (1 - t_i) \log(1 - f(x_i)). \quad (3.15)$$

1399 This is the standard loss used for logistic regression (Section 3.4.1).

1400 3.3.3 Gradient descent

Given a loss function and a parametrised model, how can we find parameters for the model that minimise the loss function? There are many optimisation strategies but if both the loss function and model are differentiable with respect to the parameters then we can employ a particularly efficient approach: *gradient descent*. Assume we have a model $f(x; \vec{w})$ parametrised by some vector \vec{w} and a loss function $\mathcal{L}(\vec{w}; T, X)$. Then the value of \vec{w} after the $k+1$ th update of gradient descent is

$$\vec{w}^{(k+1)} = \vec{w}^{(k)} - \lambda \nabla_{\vec{w}} \mathcal{L}(\vec{w}^{(k)}; T, X) \quad (3.16)$$

1401 where $\lambda > 0$ is a small scalar called the *learning rate*. With an appropriately small choice
1402 of λ \vec{w} will converge to a local minimum of \mathcal{L} . Many variations on this concept exist
1403 which attempt to avoid local minima, such as introducing a ‘momentum’ term that
1404 accumulates as multiple iterations move \vec{w} in the same direction. If the loss function
1405 is convex, then any minimum is the global minimum (there are no local minima).

1406 3.4 Models

1407 This section describes some common models for classification. There are a plethora of
1408 different classification models and variations on these models, but I will present here
1409 only those relevant to this thesis: logistic regression, decision tree ensembles, and neu-
1410 ral networks. These are, not coincidentally, also the most common models in astroinfor-
1411 matics. Logistic regression provides reliable and interpretable results. Decision tree
1412 ensembles are a fantastic off-the-shelf choice which work on a large variety of datasets.
1413 Neural networks have proved extremely effective for a wide variety of tasks, especially
1414 in computer vision.

1415 **3.4.1 Logistic regression**

1416 *Logistic regression* is a linear, binary, probabilistic classifier. Linear classifiers can only
 1417 separate classes using a hyperplane in the feature space, with objects on one side of the
 1418 plane being assigned to one class and objects on the other side being assigned to the
 1419 other. A binary classifier works on binary classification tasks. Probabilistic classifiers,
 1420 as discussed in Section 3.1.2, have outputs interpretable as class probabilities.

Logistic regression in a d -dimensional feature space is parametrised by a *weights* vector $w \in \mathbb{R}^d$. Given a set of features $x \in \mathbb{R}^d$, logistic regression is:

$$f(x; w) = \sigma(w^T x) \quad (3.17)$$

where σ is the *logistic function* or *sigmoid*, which is a monotonic and bijective function:

$$\sigma(a) = \frac{1}{1 + e^{-a}}. \quad (3.18)$$

1421 The output of logistic regression applied to an example x is the probability that x is
 1422 in the positive class. σ , and thus logistic regression, has a domain of $(-\infty, \infty)$ and a
 1423 range of $(0, 1)$. This enforces the output to be like a probability. $w^T x = 0$ defines a
 1424 d -dimensional hyperplane, called the *separating hyperplane* or *decision surface*. Logistic
 1425 regression is differentiable, which allows us to optimise its parameters w using
 1426 gradient descent. Interpreting the classifier is possible through examining the weights
 1427 vector, with a larger absolute value of a weight corresponding to a ‘more important’
 1428 feature. The weights from logistic regression can be used to produce scores or class
 1429 probability estimates (Section 3.1.2): $w^T x$ is a score, $\sigma(x^T x)$ is a class probability esti-
 1430 mate, and thresholding either the score or the class probability estimate at 0 gives a
 1431 discrete class.

1432 A limitation of logistic regression is its sensitivity to scale. Features need to be of
 1433 approximately the same order of magnitude and should have a standard deviation of
 1434 approximately 1. An implicit assumption is that each features has a mean of 0 across
 1435 the dataset. This can be enforced by normalising and scaling: subtract the mean of the
 1436 dataset and divide by the new standard deviation.

1437 We apply logistic regression to binary classification problems in Chapter 4 and
 1438 Chapter 6.

1439 **3.4.2 Decision tree ensembles**

1440 A *decision tree* is a non-linear classifier. It repeatedly splits a dataset based on binary
 1441 comparisons until every subset contains only one class (or mostly one class, with the
 1442 amount of purity left as a hyperparameter). Each split only uses one feature for the
 1443 comparison, making decision trees relatively easy to visualise and interpret. However,
 1444 because of this, each split is axis-parallel, which can be a limitation for some datasets.
 1445 They are not sensitive to scale and do not require a zero mean, making them easy to
 1446 apply without preprocessing a dataset.

1447 Key limitations of a decision tree are:

- 1448 • They can only output a prediction, not a confidence of this prediction or a score
1449 of how likely an instance is to be found within each class.
- 1450 • Small changes to the dataset or training method can result in large changes to
1451 the tree.
- 1452 • They tend toward overfitting the training set.
- 1453 • With many low-information features, decision trees have quite poor performance
1454 (Breiman, 2001).

1455 A *decision tree ensemble* aims to reduce some of these limitations by training multiple,
1456 slightly different, independently-trained decision trees. Depending on the implemen-
1457 tation each constituent decision tree may only have access to some of the features or
1458 some of the data. To predict, each tree produces a prediction and ‘votes’ for this predic-
1459 tion; the votes can be combined to produce the overall prediction (e.g. with majority
1460 voting). A simple example of such an ensemble is decision tree bagging (Breiman,
1461 1996), which trains each tree with a random subset of the training data and takes a
1462 plurality vote. Decision tree ensembles decrease variance, increase the usability of
1463 low-information features, and increase stability of the trained model (Breiman, 2001).
1464 Averaging over the classifications produced by each tree gives a number that can be
1465 interpreted as a class probability estimate.

1466 The most well-known description of decision tree ensembles is the *random forest*
1467 (Breiman, 2001), which has found common use in astronomy partly to its readily avail-
1468 able Python implementation in `scikit-learn` (Pedregosa et al., 2011). Splits are de-
1469 cided from a subset of features and training samples are randomly drawn with replace-
1470 ment from the total training set. One downside of random forests is the large number
1471 of hyperparameters that need to be set, and these vary a lot depending on the problem
1472 being addressed.

1473 We apply decision tree ensembles to binary classification problems in Chapter 4
1474 and Chapter 6.

1475 3.4.3 Convolutional neural networks

A *neural network* is a directed graph of transformations, each node representing a trans-
formation that linearly combines its inputs and applies a non-linear function called the
activation function to the result. The inputs to the graph are the features. A particularly
prominent kind of neural network is the *fully-connected neural network*, where nodes
are arranged into layers, with each node in a layer taking as input every output from
the previous layer. Each layer can then be represented by a matrix multiplication of
the outputs of the previous layer by a weight matrix, composed with the activation
function. Fully-connected K -layer neural networks have the form:

$$f(x; W_K, \dots, W_1) = h_K(x; W_K, \dots, W_1) \quad (3.19)$$

$$h_i(x; W_i, \dots, W_1) = a(W_i h_{i-1}(x; W_{i-1}, \dots, W_1)) \quad (3.20)$$

$$h_1(x; W_1) = a(W_1 x) \quad (3.21)$$

1476 where a is the activation function. h_i are called *hidden layers*. In fact, neural networks
 1477 are usually described by their layer structure rather than graph structure, with the
 1478 addition of ‘concatenation layers’ to combine outputs from previous layers. Neural
 1479 networks may be used for regression or for classification; these are structured the same
 1480 but for classification the last activation function is replaced by sigmoid (for binary
 1481 classification) or its multiclass counterpart softmax.

1482 *Convolutional neural networks* (CNN; LeCun et al., 1998) are a variant of neural net-
 1483 works that are particularly well-suited to inputs that have local structure, such as im-
 1484 ages or spectra. Layers in the network may be *dense layers* of the same form as Equa-
 1485 tion 3.20, or *convolutional layers*, where the weights are convolved with the input rather
 1486 than multiplied. These convolutional weights are called *filters* and they are small com-
 1487 pared to the dimensionality of the input. CNNs are translation-invariant (Waibel et al.,
 1488 1989) and derive features from local relationships thanks to the trainable filters.

1489 We apply CNNs to binary classification problems in Chapter 4 and Chapter 5, and
 1490 compare our results to a state-of-the-art CNN in Chapter 6.

1491 3.5 Labels

1492 As described in Section 3.2, labels are the known outputs of supervised prediction
 1493 tasks. They are used for two main, distinct purposes: training and validation. Labels
 1494 for training are used to evaluate loss and determine how to update the model. Labels
 1495 for validation are used to evaluate and characterise the model’s behaviour.

1496 3.5.1 Where do labels come from?

1497 We previously said that labels were ‘expensive’. This is to be interpreted as expen-
 1498 sive in either or both time and money: labelling can be a slow, manual process, and
 1499 labelling can be costly. Labelling is usually completed by hand, manually examining
 1500 instances and determining what class they belong to (for classification) or what target
 1501 they ought to be associated with (for regression). In astronomy this usually amounts
 1502 to expert astronomers examining imagery at multiple wavelengths and making an edu-
 1503 cated guess as to what the true label should be, but labelling may also involve follow-up
 1504 observations (perhaps at higher resolution, greater sensitivity, or at a different wave-
 1505 length).

1506 An increasingly popular option for labelling large amounts of data is *citizen science*:
 1507 asking volunteers who are interested in contributing to science to label our data. Citi-
 1508 zen science projects can be a fantastic opportunity for both science and outreach. For
 1509 example, the ABC’s ‘Stargazing Live’ television programme engaged viewers and with
 1510 their help found four exoplanets in just 48 hours (Miller, 2017) and labelled 120 000
 1511 SkyMapper images¹ in just three days (Tucker, 2017). The downside of citizen science
 1512 is that non-expert labellers may be less accurate than experts, and indeed some may
 1513 even be malicious and provide intentionally incorrect labels (Zhang et al., 2016).

¹Citizen scientists actually produced around 5 million labels—these were aggregated to 120 000 to reduce noise.

1514 Astronomers often face a large collection of unlabelled data and must choose which
1515 to label. Choosing what to label is a broad topic of research separately in machine
1516 learning (often called active learning e.g. Gilyazev & Turdakov, 2018), in astronomy
1517 ('follow-up observations'), and in citizen science project design (e.g. citizen science
1518 project Snapshot Serengeti found that showing volunteers *uninteresting* images helped
1519 retain engagement; Sieland, 2015).

1520 **3.5.2 Label noise**

1521 *Label noise* is the presence of incorrect labels in the training or validation data set. In
1522 classical machine learning there is no such thing: labels are assumed to come from
1523 some always-correct 'oracle'. In reality, though, labels can be wrong. There is intrinsic
1524 noise in data, and even expert astronomers can disagree on labels due to ambiguities
1525 (e.g. around 10 per cent of Radio Galaxy Zoo is extremely divisive amongst expert
1526 labellers; Banfield et al., 2015). All is not lost for machine learning: many optimisation
1527 targets are robust to label noise (Menon et al., 2015). One way to think about this is
1528 that the loss function for machine learning 'smooths over' or 'averages out' the noise.

1529 It is important to note that label noise affects training and validation differently.
1530 While it is perfectly possible to train a good model with noisy labels, performance
1531 measures are not as robust to label noise. Noise in the validation set can change the re-
1532 ported performance in unpredictable ways and wherever possible should be avoided.

1533 **3.6 Summary: machine learning**

1534 Machine learning is a field of study concerned with ways of automatically finding use-
1535 ful patterns in large datasets, formalising predictive problems, converting between dif-
1536 ferent statistical problems, and handling data and uncertainty. With the large volumes
1537 of data radio astronomy faces in the lead-up to the SKA, machine learning provides
1538 an avenue for enabling scientific discovery at scale. In this chapter we introduced key
1539 concepts of machine learning, including predictors, features, labels, and loss functions.
1540 We also highlighted classification tasks as a key kind of prediction problem that we will
1541 use to frame astronomical questions throughout this thesis and introduced three clas-
1542 sification models: logistic regression, decision tree ensembles, and CNNs. In the next
1543 chapter, we will frame the astronomical problem of cross-identification as binary clas-
1544 sification, allowing us to train a cross-identification algorithm using an existing cata-
1545 logue of cross-identifications and hopefully allowing us to cross-identify radio sources
1546 en masse.

REVISION DIFF

1548

Radio Cross-identification

1549 This chapter is based on my paper *Radio Galaxy Zoo: Machine learning for radio source*
1550 *host galaxy cross-identification*, by M. J. Alger, J. K. Banfield, C. S. Ong, L. Rudnick, O. I.
1551 Wong, C. Wolf, H. Andernach, R. P. Norris, and S. S. Shabala; published in the *Monthly*
1552 *Notices of the Royal Astronomical Society* in 2018. Some minor changes have been made
1553 to match the rest of this thesis.

1554

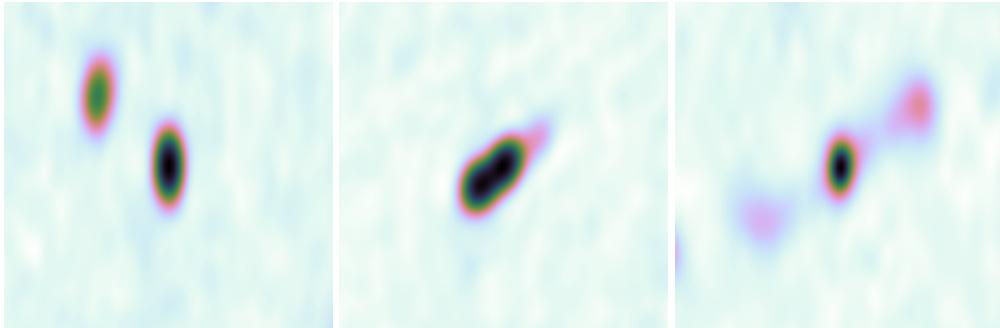
1555 In this chapter, we consider the problem of determining the host galaxies of radio
1556 sources by cross-identification. This has traditionally been done manually, which will
1557 be intractable for upcoming and ongoing wide-area radio surveys like the Evolutionary
1558 Map of the Universe (EMU). Automated cross-identification will be critical for these
1559 future surveys, and machine learning may provide the tools to develop such methods.
1560 We applied a standard approach from computer vision to cross-identification, intro-
1561 ducing one possible way of automating this problem, and explored the pros and cons
1562 of this approach. We applied our method to the 1.4 GHz Australian Telescope Large
1563 Area Survey (ATLAS) observations of the *Chandra* Deep Field South (CDFS) and the
1564 ESO Large Area ISO Survey South 1 (ELAIS-S1) fields by cross-identifying them with
1565 the *Spitzer* Wide-area Infrared Extragalactic (SWIRE) survey. We trained our method
1566 with two sets of data: expert cross-identifications of CDFS from the initial ATLAS data
1567 release and crowdsourced cross-identifications of CDFS from Radio Galaxy Zoo. We
1568 found that a simple strategy of cross-identifying a radio component with the nearest
1569 galaxy performs comparably to our more complex methods, though our estimated
1570 best-case performance is near 100 per cent. ATLAS contains 87 complex radio sources
1571 that have been cross-identified by experts, so there are not enough complex examples
1572 to learn how to cross-identify them accurately. Much larger datasets are therefore re-
1573 quired for training methods like ours. We also showed that training our method on
1574 Radio Galaxy Zoo cross-identifications gives comparable results to training on expert
1575 cross-identifications, demonstrating the value of crowdsourced training data.

1576

4.1 Introduction to cross-identification

1577

1578 Next generation radio telescopes such as the Australian SKA Pathfinder (ASKAP; John-
ston et al., 2007) and Apertif (Verheijen et al., 2008) will conduct increasingly wide,



(a) Two compact components, each a compact source.

(b) One resolved component and resolved source.

(c) Three resolved components comprising one resolved source.

Figure 4.1: Examples showing key definitions of radio emission regions used throughout this chapter. Compact and resolved components are defined by Equation 4.1.

deep, and high-resolution radio surveys, producing large amounts of data. The Evolutionary Map of the Universe (EMU; Norris et al., 2011) survey using ASKAP is expected to detect over 70 million radio sources, compared to the 2.5 million radio sources currently known (Banfield et al., 2015). An important part of processing these data is cross-identifying observed radio emission regions with observations of their host galaxy in surveys at other wavelengths.

In the presence of extended radio emission cross-identification of the host can be a difficult task. Radio emission may extend far from the host galaxy and emission regions from a single physical object may appear disconnected. As a result, the observed structure of a radio source may have a complex relationship with the corresponding host galaxy, and cross-identification in radio is much more difficult than cross-identification at shorter wavelengths. Small surveys containing a few thousand sources such as the Australia Telescope Large Area Survey (ATLAS; Middelberg et al., 2008; Norris et al., 2006) can be cross-identified manually, but this is impractical for larger surveys.

One approach to cross-identification of large numbers of sources is crowdsourcing, where volunteers cross-identify radio sources with their host galaxy. This is the premise of Radio Galaxy Zoo¹ (Banfield et al., 2015), a citizen science project hosted on the Zooniverse platform (Lintott et al., 2008). Volunteers are shown radio and infrared images and are asked to cross-identify radio sources with the corresponding infrared host galaxies. An explanation of the project can be found in Banfield et al. (2015). The first data release for Radio Galaxy Zoo will provide a large dataset of over 75 000 radio-host cross-identifications and radio source morphologies (Wong et al., in prep.). While this is a much larger number of visual cross-identifications than have been made by experts (e.g., Gendre & Wall, 2008; Grant et al., 2010; Middelberg et al., 2008; Norris et al., 2006; Taylor et al., 2007) it is still far short of the millions of radio sources expected to be detected in upcoming radio surveys (Norris, 2017b).

¹<https://radio.galaxyzoo.org>

1606 Automated algorithms have been developed for cross-identification. Fan et al. (2015)
1607 applied Bayesian hypothesis testing to this problem, fitting a three-component model
1608 to extended radio sources. This was achieved under the assumption that extended
1609 radio sources are composed of a core radio component and two lobe components.
1610 The core radio component is coincident with the host galaxy, so cross-identification
1611 amounts to finding the galaxy coincident with the core radio component in the most
1612 likely model fit. This method is easily extended to use other, more complex models, but
1613 it is purely geometric. It does not incorporate other information such as the physical
1614 properties of the potential host galaxy. Additionally, there may be new classes of radio
1615 source detected in future surveys like EMU which do not fit the model. Weston et al.
1616 (2018) developed a modification of the likelihood ratio method of cross-identification
1617 (Richter, 1975) for application to ATLAS and EMU. This method does well on non-
1618 extended radio sources with approximately 70 per cent accuracy in the ATLAS fields,
1619 but does not currently handle more complex (extended or multi-component) radio
1620 sources (Norris, 2017a).

1621 One possibility is that machine learning techniques can be developed to automatically
1622 cross-identify catalogues drawn from new surveys. Machine learning describes
1623 a class of methods that learn approximations to functions. If cross-identification can
1624 be cast as a function approximation problem, then machine learning will allow data
1625 sets such as Radio Galaxy Zoo to be generalised to work on new data. Data sets from
1626 citizen scientists have already been used to train machine learning methods. Some
1627 astronomical examples can be found in Marshall et al. (2015).

1628 In this chapter we cast cross-identification as a function approximation problem
1629 by applying an approach from computer vision literature. This approach casts cross-
1630 identification as the standard machine learning problem of binary classification by ask-
1631 ing whether a given infrared source is the host galaxy or not. We train our methods
1632 on expert cross-identifications and volunteer cross-identifications from Radio Galaxy
1633 Zoo. In Section 4.2 we describe the data we use to train our methods. In Section 4.3
1634 we discuss how we cast the radio host galaxy cross-identification problem as a ma-
1635 chine learning problem. In Section 4.4 we present results of applying our method
1636 to ATLAS observations of the *Chandra* Deep Field South (CDFS) and the ESO Large
1637 Area ISO Survey South 1 (ELAIS-S1) field. Our data, code and results are available at
1638 <https://radiogalaxyzoo.github.io/atlas-xid>.

1639 Throughout this chapter, a ‘radio source’ refers to all radio emission observed asso-
1640 ciated with a single host galaxy, and a ‘radio component’ refers to a single, contiguous
1641 region of radio emission. Multiple components may arise from a single source. A
1642 ‘compact’ source is composed of a single unresolved component. Equation 4.1 shows
1643 the definition of a resolved component. We assume that all unresolved components
1644 are compact sources, i.e. we assume that each unresolved component has its own host
1645 galaxy². An ‘extended’ source is a non-compact source, i.e. resolved single-component
1646 sources or a multi-component source. Figure 4.1 illustrates these definitions.

²This will be incorrect if the unresolved components are actually compact lobes or hotspots, but determining which components correspond to unique radio sources is outside the scope of this thesis.

Table 4.1: Catalogues of ATLAS/SWIRE cross-identifications for the CDFS and ELAIS-S1 fields. The method used to generate each catalogue is shown, along with the number of radio components cross-identified in each field.

Catalogue	Method	CDFS	ELAIS-S1
Norris et al. (2006)	Manual	784	0
Middelberg et al. (2008)	Manual	0	1366
Fan et al. (2015)	Bayesian models	784	0
Weston et al. (2018)	Likelihood ratio	3078	2113
Wong et al. (in prep)	Crowdsourcing	2460	0

1647 4.2 Data

1648 We use radio data from the Australia Telescope Large Area Survey (ATLAS; Franzen
 1649 et al., 2015; Norris et al., 2006), infrared data from the *Spitzer* Wide-area Infrared
 1650 Extragalactic survey (SWIRE; Lonsdale et al., 2003; Surace et al., 2005), and cross-
 1651 identifications of these surveys from the citizen science project Radio Galaxy Zoo (Ban-
 1652 field et al., 2015). Radio Galaxy Zoo also includes cross-identifications of sources in
 1653 Faint Images of the Radio Sky at Twenty Centimeters (FIRST; White et al., 1997) and
 1654 the AllWISE survey (Cutri et al., 2013), though we focus only on Radio Galaxy Zoo
 1655 data from ATLAS and SWIRE.

1656 4.2.1 ATLAS

1657 ATLAS is a pilot survey for the EMU (Norris et al., 2011) survey, which will cover
 1658 the entire sky south of +30 deg and is expected to detect approximately 70 million
 1659 new radio sources. 95 per cent of these sources will be single-component sources, but
 1660 the remaining 5 per cent pose a considerable challenge to current automated cross-
 1661 identification methods (Norris et al., 2011). EMU will be conducted at the same depth
 1662 and resolution as ATLAS, so methods developed for processing ATLAS data are ex-
 1663 pected to work for EMU. ATLAS is a wide-area radio survey of the CDFS and ELAIS-
 1664 S1 fields at 1.4 GHz with a sensitivity of 14 and 17 $\mu\text{Jy beam}^{-1}$ on CDFS and ELAIS-S1
 1665 respectively. CDFS covers 3.6 deg^2 and contains 3034 radio components above a signal-
 1666 to-noise ratio of 5. ELAIS-S1 covers 2.7 deg^2 and contains 2084 radio components above
 1667 a signal-to-noise ratio of 5 (Franzen et al., 2015). The images of CDFS and ELAIS-S1
 1668 have angular resolutions of 16 by 7 and 12 by 8 arcsec respectively, with pixel sizes
 1669 of $1.5 \text{ arcsec px}^{-1}$. Table 4.1 summarises catalogues that contain cross-identifications
 1670 of radio components in ATLAS with host galaxies in SWIRE. In the present work, we
 1671 train methods on CDFS³ and test these methods on both CDFS and ELAIS-S1. This
 1672 ensures helps confirm that our methods are transferable to different areas of the sky
 1673 observed by the same telescope, as will be the case for EMU.

³Radio Galaxy Zoo only contains CDFS sources and so we cannot train methods on ELAIS-S1.

1674 **4.2.2 SWIRE**

1675 SWIRE is a wide-area infrared survey at the four IRAC wavelengths 3.6, 4.5, 5.8, and
1676 8.0 μm (Lonsdale et al., 2003; Surace et al., 2005). It covers eight fields, including CDFS
1677 and ELAIS-S1. SWIRE is the source of infrared observations for cross-identification
1678 with ATLAS. SWIRE has catalogued 221,535 infrared objects in CDFS and 186,059 in-
1679 frared objects in ELAIS-S1 above a signal-to-noise ratio of 5.

1680 **4.2.3 Radio Galaxy Zoo**

1681 Radio Galaxy Zoo asks volunteers to cross-identify radio components with their in-
1682 frared host galaxies. There are a total of 2460 radio components in Radio Galaxy Zoo
1683 sourced from ATLAS observations of CDFS. These components are cross-identified by
1684 Radio Galaxy Zoo participants with host galaxies detected in SWIRE. A more detailed
1685 description can be found in Banfield et al. (2015) and a full description of how the Ra-
1686 dio Galaxy Zoo catalogue used in this work⁴ is generated can be found in Wong et al.
1687 (in prep).

1688 The ATLAS CDFS radio components that appear in Radio Galaxy Zoo are drawn
1689 from a prerelease version of the third data release of ATLAS by Franzen et al. (2015). In
1690 this release, each radio component was fit with a two-dimensional Gaussian. Depend-
1691 ing on the residual of the fit, more than one Gaussian may be fit to one region of radio
1692 emission. Each of these Gaussian fits is listed as a radio component in the ATLAS com-
1693 ponent catalogue. The brightest radio component from the multiple-Gaussian fit is
1694 called the ‘primary component’. If there was only one Gaussian fit then this Gaussian
1695 is the primary component. Each primary component found in the ATLAS component
1696 catalogue appears in Radio Galaxy Zoo. Non-primary components may appear within
1697 the image of a primary component, but do not have their own entry in Radio Galaxy
1698 Zoo. We will henceforth only discuss the primary components.

1699 **4.3 Method**

1700 The aim of this chapter is to express cross-identification in a form that will allow us
1701 to apply standard machine learning tools and methods. We use an approach from
1702 computer vision to cast cross-identification as binary classification.

1703 **4.3.1 Cross-identification as binary classification**

1704 We propose a two-step method for host galaxy cross-identification which we will de-
1705 scribe now. Given a radio component, we want to find the corresponding host galaxy.
1706 The input is a $2' \times 2'$ radio image of the sky centred on a radio component and po-
1707 tentially other information about objects in the image (such as the redshift or infrared

⁴The Radio Galaxy Zoo Data Release 1 catalogue will only include cross-identifications for which over 65 per cent of volunteers agree. However, we use a preliminary catalogue containing volunteer cross-identifications for all components.

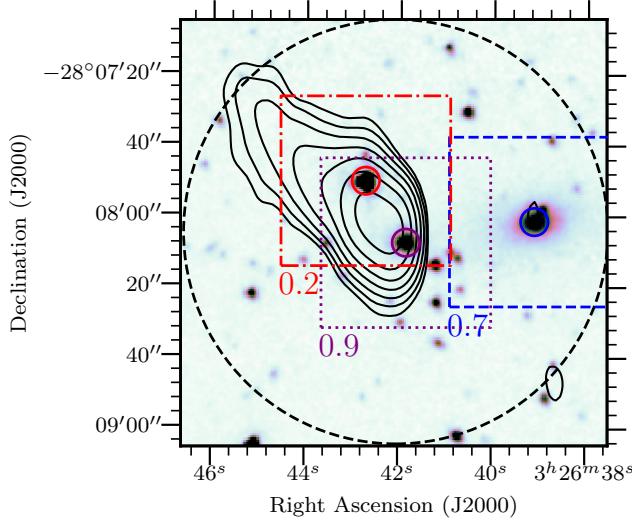


Figure 4.2: An example of finding the host galaxy of a radio source using our sliding-window method. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. Boxes represent ‘windows’ centred on candidate host galaxies, which are circled. The pixels in each window are used to represent the candidate that the window is centred on. The scores of each candidate would be calculated by a binary classifier using the window as input, and these scores are shown below each window. The scores shown are for illustration only. In this example, the galaxy coincident with the centre window would be chosen as the host galaxy, as this window has the highest score. The dashed circle shows the $1'$ radius from which candidate host galaxies are selected. For clarity, not all candidate host galaxies are shown.

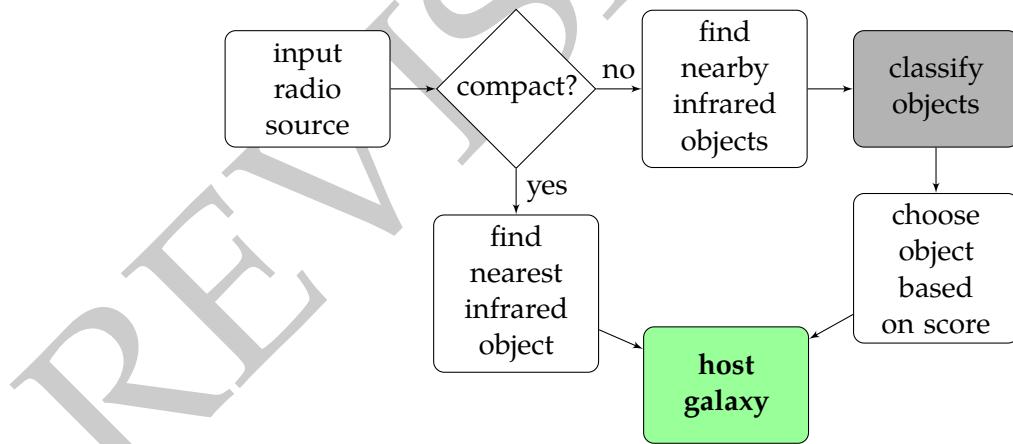


Figure 4.3: Our cross-identification method once a binary classifier has been trained. As input we accept a radio component. If the component is compact, we assume it is a compact source and select the nearest infrared object as the host galaxy. If the component is resolved, we use the binary classifier to score all nearby infrared objects and select the highest-scored object as the host galaxy. Compact and resolved components are defined in Equation 4.1.

1708 colour). Images at other wavelengths (notably infrared) might be useful, but we defer
 1709 this for now as it complicates the task. We chose a $2' \times 2'$ image to match the size of
 1710 the images used by Radio Galaxy Zoo. To avoid solving the separate task of identi-
 1711 fying which radio components are associated with the same source, we assume that
 1712 each radio image represents a single extended source⁵. Radio cross-identification can
 1713 then be formalised as follows: given a radio image centred on a radio component, lo-
 1714 cate the host galaxy of the source containing this radio component. This is a standard
 1715 computer vision problem called ‘object detection’, and we apply a common technique
 1716 called a ‘sliding-window’ (Rowley et al., 1996).

1717 In sliding-window object detection, we want to find an object in an image. We
 1718 develop a function to score each location in the image such that the highest-scored
 1719 location coincides with the desired object. Square image cutouts called ‘windows’ are
 1720 taken centred on each location and these windows are used to represent that location
 1721 in our scoring function. To find the infrared host galaxy, we choose the location with
 1722 the highest score. To improve the efficiency of this process when applied to cross-
 1723 identification, we only consider windows coincident with infrared sources detected in
 1724 SWIRE. We call these infrared sources ‘candidate host galaxies’. For this chapter, there
 1725 is no use in scoring locations without infrared sources as that would not lead to a host
 1726 identification anyway. Using candidate host galaxies instead of pixels also allows us to
 1727 include ancillary information about the candidate host galaxies, such as their infrared
 1728 colours and redshifts. We refer to the maximum distance a candidate host galaxy can
 1729 be separated from a radio component as the ‘search radius’ and take this radius to be 1
 1730 arcmin. To score each candidate host galaxy we use a ‘binary classifier’, which we will
 1731 define now.

1732 Binary classification is a common method in machine learning where objects are to
 1733 be assigned to one of two classes, called the ‘positive’ and ‘negative’ classes. This as-
 1734 signment is represented by the probability that an object is in the positive class. A
 1735 ‘binary classifier’ is a function mapping from an object to such a probability. Our
 1736 formulation of cross-identification is equivalent to binary classification of candidate
 1737 host galaxies: the positive class represents host galaxies, the negative class represents
 1738 non-host galaxies, and to cross-identify a radio component we find the candidate host
 1739 galaxy maximising the positive class probability. In other words, the binary classifier
 1740 is exactly the sliding-window scoring function. We therefore split cross-identification
 1741 into two separate tasks: the ‘candidate classification task’ where, given a candidate
 1742 host galaxy, we wish to determine whether it is a host galaxy of *any* radio component;
 1743 and the ‘cross-identification task’ where, given a specific radio component, we wish to
 1744 find its host galaxy. The candidate classification task is a traditional machine learning
 1745 problem which results in a binary classifier. To avoid ambiguity and recognise that
 1746 the values output by a binary classifier are not true probabilities, we will refer to the
 1747 outputs of the binary classifier as ‘scores’ in line with the sliding-window approach
 1748 described above. The cross-identification task maximises over scores output by this
 1749 classifier. Our approach is illustrated in Figure 4.2 and described in Algorithm 1. We

⁵Limitations of this assumption are discussed in Section 4.3.2.

Algorithm 1: Cross-identifying a radio component given a radio image of the component, a catalogue of infrared candidate host galaxies and a binary classifier. σ is a parameter of the method.

Data:

A 2×2 arcmin radio image of a radio component

A set of infrared candidate host galaxies \mathcal{G}

A binary classifier $f : \mathbb{R}^k \rightarrow \mathbb{R}$

Result: A galaxy $g \in \mathcal{G}$

$max \leftarrow -\infty;$

$host \leftarrow \emptyset;$

for $g \in \mathcal{G}$ **do**

$x \leftarrow$ a k -dimensional vector representation of g (Section 4.3.3);

$d \leftarrow$ distance between g and the radio component;

$score \leftarrow f(x) \times \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{d^2}{2\sigma^2}\right);$

if $score > max$ **then**

$max \leftarrow score;$

$host \leftarrow g;$

end

end

return $host$

1750 refer to the binary classifier scoring a candidate host galaxy as f . To implement f as a
 1751 function that accepts candidate host galaxies as input, we need to represent candidate
 1752 host galaxies by vectors. We describe this in Section 4.3.3. There are many options
 1753 for modelling f . In this chapter we apply three different models: logistic regression,
 1754 random forests and convolutional neural networks.

1755 We cross-identify each radio component in turn. The classifier f provides a score
 1756 for each candidate host galaxy. This score indicates how much the candidate looks
 1757 like a host galaxy, independent of which radio component we are currently cross-
 1758 identifying. If there are other nearby host galaxies, then multiple candidate hosts may
 1759 have high scores (e.g. Figure 4.4). This difficulty is necessary—a classifier with de-
 1760 pendence on radio object would be impossible to train. We need multiple positive
 1761 examples (i.e. host galaxies) to train a binary classifier, but for any specific radio com-
 1762 ponent there is only one host galaxy. As a result, the candidate classification task aims
 1763 to answer the general question of whether a given galaxy is the host galaxy of *any* ra-
 1764 dio component, while the cross-identification task attempts to cross-identify a *specific*
 1765 radio component. To distinguish between candidate host galaxies with high scores, we
 1766 weight the scores by a Gaussian function of angular separation between the candidates
 1767 and the radio component. The width of the Gaussian, σ , controls the influence of the
 1768 Gaussian on the final cross-identification. When σ is small, our approach is equivalent
 1769 to a nearest neighbours approach where we select the nearest infrared object to the
 1770 radio component as the host galaxy. In the limit where $\sigma \rightarrow \infty$, we maximise the score
 1771 output by the classifier as above. We take $\sigma = 30''$ as this was the best value found by

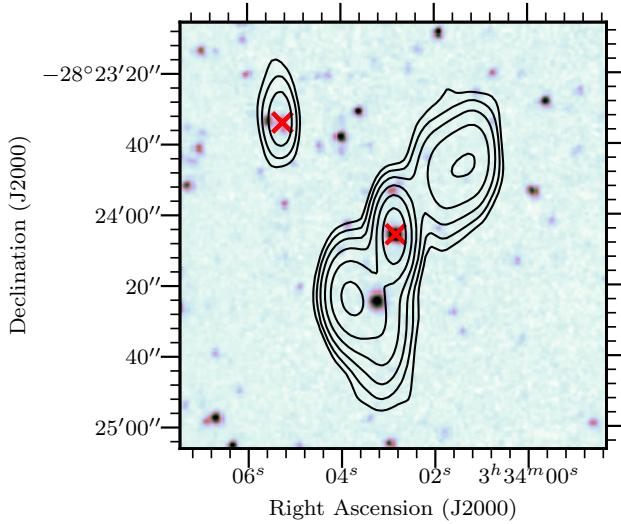


Figure 4.4: A $2'$ -wide radio image centred on ATLAS3_J033402.87-282405.8C. This radio source breaks the assumption that there are no other radio sources within 1 arcmin of the source. Another radio source is visible to the upper-left. Host galaxies found by Radio Galaxy Zoo volunteers are shown by crosses. The background image is a $3.6\text{ }\mu\text{m}$ image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2.

1772 a grid search [maximising cross-identification accuracy](#). Note that the optimum width
 1773 will depend on the density of radio sources on the sky, the angular separation of the
 1774 host galaxy and its radio components and the angular resolution of the survey.

1775 We can improve upon this method by cross-identifying compact radio sources sep-
 1776 arately from extended sources, as compact sources are much easier to cross-identify.
 1777 For a compact source, the nearest SWIRE object may be identified as the host galaxy (a
 1778 *nearest neighbours* approach), or a more complex method such as likelihood ratios may
 1779 be applied (see Weston et al., 2018). We cross-identify compact sources separately in
 1780 our pipeline and this process is shown in Figure 4.3.

1781 4.3.2 Limitations of our approach

1782 We make a number of assumptions to relate the cross-identification task to the candi-
 1783 date classification task:

- 1784 1. For any radio component, the $2' \times 2'$ image centred on the component contains
 1785 components of only one radio source.
- 1786 2. For any radio component, the $2' \times 2'$ image centred on the component contains
 1787 all components of this source.
- 1788 3. The host galaxy of a radio component is within the 1 arcmin search radius around
 1789 the component, measured from the centre of the Gaussian fit.

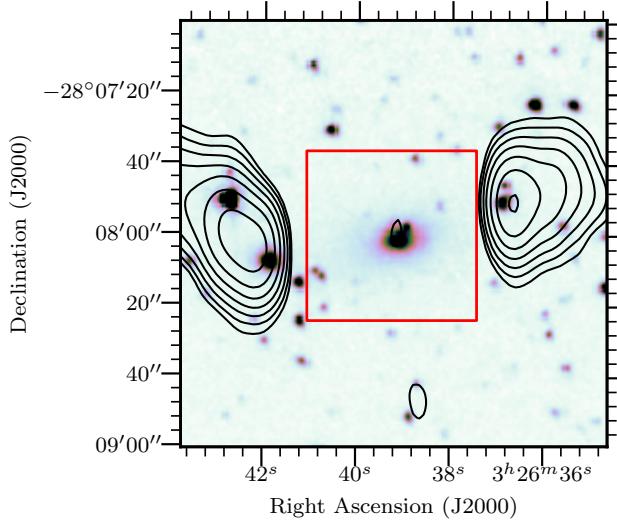


Figure 4.5: An example of a radio source where the window centred on the host galaxy, shown as a rectangle, does not contain enough radio information to correctly identify the galaxy as the host. The background image is a 3.6 μm image from SWIRE. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2.

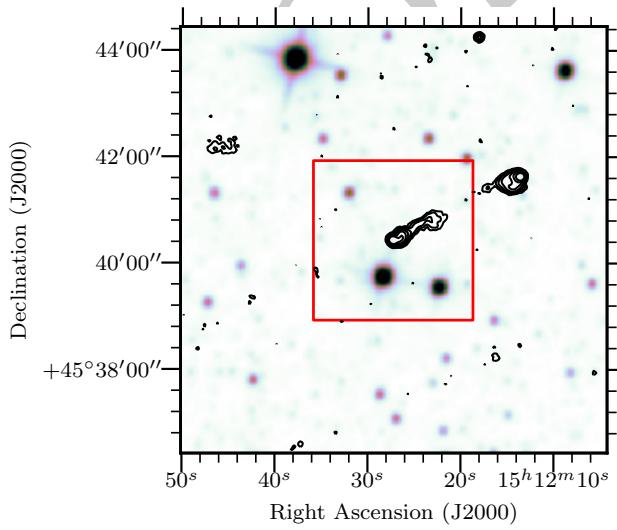


Figure 4.6: A 8'-wide radio image from FIRST, centred on FIRST J151227.2+454026. The 3'-wide red box indicates the boundaries of the image of this radio component shown to volunteers in Radio Galaxy Zoo. This radio source breaks our assumption that the whole radio source is visible in the chosen radius. As one of the components of the radio source is outside of the image, a volunteer (or automated algorithm) looking at the 3'-wide image may be unable to determine that this is a radio double or locate the host galaxy. The background image is a 3.4 μm image from WISE. The contours show FIRST radio data, starting at 4σ and increasing geometrically by a factor of 2.

1790 4. The host galaxy of a radio component is closer on the sky to the radio component
1791 than the host galaxy of any other radio component.

1792 5. The host galaxy appears in the SWIRE catalogue.

1793 These assumptions limit the effectiveness of our approach, regardless of how accurate
1794 our binary classifier may be. Examples of radio sources that break these respective
1795 assumptions are:

1796 1. A radio source less than $1'$ away from another radio source.

1797 2. A radio source with an angular size greater than $2'$.

1798 3. A radio source with a component greater than $1'$ away from the host galaxy.

1799 4. A two-component radio source with another host galaxy between a component
1800 and the true host galaxy.

1801 5. An infrared-faint radio source (as in Collier et al., 2014).

1802 The main limitations are problems of scale in choosing the candidate search radius
1803 and the size of the windows representing candidates. If the search radius is too small,
1804 we may not consider the host galaxy as a candidate. If the search radius is too large, we
1805 may consider multiple host galaxies (though this is mostly mitigated by the Gaussian
1806 weighting). If the window is too small, radio emission may extend past the edges of
1807 the window and we may miss critical information required to identify the galaxy as a
1808 host galaxy. If the window is too large, then irrelevant information will be included
1809 and it may be difficult or computationally expensive to score. We chose a window size
1810 of 32×32 pixels, corresponding to approximately $48'' \times 48''$ in ATLAS. This is shown
1811 as squares in Figure 4.2 and Figure 4.5. These kinds of size problems are difficult even
1812 for non-automated methods as radio sources can be extremely wide—for example, Ra-
1813 dio Galaxy Zoo found a radio giant that spanned over three different images presented
1814 to volunteers and the full source was only cross-identified by the efforts of citizen sci-
1815 entists (Banfield et al., 2015). An example of a radio image where part of the radio
1816 source is outside the search radius is shown in Figure 4.6.

1817 In weighting the scores by a Gaussian function of angular separation, we implicitly
1818 assume that the host galaxy of a radio component is closer to that radio component
1819 than any other host galaxy. If this assumption is not true then the incorrect host galaxy
1820 may be identified, though this is rare.

1821 We only need to require that the host galaxy appears in SWIRE to incorporate
1822 galaxy-specific features (Section 4.3.3) and to improve efficiency. Our method is ap-
1823 plicable even when host galaxies are not detected in the infrared by considering every
1824 pixel of the radio image as a candidate location as would be done in the original com-
1825 puter vision approach. If the host galaxy location does not correspond to an infrared
1826 source, the radio source would be classified as infrared-faint.

1827 Our assumptions impose an upper bound on how well we can cross-identify radio
1828 sources. We estimate this upper bound in Section 4.4.1.

1829 **4.3.3 Feature vector representation of infrared sources**

1830 Inputs to binary classifiers must be represented by an array of real values called feature
 1831 vectors. We therefore need to choose a feature vector representation of our candidate
 1832 host galaxies. Candidate hosts are sourced from the SWIRE catalogue (Section 4.2.2).
 1833 We represent each candidate host with 1034 real-valued features, combining the win-
 1834 dows centred on each candidate (Section 4.3.1) with ancillary infrared data from the
 1835 SWIRE catalogue. For a given candidate host, these features are:

- 1836 • the 6 base-10 logarithms of the ratios of fluxes of the candidate host at the four
 1837 IRAC wavelengths (the ‘colours’ of the candidate);
- 1838 • the flux of the host at 3.6 μm ;
- 1839 • the stellarity index of the host at both 3.6 and 4.5 μm ;
- 1840 • the radial distance between the candidate host and the nearest radio component
 1841 in the ATLAS catalogue; and
- 1842 • a 32×32 pixel image from ATLAS (approximately $48'' \times 48''$), centred on the
 1843 candidate host (the window).

1844 The infrared colours provide insight into the properties of the candidate host galaxy
 1845 (Grant, 2011). The 3.6 and 4.5 μm fluxes trace both galaxies with faint polycyclic aro-
 1846 matic hydrocarbon (PAH) emission (i.e. late-type, usually star-forming galaxies) and
 1847 elliptical galaxies dominated by old stellar populations. The 5.8 μm flux selects galax-
 1848 ies where the infrared emission is dominated by non-equilibrium emission of dust
 1849 grains due to active galactic nuclei, while the 8.0 μm flux traces strong PAH emission
 1850 at low redshift (Sajina et al., 2005). The stellarity index is a value in the SWIRE cata-
 1851 logue that represents how likely the object is to be a star rather than a galaxy (Surace
 1852 et al., 2005). It was estimated by a neural network in SExtractor (Bertin & Arnouts,
 1853 1996).

1854 We use the 32×32 pixels of each radio window as independent features for all
 1855 binary classification models, with the convolutional neural network automatically ex-
 1856 tracting features that are relevant. Other features of the radio components may be
 1857 used instead of just relying on the pixel values, but there has been limited research on
 1858 extracting such features: Proctor (2006) describes hand-selected features for radio dou-
 1859 bles in FIRST, and Aniyan and Thorat (2017) and Lukic et al. (2018) make use of deep
 1860 convolutional neural networks which automatically extract features as part of classifi-
 1861 cation. A more comprehensive investigation of features is a good avenue for potential
 1862 improvement in our pipeline but this is beyond the scope of this initial study.

1863 **4.3.4 Binary Classifiers**

1864 We use three different binary classification models: logistic regression, convolutional
 1865 neural networks, and random forests. These models cover three different approaches
 1866 to machine learning. Logistic regression is a probabilistic binary classification model.

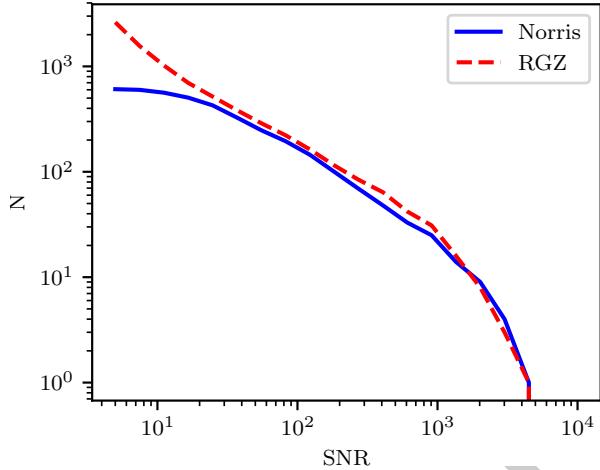


Figure 4.7: Cumulative number of radio components (N) in the expert (Norris) and Radio Galaxy Zoo (RGZ) training sets with different signal-to-noise ratios (SNR).

Table 4.2: Number of compact and resolved radio objects in each CDFS quadrant. Radio Galaxy Zoo (RGZ) has more cross-identifications than the expert catalogue (Norris et al., 2006) provides as it uses a deeper data release of ATLAS, and so has more objects in each quadrant for training.

Quadrant	Compact	Resolved	Compact (RGZ)	Resolved (RGZ)
0	126	24	410	43
1	99	21	659	54
2	61	24	555	57
3	95	18	631	51
<i>Total</i>	381	87	2255	205

1867 It is linear in the feature space and outputs the probability that the input has a positive
 1868 label (Bishop, 2006, Chap. 4). Convolutional neural networks (CNN) are biologically-
 1869 inspired prediction models with image inputs. They have recently produced good
 1870 results on large image-based datasets in astronomy (Dieleman et al., 2015; Lukic et al.,
 1871 2018, e.g.). Random forests are an ensemble of decision trees (Breiman, 2001). They
 1872 consider multiple subsamples of the training set, where each bootstrap subsample is
 1873 sampled with replacement from the training set. To classify a new data point, the ran-
 1874 dom forest takes the weighted average of all classifications produced by each decision
 1875 tree. [For a more detailed description of these models, see Appendix A.](#)

1876 4.3.5 Labels

1877 The Radio Galaxy Zoo and Norris et al. (2006) cross-identification catalogues must be
 1878 converted to binary labels for infrared objects so that they can be used to train binary

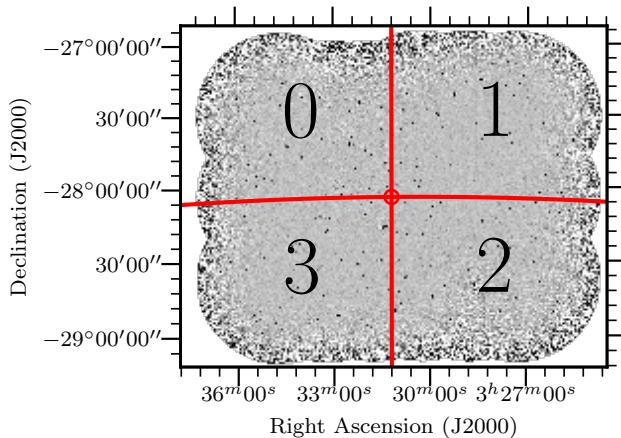


Figure 4.8: CDFS field training and testing quadrants labelled 0 – 3. The central dot is located at $\alpha = 03^{\text{h}}31^{\text{m}}12^{\text{s}}$, $\delta = -28^{\circ}06'00''$. The quadrants were chosen such that there are similar numbers of radio sources in each quadrant.

1879 classifiers. There are two challenges with this conversion:

- 1880 • We can only say that an object is *a* host galaxy, not which radio object it is associated with, and
- 1881
- 1882 • We cannot disambiguate between non-host infrared objects and host galaxies that were not in the cross-identification catalogue.
- 1883

1884 We use the Gaussian weighting described in Section 4.3.1 to address the first issue.

1885 The second issue is known as a ‘positive-unlabelled’ classification problem, which is a
1886 binary classification problem where we only observe labels for the positive class. We
1887 treat unlabelled objects as negative examples following Menon et al. (2015). That is,
1888 we make the naïve assumption that any infrared object in the SWIRE catalogue not
1889 identified as a host galaxy in a cross-identification catalogue is not a host galaxy at all.

1890 We first generate positive labels from a cross-identification catalogue. We decide
1891 that if an infrared object is listed in the catalogue, then it is assigned a positive label
1892 as a host galaxy. We then assign every other galaxy a negative label. This has some
1893 problems—an example is that if the cross-identification catalogue did not include a
1894 radio object (e.g. it was below the signal-to-noise ratio) then the host galaxy of that
1895 radio object would receive a negative label. This occurs with Norris et al. (2006) cross-
1896 identifications, as these are associated with the first data release of ATLAS. The first
1897 data release went to a 5σ flux density level of $S_{1.4} \geq 200 \mu\text{Jy beam}^{-1}$ (Norris et al.,
1898 2006), compared to $S_{1.4} \geq 85 \mu\text{Jy beam}^{-1}$ for the third data release used by Radio
1899 Galaxy Zoo (Franzen et al., 2015). The labels from Norris et al. (2006) may therefore
1900 disagree with labels from Radio Galaxy Zoo even if they are both plausible. The dif-
1901 ference in training set size at different flux cutoffs is shown in Figure 4.7. We train and
1902 test our binary classifiers on infrared objects within a 1 arcmin radius of an ATLAS
1903 radio component.

1904 **4.3.6 Experimental Setup**

1905 We trained binary classifiers on infrared objects in the CDFS field using two sets of
 1906 labels. One label set was derived from Radio Galaxy Zoo cross-identifications and
 1907 the other was derived from the Norris et al. (2006) cross-identification catalogue. We
 1908 refer to these as the ‘Radio Galaxy Zoo labels’ and the ‘expert labels’ respectively. We
 1909 divided the CDFS field into four quadrants for training and testing. The quadrants
 1910 were divided with a common corner at $\alpha = 03^{\text{h}}31^{\text{m}}12^{\text{s}}$, $\delta = -28^{\circ}06'00''$ as shown in
 1911 Figure 4.8. For each trial, one quadrant was used to extract test examples and the other
 1912 three quadrants were used for training examples.

We further divided the radio components into compact and resolved. Compact components are cross-identified by fitting a 2D Gaussian (as in Norris et al., 2006) and we would expect any machine learning approach for host cross-identification to attain high accuracy on this set. A radio component was considered resolved if

$$\ln \left(\frac{S_{\text{int}}}{S_{\text{peak}}} \right) > 2 \sqrt{ \left(\frac{\sigma_{S_{\text{int}}}}{S_{\text{int}}} \right)^2 + \left(\frac{\sigma_{S_{\text{peak}}}}{S_{\text{peak}}} \right)^2 }, \quad (4.1)$$

1913 where S_{int} is the integrated flux density, S_{peak} is the peak flux density, $\sigma_{S_{\text{int}}}$ is the un-
 1914 certainty in integrated flux density and $\sigma_{S_{\text{peak}}}$ is the uncertainty in peak flux density
 1915 (following Franzen et al., 2015).

1916 Candidate hosts were selected from the SWIRE catalogue. For a given subset of
 1917 radio components, all SWIRE objects within 1 arcmin of all radio components in the
 1918 subset were added to the associated SWIRE subset. In results for the candidate classi-
 1919 fication task, we refer to SWIRE objects within 1 arcmin of a compact radio component
 1920 as part of the ‘compact set’, and SWIRE objects within 1 arcmin of a resolved radio
 1921 component as part of the ‘resolved set’.

1922 To reduce bias in the testing data due to the expert labels being generated from a
 1923 shallower data release of ATLAS, a SWIRE object was only included in the test set if
 1924 it was within 1 arcmin of a radio object with a SWIRE cross-identification in both the
 1925 Norris et al. (2006) catalogue and the Radio Galaxy Zoo catalogue.

1926 Each binary classifier was trained on the training examples and used to score the
 1927 test examples. These scores were thresholded to generate labels which could be di-
 1928 rectly compared to the expert labels. We then computed the ‘balanced accuracy’ of
 1929 these predicted labels. Balanced accuracy is the average of the accuracy on the positive
 1930 class and the accuracy on the negative class, and is not sensitive to class imbalance. The
 1931 candidate classification task has highly imbalanced classes—in our total set of SWIRE
 1932 objects within 1 arcmin of an ATLAS object, only 4 per cent have positive labels. Our
 1933 threshold was chosen to maximise the balanced accuracy on predicted labels of the
 1934 training set. Only examples within 1 arcmin of ATLAS objects in the first ATLAS data
 1935 release (Norris et al., 2006) were used to compute balanced accuracy, as these were
 1936 the only ATLAS objects with expert labels.

1937 We then used the scores to predict the host galaxy for each radio component cross-
 1938 identified by both Norris et al. (2006) and Radio Galaxy Zoo. We followed Algo-

1939 rithm 1: the score of each SWIRE object within 1 arcmin of a given radio component
 1940 was weighted by a Gaussian function of angular separation from the radio component
 1941 and the object with the highest weighted score was chosen as the host galaxy. The
 1942 cross-identification accuracy was then estimated as the fraction of the predicted host
 1943 galaxies that matched the Norris et al. (2006) cross-identifications.

1944 **4.4 Results**

1945 In this section we present accuracies of our method trained on CDFS and applied to
 1946 CDFS and ELAIS-S1, as well as results motivating our accuracy measures and esti-
 1947 mates of upper and lower bounds for cross-identification accuracy using our method.

1948 **4.4.1 Application to ATLAS-CDFS**

1949 We can assess trained binary classifiers either by their performance on the candidate
 1950 classification task or by their performance on the cross-identification task when used
 1951 in our method. Both performances are useful: performance on the candidate classi-
 1952 fication task provides a robust and simple way to compare binary classifiers without
 1953 the limitations of our specific formulation, and performance on the cross-identification
 1954 task can be compared with other cross-identification methods. We therefore report two
 1955 sets of accuracies: balanced accuracy for the galaxy classification task and accuracy for
 1956 the cross-identification task. These accuracy measures are correlated and we show this
 1957 correlation in Figure 4.9. Fitting a line of best fit with `scipy` gives $R^2 = 0.92$ for logistic
 1958 regression and $R^2 = 0.87$ for random forests. While performance on the candidate clas-
 1959 sification task is correlated with performance on the cross-identification task, balanced
 1960 accuracy does not completely capture the effectiveness of a binary classifier applied to
 1961 the cross-identification task. This is because while our binary classifiers output real-
 1962 valued scores, these scores are thresholded to compute the balanced accuracy. In the
 1963 candidate classification task, the binary classifier only needs to ensure that host galax-
 1964 ies are scored higher than non-host galaxies. This means that after thresholding there
 1965 can be many ‘false positives’ that do not affect cross-identification. An example of this
 1966 is shown in Figure 4.10, where the classifier has identified 8 ‘host galaxies’. However,
 1967 there are only three true host galaxies in this image—one per radio component—and so
 1968 in the cross-identification task, only three of these galaxies will be identified as hosts.

1969 In Figure 4.11 we plot the balanced accuracies of our classification models on the
 1970 candidate classification task and the cross-identification accuracies of our method us-
 1971 ing each of these models. Results are shown for both the resolved and compact sets.
 1972 For comparison, we also plot the cross-identification accuracy of Radio Galaxy Zoo and
 1973 a nearest neighbours approach, as well as estimates for upper and lower limits on the
 1974 cross-identification accuracy. We estimate the upper limit on performance by assign-
 1975 ing all true host galaxies a score of 1 and assigning all other candidate host galaxies a
 1976 score of 0. This is equivalent to ‘perfectly’ solving the candidate classification task and
 1977 so represents the best possible cross-identification performance achievable with our
 1978 method. We estimate the lower limit on performance by assigning random scores to

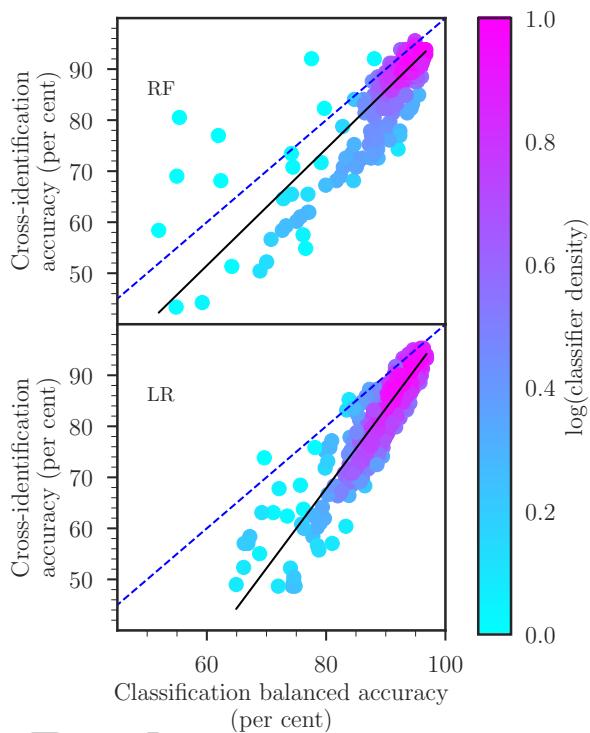


Figure 4.9: Balanced accuracy on the candidate classification task plotted against accuracy on the cross-identification task. 'RF' indicates results from random forests, and 'LR' indicates results from logistic regression. Binary classifiers were trained on random, small subsets of the training data to artificially restrict their accuracies. Colour shows the density of points on the plot estimated by a Gaussian kernel density estimate. The solid lines indicate the best linear fit; these fits have $R^2 = 0.92$ for logistic regression and $R^2 = 0.87$ for random forests. The dashed line shows the line where cross-identification accuracy and candidate classification accuracy are equal. We did not include convolutional neural networks in this test, as training them is very computationally expensive. There are 640 trials shown per classification model. These results exclude binary classifiers with balanced accuracies less than 51 per cent, as these are essentially random.

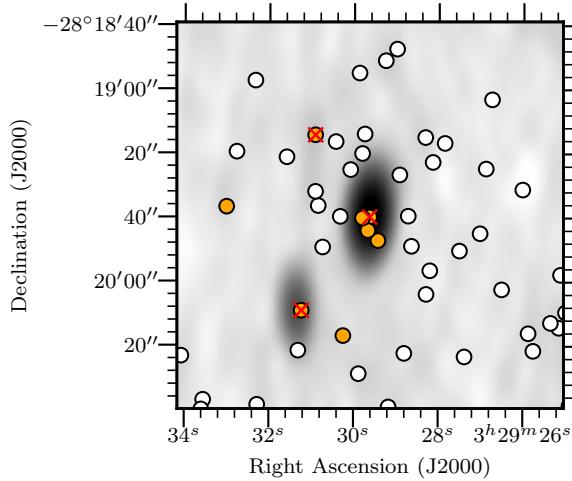


Figure 4.10: Predicted host galaxies in the candidate classification task for ATLAS3 J032929.61-281938.9. The background image is an ATLAS radio image. Radio Galaxy Zoo host galaxies are marked by crosses. SWIRE candidate host galaxies are circles coloured by the score output by a logistic regression binary classifier. The scores are thresholded to obtain labels, as when we compute balanced accuracy. Orange circles have been assigned a ‘positive’ label by a logistic regression binary classifier and white otherwise. Note that there are more predicted host galaxies than there are radio components, so not all of the predicted host galaxies would be assigned as host galaxies in the cross-identification task.

1979 each candidate host galaxy. We expect any useful binary classifier to produce better re-
1980 sults than this, so this represents the lowest expected cross-identification performance.

1981 The upper estimates, lower estimates, and nearest-neighbour accuracy are shown as
1982 horizontal lines in Figure 4.11.

1983 In Figure 4.12 we plot the performance of our method using different binary classi-
1984 fication models, as well as the performance of Radio Galaxy Zoo, nearest neighbours,
1985 and the perfect and random binary classifiers, on the full set of ATLAS DR1 radio
1986 components using the pipeline in Figure 4.3. The accuracy associated with each clas-
1987 sification model and training label set averaged across all four quadrants is shown in
1988 Appendix B.

1989 Differences between accuracies across training labels are well within one standard
1990 deviation computed across the four quadrants, with convolutional neural networks on
1991 compact objects as the only exception. The spread of accuracies is similar for both sets
1992 of training labels, with the exception of random forests. The balanced accuracies of
1993 random forests trained on expert labels have a considerably higher spread than those
1994 trained on Radio Galaxy Zoo labels, likely because of the small size of the expert train-
1995 ing set—there are less than half the number of objects in the expert-labelled training set
1996 than the number of objects in the Radio Galaxy Zoo-labelled training set (Table 4.2).

1997 Radio Galaxy Zoo-trained methods significantly outperform Radio Galaxy Zoo
1998 cross-identifications. Additionally, despite poor performance of Radio Galaxy Zoo on

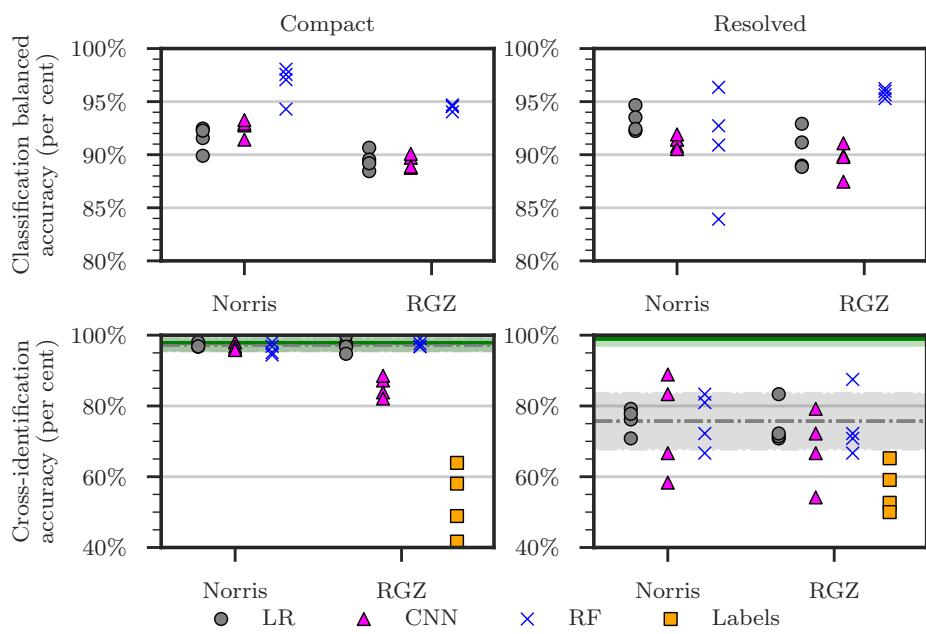


Figure 4.11: Performance of our method with logistic regression ('LR'), convolutional neural networks ('CNN') and random forest ('RF') binary classifiers. 'Norris' indicates the performance of binary classifiers trained on the expert labels and 'RGZ' indicates the performance of binary classifiers trained on the Radio Galaxy Zoo labels. One point is shown per binary classifier per testing quadrant. The training and testing sets have been split into compact (left) and resolved (right) objects. Shown for comparison is the accuracy of the Radio Galaxy Zoo consensus cross-identifications on the cross-identification task, shown as 'Labels'. The cross-identification accuracy attained by a perfect binary classifier is shown by a solid green line, and the cross-identification accuracy of a nearest neighbours approach is shown by a dashed grey line. The standard deviation of these accuracies across the four CDFS quadrants is shown by the shaded area. Note that the pipeline shown in Figure 4.3 is not used for these results.

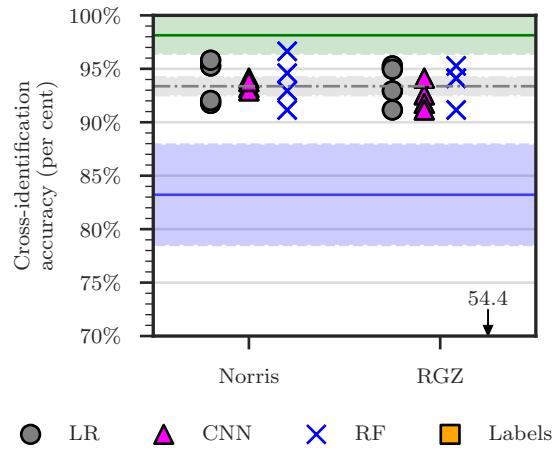


Figure 4.12: Performance of our approach using different binary classifiers on the cross-identification task. Markers and lines are as in Figure 4.11. The blue solid line indicates the performance of a random binary classifier and represents the minimum accuracy we expect to obtain. The standard deviation of this accuracy across 25 trials and 4 quadrants is shaded. The accuracy of Radio Galaxy Zoo on the cross-identification task is below the axis and is instead marked by an arrow with the mean accuracy. Note that the pipeline shown in Figure 4.3 is used here, so compact objects are cross-identified in the same way regardless of binary classifier model.

1999 the cross-identification task, methods trained on these cross-identifications still per-
 2000 form comparably to those trained on expert labels. This is because incorrect Radio
 2001 Galaxy Zoo cross-identifications can be thought of as a source of noise in the labels
 2002 which is ‘averaged out’ in training. This shows the usefulness of crowdsourced train-
 2003 ing data, even when the data is noisy.

2004 Our method performs comparably to a nearest neighbours approach. For compact
 2005 objects, this is to be expected—indeed, nearest neighbours attains nearly 100 per cent
 2006 accuracy on the compact test set. Our results do not improve on nearest neighbours
 2007 for resolved objects. However, our method does allow for improvement on nearest
 2008 neighbours with a sufficiently good binary classifier: a ‘perfect’ binary classifier attains
 2009 nearly 100 per cent accuracy on resolved sources. This shows that our method may be
 2010 useful provided that a good binary classifier can be trained. The most obvious place
 2011 for improvement is in feature selection: we use pixels of radio images directly and
 2012 these are likely not conducive to good performance on the candidate classification task.
 2013 Convolutional neural networks, which are able to extract features from images, *should*
 2014 work better, but these require far more training data than the other methods we have
 2015 applied and the small size of ATLAS thus limits their performance.

2016 We noted in Section 4.3.5 that the test set of expert labels, derived from the initial
 2017 ATLAS data release, was less deep than the third data release used by Radio Galaxy
 2018 Zoo and this chapter, introducing a source of label noise in the testing labels. Specifi-
 2019 cally, true host galaxies may be misidentified as non-host galaxies if the associated ra-

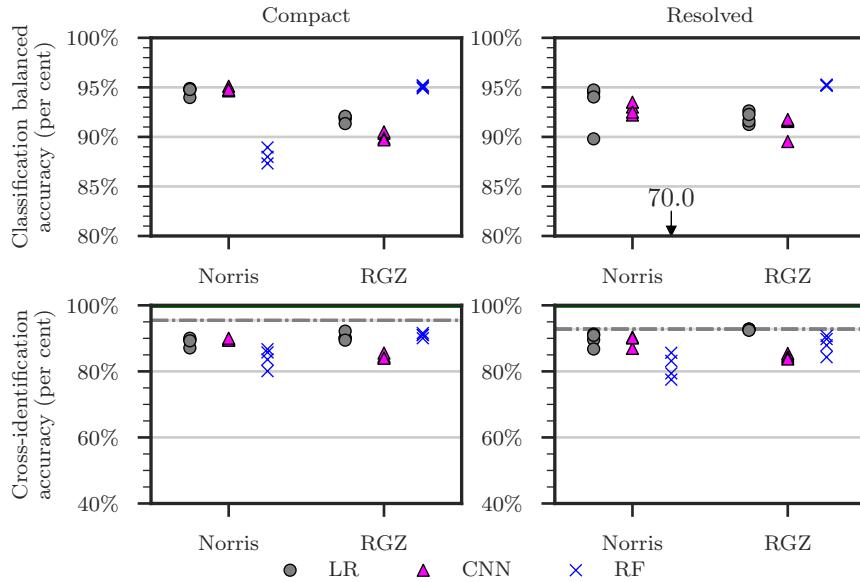


Figure 4.13: Performance of different classification models trained on CDFS and tested on resolved and compact sources in ELAIS-S1. Points represent classification models trained on different quadrants of CDFS, with markers, lines and axes as in Figure 4.11. The balanced accuracy of expert-trained random forest binary classifiers falls below the axis and the corresponding mean accuracy is shown by an arrow. The estimated best attainable accuracy is almost 100 per cent.

2020 dio source was below the 5 signal-to-noise limit in ATLAS DR1 but not in ATLAS DR3.
 2021 This has the effect of reducing the accuracy for Radio Galaxy Zoo-trained classifiers.
 2022 We report the scores predicted by each classifier for each SWIRE object in Ap-
 2023 pendix C and the predicted cross-identification for each ATLAS object in Appendix D.
 2024 Scores reported for a given object were predicted by binary classifiers tested on the
 2025 quadrant containing that object. The reported scores are not weighted.
 2026 In Figure 4.17 we show 5 resolved sources where the most classifiers disagreed on
 2027 the correct cross-identification.

2028 4.4.2 Application to ATLAS-ELAIS-S1

2029 We applied the method trained on CDFS to perform cross-identification on the ELAIS-
 2030 S1 field. Both CDFS and ELAIS-S1 were imaged by the same radio telescope to similar
 2031 sensitivities and angular resolution for the ATLAS survey. We can use the SWIRE
 2032 cross-identifications made by Middelberg et al. (2008) to derive another set of expert
 2033 labels, and hence determine how accurate our method is. If our method generalises
 2034 well across different parts of the sky, then we expect CDFS-trained classifiers to have
 2035 comparable performance between ELAIS-S1 and CDFS. In Figure 4.13 we plot the per-
 2036 formance of CDFS-trained classification models on the candidate classification task

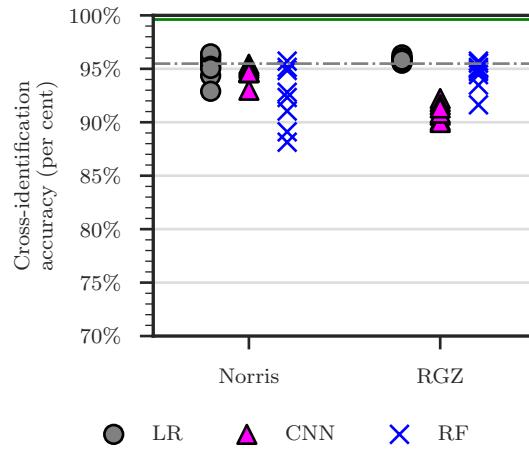


Figure 4.14: Performance of different classifiers trained on CDFS and tested on ELAIS-S1. Markers are as in Figure 4.12 and horizontal lines are as in Figure 4.13. Note that the pipeline shown in Figure 4.3 is used here, so compact objects are cross-identified in the same way regardless of binary classifier model.

2037 and the performance of our method on the cross-identification task using these mod-
 2038 els. We also plot the cross-identification accuracy of a nearest neighbours approach⁶.
 2039 In Figure 4.14 we plot the performance of our method on the full set of ELAIS-S1 AT-
 2040 LAS DR1 radio components using the pipeline in Figure 4.3. We list the corresponding
 2041 accuracies in Appendix B.

2042 Cross-identification results from ELAIS-S1 are similar to those for CDFS, showing
 2043 that our method trained on CDFS performs comparably well on ELAIS-S1. However,
 2044 nearest neighbours outperforms most methods on ELAIS-S1. This is likely because
 2045 there is a much higher percentage of compact objects in ELAIS-S1 than in CDFS. The
 2046 maximum achievable accuracy we have estimated for ELAIS-S1 is very close to 100
 2047 per cent, so (as for CDFS) a very accurate binary classifier would outperform nearest
 2048 neighbours.

2049 One interesting difference between the ATLAS fields is that random forests trained
 2050 on expert labels perform well on CDFS but poorly on ELAIS-S1. This is not the case for
 2051 logistic regression or convolutional neural networks trained on expert labels, nor is it
 2052 the case for random forests trained on Radio Galaxy Zoo. We hypothesise that this is
 2053 because the ELAIS-S1 cross-identification catalogue (Middelberg et al., 2008) labelled
 2054 fainter radio components than the CDFS cross-identification catalogue (Norris et al.,
 2055 2006) due to noise from the very bright source ATCDFS_J032836.53-284156.0 in CDFS.
 2056 Classifiers trained on CDFS expert labels may thus be biased toward brighter radio
 2057 components compared to ELAIS-S1. Radio Galaxy Zoo uses a preliminary version of
 2058 the third data release of ATLAS (Franzen et al., 2015) and so classifiers trained on the

⁶We cannot directly compare our method applied to ELAIS-S1 with Radio Galaxy Zoo, as Radio Galaxy Zoo does not include ELAIS-S1.

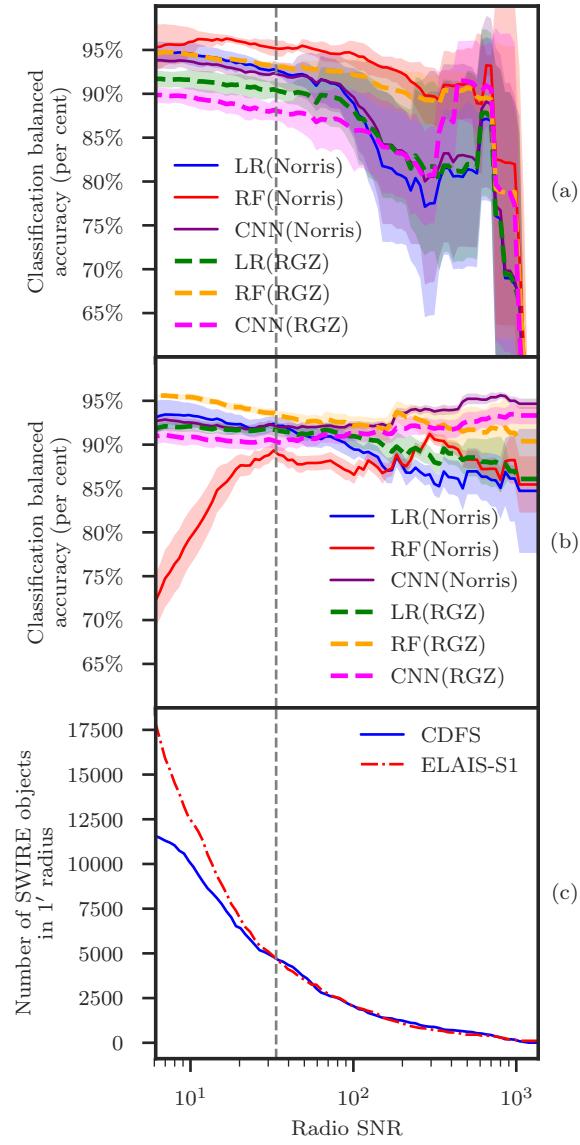


Figure 4.15: (a) Balanced accuracies of classifiers trained and tested on CDFS with different signal-to-noise ratio (SNR) cutoffs for the test set. A SWIRE object is included in the test set if it is within $1'$ of a radio component with greater SNR than the cutoff. Lines of different colour indicate different classifier/training labels combinations, where LR is logistic regression, RF is random forests, CNN is convolutional neural networks, and Norris and RGZ are the expert and Radio Galaxy Zoo label sets respectively. Filled areas represent standard deviations across CDFS quadrants. (b) Balanced accuracies of classifiers trained on CDFS and tested on ELAIS-S1. (c) A cumulative distribution plot of SWIRE objects associated with a radio object with greater SNR than the cutoff. The grey dashed line shows the SNR level at which the number of SWIRE objects above the cutoff is equal for CDFS and ELAIS-S1. This cutoff level is approximately at a SNR of 34.

2059 Radio Galaxy Zoo labels may be less biased toward brighter sources compared to those
 2060 trained on the expert labels. To test this hypothesis we tested each classification model
 2061 against test sets with a signal-to-noise ratio (SNR) cutoff. A SWIRE object was only
 2062 included in the test set for a given cutoff if it was located within $1'$ of a radio component
 2063 with a SNR above the cutoff. The balanced accuracies for each classifier at each cutoff
 2064 are shown in Figure 4.15(a) and (b) and the distribution of test set size for each cutoff
 2065 is shown in Figure 4.15(c). Figure 4.15(c) shows that ELAIS-S1 indeed has more faint
 2066 objects in its test set than the CDFS test set, with the SNR for which the two fields reach
 2067 the same test set size (approximately 34) indicated by the dashed vertical line on each
 2068 plot. For CDFS, all classifiers perform reasonably well across cutoffs, with performance
 2069 dropping as the size of the test set becomes small. For ELAIS-S1, logistic regression and
 2070 convolutional neural networks perform comparably across all SNR cutoffs, but random
 2071 forests do not. While random forests trained on Radio Galaxy Zoo labels perform
 2072 comparably to other classifiers across all SNR cutoffs, random forests trained on expert
 2073 labels show a considerable drop in performance below the dashed line.

2074 4.5 Discussion

2075 Based on the ATLAS sample, our main result is that it is possible to cast radio host
 2076 galaxy cross-identification as a machine learning task for which standard methods can
 2077 be applied. These methods can then be trained with a variety of label sets derived
 2078 from cross-identification catalogues. While our methods have not outperformed near-
 2079 est neighbours, we have demonstrated that for a very accurate binary classifier, good
 2080 cross-identification results can be obtained using our method. Future work could com-
 2081 bine multiple catalogues or physical priors to boost performance.

2082 Nearest neighbours approaches outperform most methods we investigated, no-
 2083 tably including Radio Galaxy Zoo. This is due to the large number of compact or
 2084 partially-resolved objects in ATLAS. This result shows that for compact and partially-
 2085 resolved objects methods that do not use machine learning such as a nearest neigh-
 2086 bours approach or likelihood ratio (Weston et al., 2018) should be preferred to machine
 2087 learning methods. It also shows that ATLAS is not an ideal data set for developing ma-
 2088 chine learning methods like ours. Our use of ATLAS is motivated by its status as a
 2089 pilot survey for EMU, so methods developed for ATLAS should also work for EMU.
 2090 New methods developed should work well with extended radio sources, but this goal
 2091 is almost unsupported by ATLAS as it has very few examples of such sources. This
 2092 makes both training and testing difficult—there are too few extended sources to train
 2093 on and performance on such a small test set may be unreliable. Larger data sets with
 2094 many extended sources like FIRST exist, but these are considerably less deep than and
 2095 at a different resolution to EMU, so there is no reason to expect methods trained on
 2096 such data sets to be applicable to EMU.

2097 The accuracies of our trained cross-identification methods generally fall far below
 2098 the estimated best possible accuracy attainable using our approach, indicated by the
 2099 green-shaded areas in Figures 4.12 and 4.14. The balanced accuracies attained by our

2100 binary classifiers indicate that there is significant room for improvement in classifica-
2101 tion. The classification accuracy could be improved by better model selection and more
2102 training data, particularly for convolutional neural networks. There is a huge variety
2103 of ways to build a convolutional neural network, and we have only investigated one ar-
2104 chitecture. For an exploration of different convolutional neural network architectures
2105 applied to radio astronomy, see Lukic et al. (2018). Convolutional neural networks
2106 generally require more training data than other machine learning models and we have
2107 only trained our networks on a few hundred sources. We would expect performance
2108 on the classification task to greatly increase with larger training sets.

2109 Another problem is that of the window size used to select radio features. Increasing
2110 window size would increase computational expense, but provide more information to
2111 the models. Results are also highly sensitive to how large the window size is compared
2112 to the size of the radio source we are trying to cross-identify, with large angular sizes
2113 requiring large window sizes to ensure that the features contain all the information
2114 needed to localise the host galaxy. An ideal implementation of our method would
2115 most likely represent a galaxy using radio images taken at multiple window sizes, but
2116 this is considerably more expensive.

2117 Larger training sets, better model selection, and larger window sizes would im-
2118 prove performance, but only so far: we would still be bounded above by the estimated
2119 ‘perfect’ classifier accuracy. From this point, the performance can only be improved by
2120 addressing our broken assumptions. We detailed these assumptions in Section 4.3.2,
2121 and we will discuss here how our method could be adapted to avoid these assump-
2122 tions. Our assumption that the host galaxy is contained within the search radius could
2123 be improved by dynamically choosing the search radius, perhaps based on the angu-
2124 lar extent of the radio emission, or the redshift of candidate hosts. Radio morphology
2125 information may allow us to select relevant radio data and hence relax the assump-
2126 tion that a 1'-wide radio image represents just one, whole radio source. Finally, our
2127 assumption that the host galaxy is detected in infrared is technically not needed, as
2128 the sliding-window approach we have employed will still work even if there are no de-
2129 tected host galaxies—instead of classifying candidate hosts, simply classify each pixel
2130 in the radio image. The downside of removing candidate hosts is that we are no longer
2131 able to reliably incorporate host galaxy information such as colour and redshift, though
2132 this could be resolved by treating pixels as potentially undetected candidate hosts with
2133 noisy features.

2134 We observe that Radio Galaxy Zoo-trained methods perform comparably to meth-
2135 ods trained on expert labels. This shows that the crowdsourced labels from Radio
2136 Galaxy Zoo will provide a valuable source of training data for future machine learn-
2137 ing methods in radio astronomy.

2138 Compared to nearest neighbours, cross-identification accuracy on ELAIS-S1 is lower
2139 than on CDFS. Particularly notable is that our performance on compact objects is very
2140 low for ELAIS-S1, while it was near-optimal for CDFS. These differences may be for
2141 a number of reasons. ELAIS-S1 has beam size and noise profile different from CDFS
2142 (even though both were imaged with the same telescope), so it is possible that our
2143 methods over-adapted to the beam and noise of CDFS. Additionally, CDFS contains a

very bright source which may have caused artefacts throughout the field that are not present in ELAIS-S1. Further work is required to understand the differences between the fields and their effect on performance.

Figure 4.15 reveals interesting behaviour of different classifier models at different flux cutoffs. Logistic regression and convolutional neural networks seem relatively independent of flux, with these models performing well on the fainter ELAIS-S1 components even when they were trained on the generally brighter components in CDFS. Conversely, random forests were sensitive to the changes in flux distribution between datasets. This shows that not all models behave similarly on radio data, and it is therefore important to investigate multiple models when developing machine learning methods for radio astronomy.

Appendix E (see Figure 4.17) shows examples of incorrectly cross-identified components in CDFS. On no such component do all classifiers agree. This raises the possibility of using the level of disagreement of an ensemble of binary classifiers as a measure of the difficulty of cross-identifying a radio component, analogous to the consensus level for Radio Galaxy Zoo volunteers.

Our methods can be easily incorporated into other cross-identification methods or used as an extra data source for source detection. For example, the scores output by our binary classifiers could be used to disambiguate between candidate host galaxies selected by model-based algorithms, or used to weight candidate host galaxies while a source detector attempts to associate radio components. Our method can also be extended using other data sources: for example, information from source identification algorithms could be incorporated into the feature set of candidate host galaxies.

4.6 Summary

We presented a machine learning approach for cross-identification of radio components with their corresponding infrared host galaxy. Using the CDFS field of ATLAS as a training set we trained our methods on expert and crowdsourced cross-identification catalogues. Applying these methods on both fields of ATLAS, we found that:

- Our method trained on ATLAS observations of CDFS generalised to ATLAS observations of ELAIS-S1, demonstrating that training on a single patch of sky is a feasible option for training machine learning methods for wide-area radio surveys;
- Performance was comparable to nearest neighbours even on resolved sources, showing that nearest neighbours is useful for datasets consisting mostly of unresolved sources such as ATLAS and EMU;
- Radio Galaxy Zoo-trained models performed comparably to expert-trained models and outperformed Radio Galaxy Zoo, showing that crowdsourced labels are useful for training machine learning methods for cross-identification even when these labels are noisy;

- 2183 • ATLAS does not contain sufficient data to train or test machine learning cross-
2184 identification methods for extended radio sources. This suggests that if machine
2185 learning methods are to be used on EMU, a larger area of sky will be required
2186 for training and testing these methods. However, existing surveys like FIRST are
2187 likely too different from EMU to expect good generalisation.

2188 While our cross-identification performance is not as high as desired, we make no
2189 assumptions on the binary classification model used in our methods and so we expect
2190 the performance to be improved by further experimentation and model selection. Our
2191 method provides a useful framework for generalising cross-identification catalogues
2192 to other areas of the sky from the same radio survey and can be incorporated into
2193 existing methods. We have shown that citizen science can provide a useful dataset for
2194 training machine learning methods in the radio domain. Chapter 5 will extend this
2195 approach and confirm that dataset size is a key limitation by successfully applying it
2196 to a considerably larger dataset: FIRST.

2197 4.7 Acknowledgements

2198 This chapter and corresponding publication was made possible by the participation of
2199 more than 11 000 volunteers in the Radio Galaxy Zoo project. Their contributions are
2200 individually acknowledged at <http://rgzauthors.galaxyzoo.org>. Parts of this research
2201 were conducted by the Australian Research Council Centre of Excellence for All-sky
2202 Astrophysics (CAASTRO), through project number CE110001020. Partial support for
2203 LR was provided by U.S. National Science Foundation grants AST1211595 and 1714205
2204 to the University of Minnesota. HA benefitted from grant 980/2016-2017 of Universi-
2205 dad de Guanajuato. We thank A. Tran and the [reviewer anonymous referee](#) for their
2206 comments on the manuscript this chapter was based on. Radio Galaxy Zoo makes
2207 use of data products from the Wide-field Infrared Survey Explorer and the Very Large
2208 Array. The Wide-field Infrared Survey Explorer is a joint project of the University of
2209 California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Tech-
2210 nology, funded by the National Aeronautics and Space Administration. The National
2211 Radio Astronomy Observatory is a facility of the National Science Foundation oper-
2212 ated under cooperative agreement by Associated Universities, Inc. The figures in this
2213 work made use of Astropy, a community-developed core Python package for Astron-
2214 omy (The Astropy Collaboration et al., 2018). The Australia Telescope Compact Array
2215 is part of the Australia Telescope, which is funded by the Commonwealth of Australia
2216 for operation as a National Facility managed by CSIRO. [We acknowledge the Gomeroi](#)
2217 [people as the traditional custodians of the Observatory site.](#)

2218 A [Classification models](#)

2219 [This appendix describes the three different models we used for binary classification in](#)
2220 [this chapter: logistic regression, convolutional neural networks, and random forests.](#)

2221 **A.1 Logistic Regression**

2222 Logistic regression is linear in the feature space and outputs the probability that the
2223 input has a positive label. The model is (Bishop, 2006):

$$f(\vec{x}) = \sigma(\vec{w}^T \vec{x} + b) , \quad (4.2)$$

where $\vec{w} \in \mathbb{R}^D$ is a vector of parameters, $b \in \mathbb{R}$ is a bias term, $\vec{x} \in \mathbb{R}^D$ is the feature vector representation of a candidate host, and $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is the logistic sigmoid function:

$$\sigma(a) = (1 + \exp(-a))^{-1} . \quad (4.3)$$

2224 The logistic regression model is fully differentiable, and the parameters \vec{w} can therefore
2225 be learned using gradient-based optimisation methods. We used the scikit-learn
2226 (Pedregosa et al., 2011) implementation of logistic regression with balanced classes.

2227 **A.2 Convolutional neural networks**

2228 Convolutional neural networks (CNN) are a biologically-inspired prediction model
2229 for prediction with image inputs. The input image is convolved with a number of
2230 filters to produce output images called feature maps. These feature maps can then
2231 be convolved again with other filters on subsequent layers, producing a network of
2232 convolutions. The whole network is differentiable with respect to the values of the
2233 filters and the filters can be learned using gradient-based optimisation methods. The
2234 final layer of the network is logistic regression, with the convolved outputs as input
2235 features. For more detail, see subsection II.A, LeCun et al. (1998). We used KERAS
2236 (Chollet et al., 2015) to implement our CNN, accounting for class imbalance by reweighting
2237 the classes.

2238 CNNs have recently produced good results on large image-based datasets in astronomy
2239 (Dieleman et al., 2015; Lukic et al., 2018, e.g.). We employed only a simple CNN model
2240 in Chapter 4 as a proof of concept that CNNs may be used for class probability prediction
2241 on radio images. The model architecture we used is shown in Figure 4.16.

2242 **A.3 Random Forests**

2243 Random forests are an ensemble of decision trees (Breiman, 2001). They consider
2244 multiple subsamples of the training set, where each subsample is sampled with replacement
2245 from the training set. For each subsample a decision tree classifier is constructed by
2246 repeatedly making axis-parallel splits based on individual features. In a random forest
2247 the split decision is taken based on a random subset of features. To classify a new data
2248 point, the random forest takes the weighted average of all classifications produced by
2249 each decision tree. In Chapter 4 we used the scikit-learn (Pedregosa et al., 2011)
2250 implementation of random forests with 10 trees, the information entropy split criterion,
2251 a minimum leaf size of 45, and balanced classes.

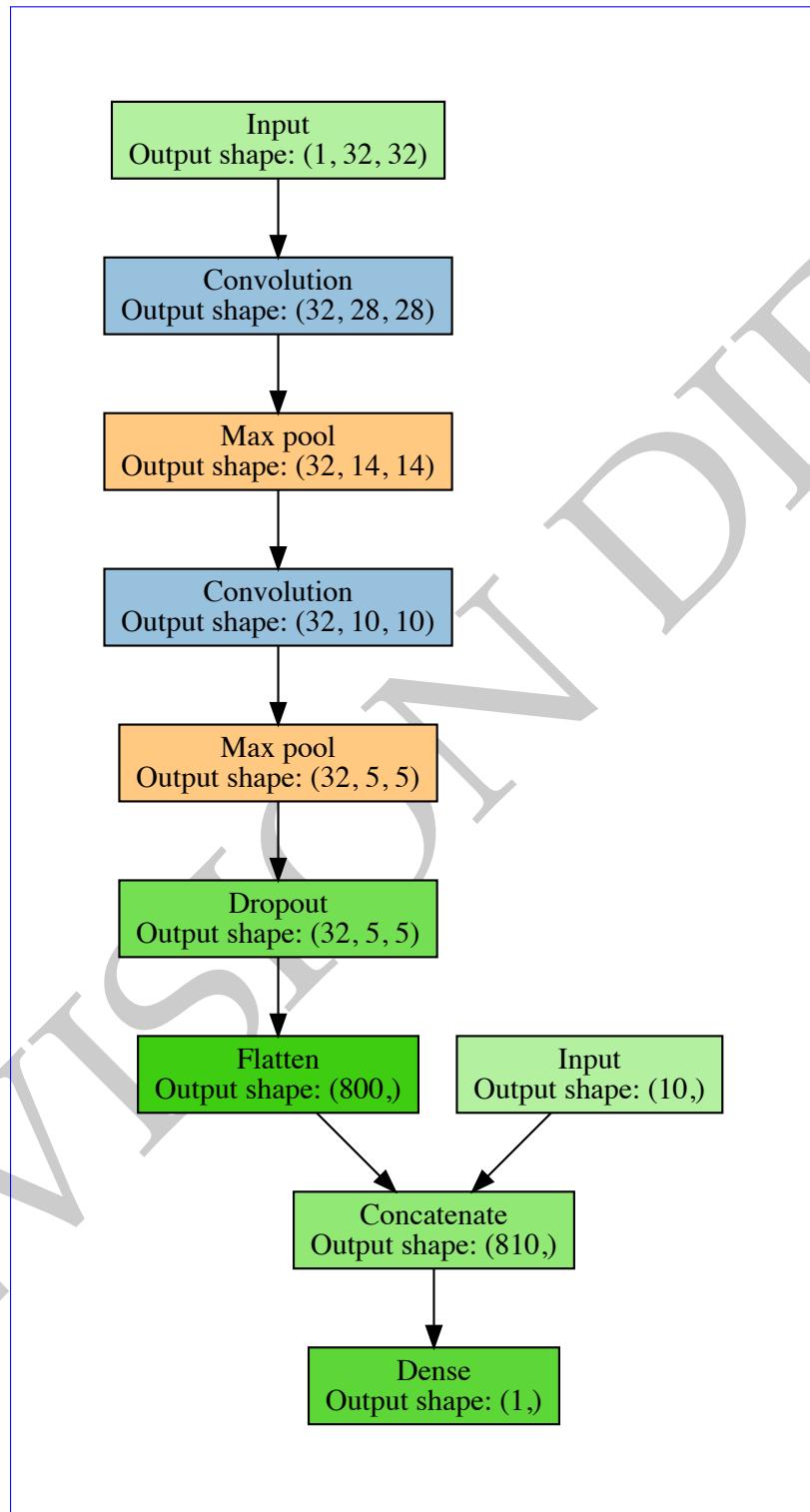


Figure 4.16: Architecture of our CNN. Parenthesised numbers indicate the size of output layers as a tuple (width, height, depth). The concatenate layer flattens the output of the previous layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux ratios, stellarity indices, and distance. The dropout layer randomly sets 25% of its inputs to zero during training to prevent overfitting. Diagram based on <https://github.com/dnouri/nolearn>.

Table 4.3: Balanced accuracies for different binary classification models trained and tested on SWIRE objects in CDFS. The ‘Labeller’ column states what set of training labels were used to train the classifier, and the ‘Classifier’ column states what classification model was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression and ‘RF’ is random forests. Accuracies are evaluated against the expert label set derived from Norris et al. (2006). The standard deviation of balanced accuracies evaluated across the four quadrants of CDFS (Figure 4.8) is also shown. The ‘compact’ set refers to SWIRE objects within $1'$ of a compact radio component, the ‘resolved’ set refers to SWIRE objects within $1'$ of a resolved radio component, and ‘all’ is the union of these sets.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accuracy (per cent)
Norris	LR	91.5 ± 1.0	93.2 ± 1.0	93.0 ± 1.2
	CNN	92.6 ± 0.7	91.2 ± 0.5	92.0 ± 0.6
	RF	96.7 ± 1.5	91.0 ± 4.5	96.0 ± 2.5
RGZ	LR	89.5 ± 0.8	90.5 ± 1.7	90.2 ± 0.8
	CNN	89.4 ± 0.6	89.6 ± 1.3	89.4 ± 0.5
	RF	94.5 ± 0.2	95.8 ± 0.4	94.7 ± 0.3

2252 B Accuracy tables

2253 This section contains tables of accuracy for our cross-identification method applied
 2254 to CDFS and ELAIS-S1. In Table 4.3 and Table 4.4 we list the balanced accuracies
 2255 of our Chapter 4 classifiers on the cross-identification task for CDFS and ELAIS-S1
 2256 respectively, averaged over each set of training quadrants. In Table 4.5 and Table 4.6
 2257 we list the balanced accuracies of classifiers on the cross-identification task for CDFS
 2258 and ELAIS-S1 respectively, averaged over each set of training quadrants.

2259 C SWIRE object scores

2260 This appendix contains scores predicted by our Chapter 4 binary classifiers for each
 2261 SWIRE object within $1'$ of a radio component in CDFS and ELAIS-S1. Scores for SWIRE CDFS
 2262 objects are shown in Table 4.7 and scores for SWIRE ELAIS-S1 are shown in Table 4.8.
 2263 For CDFS, the score for an object in a quadrant is predicted by binary classifiers trained
 2264 on all other quadrants. For ELAIS-S1, we show the scores predicted by binary classifiers
 2265 trained on each CDFS quadrant. Note that these scores have *not* been weighted by
 2266 Gaussians. These are partial tables, and the full tables are available online at the
 2267 *Monthly Notices of the Royal Astronomical Society* website⁷.

2268 The columns of the score tables are defined as follows:

- 2269 • *SWIRE*—SWIRE designation for candidate host galaxy.
- 2270 • *RA*—Right ascension (J2000).

⁷<https://doi.org/10.1093/mnras/sty1308>

Table 4.4: Balanced accuracies for different binary classification models trained on SWIRE objects in CDFS and tested on SWIRE objects in ELAIS-S1. Columns and abbreviations are as in Table 4.3. Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008). The standard deviations of balanced accuracies of models trained on the four subsets of CDFS (Figure 4.8) are also shown.

Labeller	Classifier	Mean 'Compact' accuracy (per cent)	Mean 'Resolved' accuracy (per cent)	Mean 'All' accuracy (per cent)
Norris	LR	94.6 ± 0.4	93.3 ± 2.0	95.3 ± 0.1
	CNN	94.8 ± 0.2	92.8 ± 0.5	94.4 ± 0.2
	RF	85.9 ± 3.8	70.0 ± 2.8	86.6 ± 3.2
RGZ	LR	91.8 ± 0.3	91.9 ± 0.5	92.0 ± 0.2
	CNN	90.1 ± 0.3	91.1 ± 0.9	90.2 ± 0.3
	RF	95.1 ± 0.1	95.2 ± 0.0	95.2 ± 0.3

Table 4.5: Cross-identification accuracies for different classification models on CDFS. The 'Labeller' column states what set of training labels were used to train the method, and the 'Classifier' column states what classification model was used. 'CNN' is a convolutional neural network, 'LR' is logistic regression, 'RF' is random forests, and 'Labels' is the accuracy of the label set itself. 'Perfect' indicates that the true labels of the test set were used and hence represents an upper bound on cross-identification accuracy with our method. 'NN' is a nearest neighbours approach. Accuracies are evaluated against the expert label set, so 'Norris' labels are 100 per cent accurate by definition. The standard deviation of accuracies evaluated across the four quadrants of CDFS (Figure 4.8) is also shown.

Labeller	Classifier	Mean 'Compact' accuracy (per cent)	Mean 'Resolved' accuracy (per cent)	Mean 'All' accuracy (per cent)
—	NN	97.2 ± 1.7	75.7 ± 7.9	93.4 ± 0.8
	Random	97.9 ± 2.2	22.3 ± 9.2	83.2 ± 4.7
	Labels	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	Perfect	97.9 ± 2.2	99.0 ± 1.8	98.1 ± 1.7
Norris	LR	97.3 ± 0.5	76.0 ± 3.2	93.7 ± 1.8
	CNN	96.6 ± 0.9	74.3 ± 12.3	93.5 ± 0.5
	RF	96.1 ± 1.4	75.8 ± 6.7	93.8 ± 2.0
	Labels	53.1 ± 8.5	56.7 ± 5.9	54.4 ± 5.9
	LR	97.3 ± 1.9	74.5 ± 5.1	93.6 ± 1.7
	CNN	85.4 ± 2.6	68.1 ± 9.2	92.4 ± 1.1
	RF	97.5 ± 0.9	74.3 ± 7.9	93.7 ± 1.5
	—	—	—	—
RGZ	NN	97.2 ± 1.7	75.7 ± 7.9	93.4 ± 0.8
	Random	97.9 ± 2.2	22.3 ± 9.2	83.2 ± 4.7
	Labels	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	Perfect	97.9 ± 2.2	99.0 ± 1.8	98.1 ± 1.7

Table 4.6: Cross-identification accuracies for different classification models on ELAIS-S1. Columns and abbreviations are as in Table 4.5. Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008) cross-identifications. The standard deviation of accuracies evaluated across models trained on the four quadrants of CDFS (Figure 4.8) is also shown.

Labeller	Classifier	Mean ‘Compact’ accuracy (per cent)	Mean ‘Resolved’ accuracy (per cent)	Mean ‘All’ accuracy (per cent)
Middelberg	NN	95.5 ± 0.0	92.8 ± 0.0	95.5 ± 0.0
	Random	61.9 ± 1.1	26.6 ± 2.1	61.9 ± 1.1
	Perfect	99.6 ± 0.0	99.8 ± 0.0	99.6 ± 0.0
	LR	89.0 ± 1.1	89.7 ± 1.8	94.4 ± 0.9
	CNN	89.7 ± 0.3	89.4 ± 1.4	94.3 ± 0.7
	RF	83.8 ± 5.6	82.3 ± 4.1	90.6 ± 2.1
Norris	LR	90.5 ± 1.0	92.7 ± 0.2	95.9 ± 0.1
	CNN	84.6 ± 0.6	84.6 ± 0.6	91.8 ± 0.3
	RF	91.3 ± 1.0	90.3 ± 2.4	94.7 ± 1.2
RGZ	LR			
	CNN			
	RF			

- *Dec*—Declination (J2000).
- *Expert host*—Whether the candidate host galaxy is a host galaxy according to Norris et al. (2006) or Middelberg et al. (2008) cross-identifications of CDFS and ELAIS-S1 respectively.
- *RGZ host*—Whether the candidate host galaxy is a host galaxy according to Radio Galaxy Zoo cross-identifications (Wong et al., in prep.). This is always ‘no’ for ELAIS-S1 objects.
- *C/L/D*—Score assigned by binary classifier *C* trained on label set *L* of *D* candidate host galaxies. *C* may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or random forests respectively. *L* may be ‘Norris’ or ‘RGZ’ for expert and Radio Galaxy Zoo labels respectively. *D* may be ‘All’, ‘Compact’ or ‘Resolved’ for each respective subset defined in Section 4.3.6.

D ATLAS component cross-identifications

This section contains cross-identifications predicted by our cross-identification method for each ATLAS radio component in CDFS and ELAIS-S1. Cross-identifications for ATLAS CDFS components are shown in Table 4.9 and cross-identifications for ATLAS ELAIS-S1 are shown in Table 4.10. For CDFS, the cross-identification for a component in a quadrant is predicted using our method with binary classifiers trained on all other quadrants. For ELAIS-S1, we show the cross-identifications predicted by our method using binary classifiers trained on each CDFS quadrant. For CDFS, we also show the Radio Galaxy Zoo consensus, which is a proxy for the difficulty of cross-identifying a component

Table 4.7: Scores output by our trained classifiers for SWIRE CDFS candidate host galaxies. Columns are defined in Appendix C. Full table electronic.

Table 4.8: Scores output by our trained classifiers for SWIRE-ELAIS-S1 candidate host galaxies. Columns are defined in Appendix C. Full table electronic.

SWIRE	RA	Dec	Expert host	RGZ host	CNN			RGZ		
					All	Compact	Norris	All	Compact	Resolved
J002925.73-440256.2	7.3572	-44.0490	yes	no	0.9537	0.8638	0.5552	0.9195	0.9037	0.9371
J002926.14-440249.0	7.3590	-44.0470	no	no	0.7361	0.8752	0.5640	0.7740	0.7474	0.7952
J002926.52-440247.0	7.3605	-44.0464	no	no	0.3390	0.8338	0.5556	0.7275	0.6894	0.7197
J002926.63-440301.1	7.3610	-44.0503	no	no	0.2108	0.8251	0.5623	0.3434	0.3306	0.3292
J002927.13-440232.6	7.3631	-44.0424	no	no	0.0339	0.8479	0.5669	0.5853	0.5148	0.5159
J002927.28-440245.3	7.3637	-44.0459	no	no	0.0406	0.8345	0.5540	0.2702	0.2340	0.2133
J002927.44-440238.5	7.3644	-44.0440	no	no	0.0116	0.8267	0.5746	0.2228	0.2182	0.2028
J002928.08-440230.3	7.3670	-44.0418	no	no	0.0024	0.8626	0.5791	0.2297	0.1963	0.1549
J002928.11-440312.7	7.3671	-44.0535	no	no	0.0011	0.8159	0.5514	0.0377	0.0384	0.0271
J002928.80-440306.8	7.3700	-44.0519	no	no	0.0003	0.8405	0.5668	0.0236	0.0226	0.0136
LR				RF				RF		
All	Compact	Resolved	All	Compact	Resolved	All	Compact	All	Compact	Resolved
0.9722	0.9955	0.8769	0.9933	0.9934	0.9658	0.8824	0.9664	0.7950	0.8078	0.9227
0.4669	0.0111	0.4249	0.3926	0.2220	0.5947	0.2077	0.0000	0.1613	0.1876	0.0852
0.2264	0.0254	0.2389	0.6275	0.3033	0.6812	0.1347	0.0857	0.0399	0.3582	0.4854
0.0603	0.0007	0.0734	0.0688	0.0141	0.1581	0.0917	0.0000	0.0399	0.2846	0.1245
0.0248	0.0334	0.0301	0.5735	0.5065	0.5265	0.1977	0.1507	0.0000	0.3334	0.6593
0.0173	0.0016	0.0359	0.1056	0.0492	0.1456	0.0000	0.0000	0.0000	0.0000	0.3995
0.0064	0.0049	0.0187	0.1981	0.1534	0.1493	0.0000	0.0000	0.0000	0.0000	0.0287
0.0020	0.0005	0.0239	0.1337	0.1001	0.1310	0.0000	0.0000	0.0358	0.0000	0.1634
0.0008	0.0013	0.0280	0.0361	0.0205	0.1171	0.0000	0.0000	0.0000	0.0000	0.1284
0.0004	0.0014	0.0095	0.0339	0.0408	0.0136	0.0000	0.0000	0.0000	0.0000	0.1584

2292 (Wong et al., in prep.). These are partial tables, and the full tables are available online
2293 at the *Monthly Notices of the Royal Astronomical Society* website⁸.

2294 The columns of the cross-identification tables are defined as follows:

- 2295 • *ATLAS*—ATLAS designation for radio component.
- 2296 • *RA*—Right ascension of radio component (J2000).
- 2297 • *Dec*—Declination of radio component (J2000).
- 2298 • *CID*—Radio Galaxy Zoo component ID.
- 2299 • *Zooniverse ID*—Radio Galaxy Zoo Zooniverse ID.
- 2300 • *Norris/Middleberg*—Designation of SWIRE cross-identification from Norris et al. (2006)
2301 or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- 2302 • *Norris/Middleberg RA*—Right ascension of SWIRE cross-identification from Norris et al. (2006)
2303 or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- 2304 • *Norris/Middleberg Dec*—Right ascension of SWIRE cross-identification from Norris et al. (2006)
2305 or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.
- 2306 • *RGZ*—Designation of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.)
2307 :
2308 • *RGZ RA*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo
2309 (Wong et al., in prep.).
- 2310 • *RGZ Dec*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo
2311 (Wong et al., in prep.).
- 2312 • *RGZ radio consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on
2313 the radio component configuration.
- 2314 • *RGZ IR consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on the
2315 host galaxy of this radio component.
- 2316 • *C / L / D*—Designation of SWIRE cross-identification made by our method using
2317 classification model *C* trained on label set *L* of *D* candidate host galaxies. *C* may
2318 be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or random forests respectively.
2319 *L* may be ‘Norris’ or ‘RGZ’ for expert and Radio Galaxy Zoo labels respectively.
2320 *D* may be ‘All’, ‘Compact’ or ‘Resolved’ for each respective subset defined in Sec-
2321 tion 4.3.6.
- 2322 • *C / L / D RA*—Right ascension (J2000) of SWIRE cross-identification made by
2323 our method using classification model *C* trained on label set *L* of *D* candidate
2324 host galaxies. *C*, *L* and *D* are defined as for designation.

⁸<https://doi.org/10.1093/mnras/sty1308>

- 2325 • C/L/D Dec—Declination (J2000) of SWIRE cross-identification made by our
2326 method using classification model C trained on label set L of D candidate host
2327 galaxies. C, L and D are defined as for designation.

2328 **E Cross-identification figures**

2329 Figure 4.17 shows figures of our cross-identifications of each ATLAS radio component
2330 in CDFS and ELAIS-S1. There are just five examples shown here, but all 469 examples
2331 are available online at the *Monthly Notices of the Royal Astronomical Society* website⁹.

⁹<https://doi.org/10.1093/mnras/sty1308>

Table 4.9: Cross-identifications for ATLAS CDFS components. Columns are defined in Appendix D. Full table electronic.

ATLAS	RA	Dec	CJD	Zoomwise ID	Norris			RGZ			RCZ			
					RA	Dec	RA	Dec	RA	Dec	radio	IR	consensus	
J032602.82-284708.1C	51.5117	-28.7826	C10412	ARC0003hb2	032602.13.15-284708.5	51.5132	-28.7837	032602.36-284711.5	51.5098	-28.7865	032602.36-284711.5	51.5098	-28.7865	
J032615.49-284629.4C	51.5646	-28.7749	C10614	ARC0005hf	032615.41-284630.7	51.5642	-28.7752	032615.41-284630.7	51.5642	-28.7752	032615.41-284630.7	51.5642	-28.7752	
J032615.55-280559.8C	51.5648	-28.1000	C10320	ARC0003s8s	032615.52-280559.8	51.5647	-28.1000	032615.52-280559.8	51.5647	-28.1000	032615.52-280559.8	51.5647	-28.1000	
J032617.35-280710.2C	51.5723	-28.1195	C102617	ARC0003s2	032617.39-280707.2	51.5746	-28.1187	032617.89-280707.2	51.5746	-28.1187	032617.89-280707.2	51.5746	-28.1187	
J032615.19-280909.8C	51.6047	-28.1952	C10409	ARC0003az	032615.19-280909.1	51.6050	-28.1958	032615.19-280909.1	51.6050	-28.1958	032615.19-280909.1	51.6050	-28.1958	
J032629.10-280650.1C	51.6213	-28.1139	C10963	ARC0003o4	032629.9.13-280650.7	51.6214	-28.1141	032626.74-280636.7	51.6114	-28.1102	032626.74-280636.7	51.6114	-28.1102	
J032629.61-284052.7C	51.6234	-28.6813	C10304	ARC0003o8e	032629.9.54-284055.8	51.6231	-28.6822	032629.54-284055.8	51.6231	-28.6822	032629.54-284055.8	51.6231	-28.6822	
J032629.92-284753.5C	51.6247	-28.7982	C10120	ARC0003s3o	032629.9.81-284754.4	51.6232	-28.7985	032629.81-284754.4	51.6242	-28.7985	032629.81-284754.4	51.6242	-28.7985	
J032630.66-283657.3C	51.6278	-28.6159	C10172	ARC0003s5s	032630.64-283658.0	51.6277	-28.6161	032628.56-283744.8	51.6119	-28.6291	032628.56-283744.8	51.6119	-28.6291	
J032634.59-282022.8C	51.6441	-28.3397	C10757	ARC0006o2	032634.58-282022.8	51.6441	-28.3397	032631.96-281941.0	51.6332	-28.281	032631.96-281941.0	51.6332	-28.281	
CNN														
Compact	RA	Dec	Resolved	RA	Dec	Resolved	RA	Dec	Resolved	RA	Dec	Resolved	RA	
J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865
J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752
J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000
J032617.69-280707.2	51.5746	-28.1187	J032617.89-280707.2	51.5746	-28.1187	J032617.89-280707.2	51.5746	-28.1187	J032617.89-280707.2	51.5746	-28.1187	J032617.89-280707.2	51.5746	-28.1187
J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032624.50-280905.9	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528
J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141
J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811
J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985
J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161
J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397
LR														
Compact	RA	Dec	Resolved	RA	Dec	Resolved	RA	Dec	Resolved	RA	Dec	Resolved	RA	
J032604.58-284650.9	51.5191	-28.7808	J032602.08-284713.1	51.5087	-28.7887	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865
J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752
J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000
J032615.86-280288.8	51.5961	-28.0860	J032615.16-280742.2	51.5932	-28.0860	J032615.16-280742.2	51.5932	-28.0860	J032615.16-280742.2	51.5932	-28.0860	J032615.16-280742.2	51.5932	-28.0860
J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528
J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141
J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811
J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985
J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161	J032630.64-283658.0	51.6677	-28.6161
J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397	J032634.58-282022.8	51.6441	-28.3397
RF														
Compact	RA	Dec	Resolved	RA	Dec	Resolved	RA	Dec	Resolved	RA	Dec	Resolved	RA	
J032603.15-284708.5	51.5132	-28.7857	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865	J032602.36-284711.5	51.5098	-28.7865
J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752	J032615.41-284630.7	51.5642	-28.7752
J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000	J032615.52-280559.8	51.5647	-28.1000
J032617.89-280707.2	51.5946	-28.1000	J032617.89-280707.2	51.5946	-28.1000	J032617.89-280707.2	51.5946	-28.1000	J032617.89-280707.2	51.5946	-28.1000	J032617.89-280707.2	51.5946	-28.1000
J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528	J032625.19-280910.1	51.6050	-28.1528
J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141	J032629.13-280650.7	51.6214	-28.1141
J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811	J032629.54-284051.9	51.6231	-28.6811
J032630.12-284751.2	51.6255	-28.7976	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985	J032629.81-284754.4	51.6242	-28.7985
J032630.64-283658.0	51.6277	-28.6161	J032630.64-283658.0	51.6277	-28.6161	J03263								

Table 4.10: Cross-identifications for ATLAS ELAIS-S1 components. Columns are defined in Appendix D. Full table electronic.

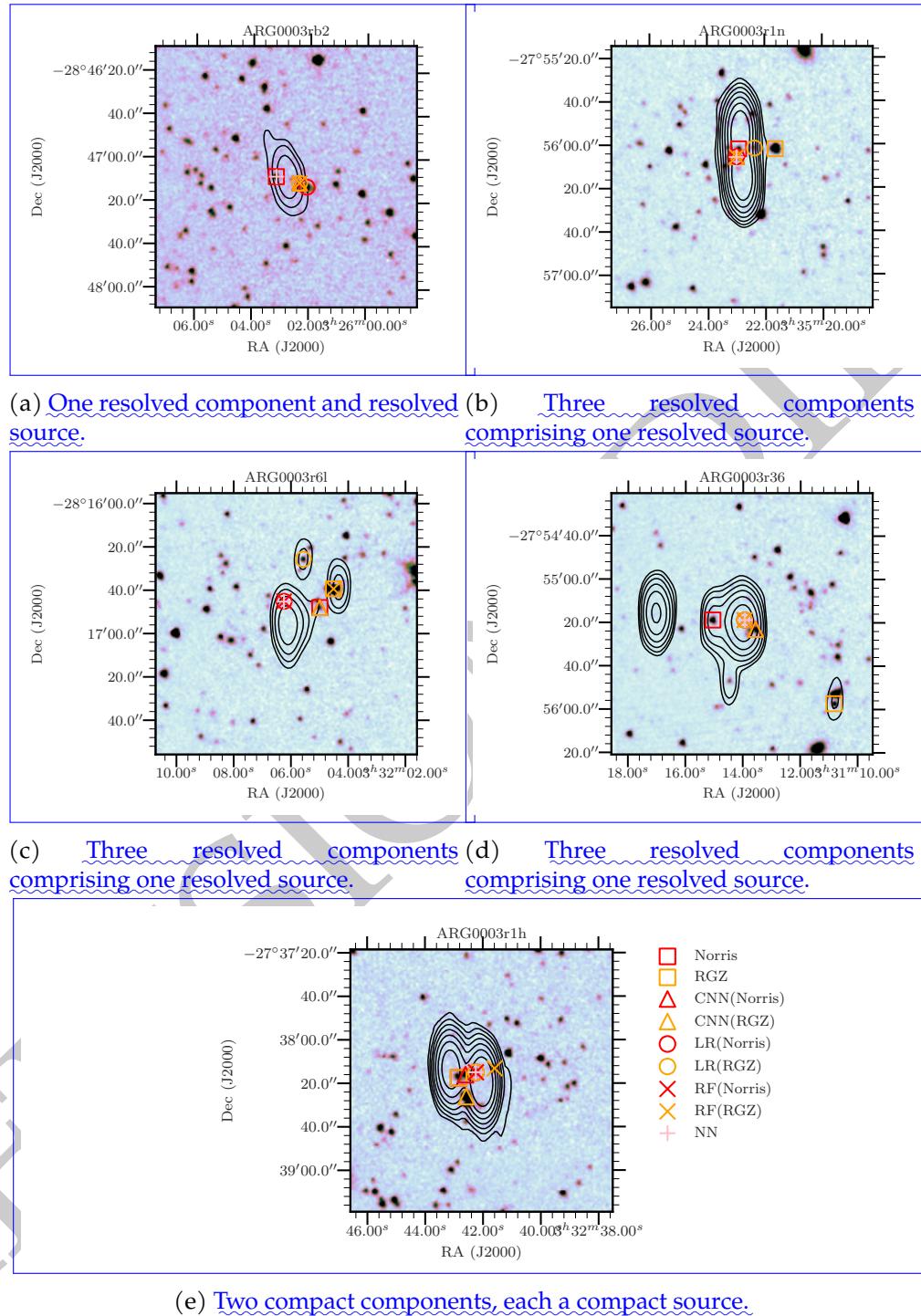


Figure 4.17: Examples of resolved sources with high disagreement between cross-identifiers. The contours show ATLAS radio data and start at 4σ , increasing geometrically by a factor of 2. The background image is the $3.6 \mu\text{m}$ SWIRE image. Binary classifier model/training set combinations are denoted $C(S)$ where C is the binary classifier model and S is the training set. ‘LR’ is logistic regression, ‘CNN’ is convolutional neural networks, and ‘RF’ is random forests. ‘Norris’ refers to the expert labels and ‘RGZ’ refers to the Radio Galaxy Zoo labels. The cross-identification made by nearest neighbours is shown by ‘NN’.

REVISION DIFF

2333

Radio Luminosity Functions

2334 This chapter is based on my paper *Radio Galaxy Zoo: Radio Luminosity Functions of Ex-
2335 tended Sources*, by M. J. Alger, O. I. Wong, C. S. Ong, N. M. McClure-Griffiths, H. An-
2336 dernach, L. Rudnick, S. S. Shabala, A. F. Garon, J. K. Banfield, A. D. Kapińska, R. P.
2337 Norris, and A. J. M. Thomson; to be submitted. Some minor changes have been made
2338 to match the rest of this thesis.

2339

2340 In Chapter 4, we concluded that our machine learning cross-identification approach
2341 was sound in principle, but the small size and lack of morphological complexity in AT-
2342 LAS prevented training a good classifier. To demonstrate this, we turned our attention
2343 to the Faint Images of the Radio Sky at Twenty Centimeters (FIRST) survey, which con-
2344 tained many more sources and many more morphologically complex sources due to
2345 its higher resolution and sample size. If cross-identification of this dataset was success-
2346 ful, we could use the resulting cross-identification catalogue to investigate statistical
2347 properties of radio galaxies at larger scales than previously possible. The object of in-
2348 terest would be the radio luminosity function (RLF), a statistical description of how
2349 common radio galaxies of various luminosities are throughout the universe. The RLF
2350 would be specific to only extended radio galaxies, and it is this class of galaxies that we
2351 would expect to maximally impact its local environment through direct, mechanical
2352 energy input.

2353 In this chapter, we use a machine learning model trained on the Radio Galaxy Zoo
2354 citizen science project to cross-identify 244 846 radio components of extended radio
2355 sources from FIRST with 158 337 host galaxies observed in the mid-infrared from the
2356 AllWISE source catalogue. 34 305 hosts have a spectroscopic redshift in the Sloan Digi-
2357 tal Sky Survey (SDSS). This is the largest available sample of cross-identified extended
2358 radio sources to date at the time of writing. The large sample size allows us to estimate
2359 the RLF of extended radio galaxies and to further characterise the sample with frac-
2360 tional radio luminosity functions. The extended source RLF can be used to estimate the
2361 mechanical energy injected into the local environment of active galactic nuclei (AGN).
2362 We estimate from the observed radio emission that inferred extended radio jets from
2363 AGN contribute between 1.3×10^{30} and 1.2×10^{32} W Mpc $^{-3}$ of mechanical energy to
2364 their environment throughout the low-redshift ($z < 0.6$) Universe. By further visual
2365 verification we also find 40 radio galaxies with projected size larger than 1 Mpc. Our re-

2366 sults directly demonstrate the impact that automated machine learning methods could
 2367 have on future wide-area radio surveys.

2368 **1 Introduction**

2369 Radio active galactic nuclei (AGN) are some of the most violent and energetic objects
 2370 in the Universe. Supermassive black holes at the centre of galaxies can produce large,
 2371 luminous radio jets that may feed lobes extending up to megaparsec scales. As a vital
 2372 part of galaxy evolution, we want to understand how radio AGN interact with their
 2373 host galaxies, in what kinds of galaxies these AGN are found, and how these popula-
 2374 tions change over cosmic time.

2375 Mechanical energy from AGN is needed to explain the star-formation history of
 2376 galaxies throughout the Universe (e.g. Hardcastle & Croston, 2020; Raouf et al., 2017).
 2377 The radio luminosity function (RLF) of extended radio sources would characterise the
 2378 energy output of galaxies capable of significant energy impact into their local environ-
 2379 ment. The local RLF of primarily compact sources is estimated and described in detail
 2380 by Mauch and Sadler (2007), Pracy et al. (2016) and Condon et al. (2019), each with
 2381 complete volume-limited samples. However, the RLF of extended radio sources has
 2382 not been characterised as a separate sample, because estimating this RLF would need
 2383 a large number of extended sources with known redshifts. Such redshifts are usually
 2384 obtained by cross-matching the radio sources with their host galaxies in the optical
 2385 or infrared. While this cross-matching is relatively straightforward for compact/unre-
 2386 solved sources, and can therefore be automated (e.g. Kimball & Ivezić, 2008), extended
 2387 radio emission may be complex and may not be co-located with the host galaxy. Cross-
 2388 identification of complex, extended sources is therefore usually done manually, limit-
 2389 ing the sample size.

2390 Wide-area radio surveys like the Very Large Array Faint Images of the Sky at Twenty
 2391 Centimeters survey (FIRST; White et al., 1997) have led to catalogues of hundreds of
 2392 thousands of radio objects, most of which are associated with radio AGN (Sadler et
 2393 al., 2002). While individual galaxy properties vary significantly between galaxies, the
 2394 large sample size of such surveys allows us to examine reliable bulk statistics that may
 2395 tightly constrain models and theory of radio galaxy evolution and formation (Condon,
 2396 1992). With a sufficiently large sample (of size N such that $N \gg \sqrt{N}$; Condon,
 2397 1991) we can also divide the RLF into fractions based on the physical properties of
 2398 each galaxy. These *fractional RLFs* show how different physical processes comprise the
 2399 luminosity distribution, and can be used to investigate how these processes relate to
 2400 the properties of the AGN.

In this chapter we calculate the RLF for extended radio sources in FIRST. Throughout this chapter we define an *extended radio source* as a collection of extended radio components with the same host galaxy, and following Banfield et al. (2015) we define an *extended radio component* as a radio component which fulfils Equation 5.1:

$$\frac{S_{\text{peak}}}{S_{\text{int}}} < 1 - \frac{0.1}{\log_{10}(S_{\text{peak}}/1 \text{ mJy})}, \quad (5.1)$$

where S_{peak} is the peak radio flux density and S_{int} is the integrated radio flux density. We define *radio components* as Gaussians fit to radio emission, *radio islands* as connected patches of radio emission above a local 4σ value, and *radio sources* as sets of radio islands or components associated with the same galaxy. The radio flux density of a source is the sum of the flux densities of the components according to the FIRST catalogue. The (*infrared*) *host galaxy* of a radio source or component is defined as the infrared galaxy associated with the radio emitter. A *cross-identification* is an association of a host galaxy with one or more radio components or islands. We define *candidate host galaxies* (or simply *candidates*) as infrared objects that are near a radio component on the sky and thus may potentially be the host galaxy of that component.

Upcoming radio surveys such as the Evolutionary Map of the Universe (EMU; Norris et al., 2011) are expected to increase the number of complex radio sources to around 7 million (Banfield et al., 2015). Manual, expert cross-identification for such surveys will be impractical. One way forward could be to ask non-expert volunteers for help with manual cross-identification, which is the approach taken by Radio Galaxy Zoo (RGZ; Banfield et al., 2015). This is called *citizen science* and has been employed successfully in many fields with large datasets (Marshall et al., 2015). Even this approach is not sufficient for 7 million sources, though, with RGZ cross-identifying around 75 000 sources in four years. Machine learning provides a potential pathway to obtaining useful physics from such large samples.

In this chapter, we train a machine learning model based on Alger et al. (2018), using RGZ as training data, and use this model to automatically cross-identify 244 846 radio components catalogued by FIRST. This results in 34 305 sources with spectroscopic redshifts. This is the largest available catalogue of extended radio source cross-identifications. We call our catalogue *RGZ-extrapolated* or *RGZ-Ex*. Due to our large sample size we are able to further divide the source population by properties of the host galaxies. We refer to the trained machine learning model and the associated cross-identification algorithm jointly as *binary cross-identification*, or *BXID*. Note that there are two stages to our automated approach: first, we train BXID using an existing catalogue of cross-identified sources; second, we generate new cross-identifications for radio components not in the training catalogue. We refer to these stages as *training* and *prediction* respectively. Our data sources are described in Section 2. Our approach is discussed in Section 3. Radio luminosity functions of extended sources are presented in Section 4 and we discuss these functions in Section 5.

Throughout this chapter we assume a flat Λ -CDM cosmology of $H_0 = 69.3 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.287$. These are the cosmological parameters from the Nine-year Wilkinson Microwave Anisotropy Probe (WMAP9; Hinshaw et al., 2013).

2 Data

In this section we describe how we obtain our training and prediction data. To enable the estimation of the extended RLF, we apply a number of selection criteria to the data which are shown in Figure 5.6.

2442 2.1 RGZ

2443 RGZ is a citizen science project that aims to cross-identify complex radio sources with
 2444 mid-infrared host galaxies with the help of volunteers. The first RGZ data release con-
 2445 tains around 75 000 cross-identifications of a random subset of extended sources in
 2446 FIRST with their host galaxies in AllWISE. For more details on RGZ see Banfield et al.
 2447 (2015). This catalogue has also been used in other machine learning contexts, includ-
 2448 ing supervised learning for source aggregation (Wu et al., 2019) and unsupervised
 2449 learning methods (Galvin et al., 2019; Ralph et al., 2019). We discard the RGZ sources
 2450 without a host galaxy detected in AllWISE and use the remaining 41 446 sources for
 2451 training.

2452 The RGZ catalogue only contains sources with at least 0.65 weighted volunteer
 2453 agreement, equivalent to approximately 80 per cent reliability. This implicitly selects
 2454 for less complex sources, since the volunteer agreement is a proxy for the difficulty of
 2455 cross-identifying a source (Wong et al. in prep).

2456 2.2 FIRST

2457 We use radio imagery and select radio components from the FIRST survey and associ-
 2458 ated catalogue respectively (Helfand et al., 2015; White et al., 1997). FIRST is a 1.4 GHz
 2459 radio survey covering 10 575 deg² of the sky north of Dec = −10° with an angular res-
 2460 olution of 5.4''. At a detection limit of 1 mJy, the catalogue contains 946 432 radio
 2461 components.

2462 For both training and prediction, we make use of image cutouts from FIRST cen-
 2463 tered on mid-infrared candidate hosts. We predict host galaxies for the 244 846 ex-
 2464 tended FIRST components detected at > 10 σ (about 1.5 mJy beam^{−1}, per Banfield
 2465 et al., 2015) that have complete radio imagery for all candidate hosts. The numbers of
 2466 objects removed by our quality filters are shown in Appendix F.

2467 2.3 AllWISE

2468 We use the sky coordinates and magnitudes from the AllWISE (Cutri et al., 2013) cata-
 2469 logue during training and prediction. AllWISE is an all-sky catalogue of mid-infrared
 2470 objects detected by the *Wide-field Infrared Survey Explorer* (WISE Wright et al., 2010) at
 2471 3.4, 4.6, 12 and 22 μm wavelengths (called W1–W4, respectively). AllWISE contains
 2472 over 747 million objects detected above 5 σ at 3.4 and 4.6 μm. WISE has an angular
 2473 resolution of 6.1'' and 6.4'' at these wavelengths, respectively.

2474 We store AllWISE objects in a MongoDB¹ database with a geospatial index on the
 2475 right ascension and declination. This allows us to perform fast spatial lookups. The
 2476 geospatial index in MongoDB assumes a perfectly spherical Earth with a fixed radius,
 2477 allowing us to use it for sky coordinate searches. We use the right ascension and decli-
 2478 nation of AllWISE sources to generate candidate hosts by searching for infrared sources
 2479 near FIRST components.

¹<https://www.mongodb.com/>

Table 5.1: Medians and standard deviations used to normalise input features [for our classifiers](#).

Feature	Median	Standard deviation
Radio image	13.2 μ Jy	3.01 mJy
W1 – W2	0.289 mag	0.378 mag
W1 – W3	4.350 mag	1.067 mag
W1 – W4	7.853 mag	1.144 mag
W2 – W3	4.016 mag	0.958 mag
W2 – W4	7.541 mag	1.046 mag
W3 – W4	3.518 mag	0.409 mag
W1	16.659 mag	1.154 mag

2480 2.4 SDSS

2481 While we do not use data from the Sloan Digital Sky Survey Data Release 15 (SDSS;
 2482 Aguado et al., 2019) for training or prediction, we do use SDSS for spectroscopic red-
 2483 shifts of our host galaxies. These redshifts are required to calculate the radio luminos-
 2484 ity of our sources. We use CDS X-Match² to match each infrared host galaxy to the
 2485 closest source imaged by SDSS to within 5''. This results in 34 305 spectroscopic red-
 2486 shifts for our 158 337 total host galaxies. To estimate the rate of false association with
 2487 SDSS sources, we add a 1' offset to all host positions and redo the matching process.
 2488 With this method we estimate a 0.4 per cent rate of false association.

2489 3 Method

2490 We apply the binary classification cross-identification method (BXID) following Alger
 2491 et al. (2018). This method casts cross-identification as a classification problem where
 2492 infrared ‘candidate host galaxies’ are classified as either being host galaxies or not. A
 2493 classifier is trained on examples of host galaxies and non-host galaxies drawn from
 2494 a cross-identification catalogue, for which we use RGZ. Other related algorithms de-
 2495 veloped to automatically cross-identify radio objects include Bayesian methods (Fan
 2496 et al., 2015), likelihood ratio (Weston et al., 2018), positional matching (e.g. Kimball &
 2497 Ivezić, 2008; Middelberg et al., 2008; Norris et al., 2006), and positional/image hybrid
 2498 approaches (van Velzen et al., 2012), but these methods do not make use of existing
 2499 cross-identification catalogues and most assume compact radio sources or that the pro-
 2500 jected radio emission overlaps the host galaxy.

We represent candidate host galaxies by a $2' \times 2'$ radio image from FIRST centred on that galaxy, the 3.4 μ m magnitude, and the six colours (magnitude differences) derived from the four *WISE* wavelengths. Unknown values of infrared flux were set to their upper limits in AllWISE. We note that many W3 – W4 colours are missing, so this feature may be less useful than the others. We normalise the colours and magnitude by subtracting the median and dividing by the standard deviation. We normalise

²<http://cdsxmatch.u-strasbg.fr/>

each pixel in each radio image by subtracting the median, dividing by the standard deviation, and applying a logistic function (σ ; Equation 5.2) to account for the high dynamic range of radio images.

$$\sigma(a) = \frac{1}{1 + \exp(-a)} \quad (5.2)$$

2501 The medians and standard deviations are reported in Table 5.1. These values are com-
2502 puted across the training set.

2503 For each FIRST component we generate a set of candidate host galaxies. An AllWISE
2504 object is considered a candidate host for a radio component if it is within $\sqrt{2} \times 1.5'$ of
2505 the centre of the two-dimensional Gaussian fit for that component. This search radius
2506 is the maximum angular distance that a host galaxy can be located in RGZ due to the 3'-
2507 wide square images shown to volunteers. Candidate hosts are assigned binary labels:
2508 All candidates identified as host galaxies in RGZ are assigned a positive label and all
2509 others are assigned a negative label. Following Alger et al. (2018) we train a convolutional
2510 neural network (CNN) on the labelled candidate hosts. We base our model on
2511 ResNet18 (He et al., 2016) pretrained on the ImageNet classification task, with the final
2512 layer removed and replaced by a logistic regression model. Non-image features (i.e.
2513 colours and the 3.6 μm magnitude) are concatenated with the features that are output
2514 by the final ResNet18 layer. Using Adam (Kingma & Ba, 2015) to optimise our weights,
2515 we train this model on our task until binary cross-entropy loss starts to increase on a
2516 randomly-selected 20 per cent validation set. We use PyTorch (Paszke et al., 2017) to
2517 implement this model. The scores of each FIRST component are weighted by a one-
2518 dimensional Gaussian function of angular separation, and the candidate maximising
2519 this weighted score is selected as the host galaxy. We set the standard deviation of the
2520 Gaussian to 120'' as this provides good empirical results.

2521 92 per cent of the host galaxies in RGZ are also detected as host galaxies in RGZ-Ex.
2522 The mean volunteer agreement on all RGZ sources with detected hosts was 95^{+5}_{-13} per
2523 cent, compared to 88^{+12}_{-17} per cent for sources with hosts not in RGZ-Ex. Incorrect cross-
2524 identifications can be considered a source of noise in the statistics. In future work
2525 we will design a way for BXID to output an ‘uncertainty’ so sources with uncertain
2526 cross-identifications can be removed from calculations. We quantify the reliability of
2527 RGZ-Ex in Section 3.1.

2528 We note that BXID necessarily identifies a host galaxy for all radio emission, even
2529 when it does not make sense to do so. In our current work we treat this as a source of
2530 noise. Future extensions to BXID will allow it to output ‘no detected host’.

2531 3.1 Visual verification

2532 To quantify the reliability of RGZ-Ex, ~~some of the authors we~~ (M.A., O.W., A.K., N.M.,
2533 and A.T.) visually verified a randomly selected set of 200 radio components/host galaxy
2534 pairs in RGZ-Ex. For each pair we decided whether the radio component matched its
2535 identified host or not. This allowed us to estimate the accuracy of radio component-
2536 infrared host pairs identified by BXID. Verification of component-host pairs is noisy

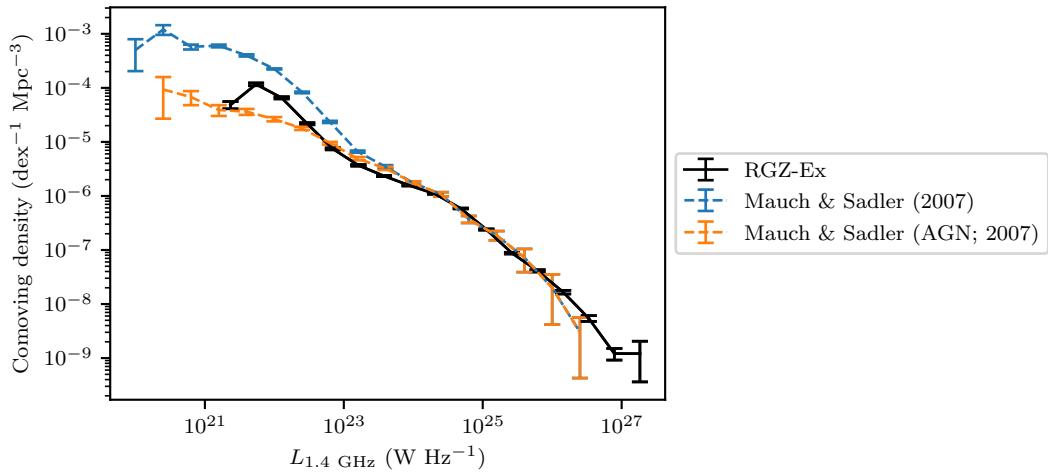


Figure 5.1: RGZ-Ex radio luminosity function compared with the RLFs of Mauch and Sadler (2007).

and sometimes even subjective, so each author looked at all 200 components. If an author was unsure about a pair, they were allowed to ignore it and the verification was treated as missing label data. We then aggregated these verifications following the aggregation approach introduced by Dawid and Skene (1979). This approach jointly estimates the accuracy of each author along with the aggregated verification by maximising the joint likelihood. Note that this approach is able to handle missing label data. We propagated the uncertainty in the accuracy with a Monte Carlo approach, sampling aggregated verifications from the probabilities resulting from the Dawid and Skene model. The estimated accuracy of BXID is (89.5 ± 0.8) per cent. It is interesting to note from our results that even astronomers disagree significantly on the radio cross-identification task, with agreement between authors ranging from 50 to 100 per cent across the verification components. We report the full verification set in Appendix G.

4 Radio luminosity functions

In this section we present our radio luminosity functions (RLFs) derived from the RGZ-Ex catalogue. We first eliminate sources and components with high sidelobe probability according to the FIRST catalogue and low BXID scores (see Appendix F for a diagram of this filtering process). To calculate each RLF we followed the $1/V_{\max}$ method (Schmidt, 1968). This method accounts for the effects of Malmquist bias, which is a systematic bias against sources at greater distances. We describe this approach in Appendix H. We limit our sample to radio sources with 1.4 GHz integrated flux density of at least 2 mJy associated with host galaxies brighter than magnitude 17 at $3.4 \mu\text{m}$, a spectroscopic redshift $0.02 \leq z \leq 0.6$, and an i -band magnitude < 20 . We chose these limits based on the distribution of redshifts and infrared magnitudes as well as the sensitivity of FIRST. We then remove sources with unusually high or low

W1 magnitude for their redshift (more than 3 standard deviations from the mean) because many such sources have incorrect spectroscopic redshifts, e.g. blazars. There are 24 743 sources matching all criteria. We assume a spectral index of $\alpha = -0.7$ (as is common in literature, e.g. Condon et al., 2002) with flux density $f \propto \nu^\alpha$ where ν is the frequency. We calculate the k -corrected radio luminosity (Kochanek et al., 2001) as follows:

$$L = \frac{4\pi f d^2}{1+z} (1+z)^{-\alpha} \quad (5.3)$$

where z is redshift and d is luminosity distance (a function of z). Uncertainties in comoving density are estimated as described in Appendix H. Completeness estimates are shown in Appendix I. We discuss biases in our methods and results in Section 5.1.

We compare our RLFs with Mauch and Sadler (2007), who estimated RLFs from 7 824 manually cross-identified radio sources in the NRAO VLA Sky Survey (NVSS; Condon et al., 1998). Their RLFs were split into AGN and star-forming radio sources. While we do not make this split explicitly in our catalogue, we expect both RGZ-Ex and RGZ to be dominated by AGN due to the selection criterion of being extended in the selected redshift volume. We note that the redshift range used in our work, $0.02 < z < 0.6$, differs from the $0.003 < z < 0.3$ range used by Mauch and Sadler (2007).

In Figure 5.1 we show the RLF derived from RGZ-Ex along with the RLFs from Mauch and Sadler (2007). There is good agreement between all three luminosity functions for luminosities greater than 10^{23} W Hz $^{-1}$ and below this luminosity the RGZ-Ex RLF is bounded above by the Mauch and Sadler (2007) RLF. RGZ-Ex generally finds less comoving density than Mauch and Sadler (2007), which we attribute to our requirement for extent. We suggest that the peak in RGZ-Ex RLF at approximately 10^{22} W Hz $^{-1}$ is due to our sample containing a small fraction of star-forming galaxies. Our criterion, however, does cut out most star-forming regions as these are often compact, which is why we report lower densities than the star-forming RLF of Mauch and Sadler (2007).

The *WISE* colour-colour plot, shown for RGZ-Ex in Figure 5.2, is often used to categorise galaxies at different evolutionary stages into four mid-infrared colour regions that are typically populated by 1) spheroidals or elliptical galaxies; 2) quasi-stellar objects (QSOs), Seyferts or powerful AGN; 3) starbursting or luminous infrared galaxies (LIRGs); and 4) the intermediate region where the other three regions overlap. The horizontal axis, $W2 - W3$, separates early- and late-type galaxies, with the star-forming late-type galaxies appearing redder (further to the right) (Wright et al., 2010). The vertical axis, $W1 - W2$, separates inactive galaxies from AGN with strongly radiating accretion discs (Sadler et al., 2014). In Figure 5.3 we show the radio luminosity function split by host galaxy location in the mid-infrared colour-colour plot as defined by Jarrett et al. (2017).

Many sources have $W3$ detections with low signal-to-noise, limiting our ability to subdivide our sample. We plot both the RLFs for the sample with only $W3 \geq 3\sigma$ as well as the RLFs for the full sample in Figure 5.3. For the full sample we use the lower magnitude limit from AllWISE as the $W3$ magnitude (which is an upper flux limit).

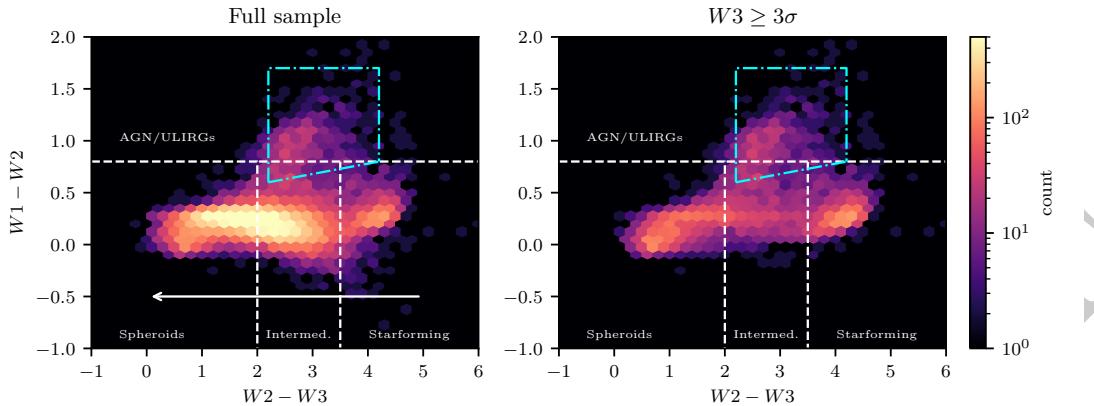


Figure 5.2: *WISE* colour-colour distributions. The dashed grey lines show simple host galaxy class divisions from Jarrett et al. (2017). These classes are labelled in the plot. The blue dot-dashed line shows the empirical optical/infrared AGN criteria from Jarrett et al. (2011). The arrow shows the direction that galaxies would shift with fainter $W3$ magnitudes. The right plot limits the sample to only sources with $W3 \geq 3\sigma$.

Using the upper flux limit as the real $W3$ flux has the effect of increasing $W2 - W3$ compared to a real detection, so objects appear further to the right of the colour-colour diagram (Figure 5.2) than they ought to. This means that due to $W3$ limits, objects that should be in the spheroid set will appear in the intermediate and star-forming sets, and objects from the intermediate set will appear in the star-forming set.

At low luminosities, our extended source RLF is dominated by galaxies with infrared colours consistent with star formation. The fraction of the RLF composed of the star-forming set drops off rapidly for $L_{1.4 \text{ GHz}} > 10^{22} \text{ W Hz}^{-1}$, as expected for galaxies with radio emission dominated by star formation (e.g. Mauch & Sadler, 2007). However, the RLF slope flattens out again beyond $10^{24} \text{ W Hz}^{-1}$, suggesting a second source population. This population has many missing $W3$ measurements, and these are likely intermediates or spheroids incorrectly included in the star-forming set. We therefore suggest that the low-luminosity RGZ-Ex sample mostly contains nearby galaxies with radio emission due to star formation, which appear extended in FIRST as they are close enough for FIRST to resolve their disc (greater than 20 kpc at $z = 0.2$). The remaining fraction of star-forming sources found by Mauch and Sadler (2007), shown in Figure 5.1, would not be resolved in FIRST, as they are small or distant.

Spheroids, which are hosts in the mid-infrared region corresponding to ellipticals and stars (Wright et al., 2010), comprise the majority of radio galaxies at $10^{23} \text{ W Hz}^{-1}$, and have a peak density at $10^{22} \text{ W Hz}^{-1}$. The common host galaxies for radio-loud AGN tend to be passively-evolving spheroids. It is not surprising that they are more common than star-forming galaxies at luminosities greater than $10^{22} \text{ W Hz}^{-1}$. Above $10^{25} \text{ W Hz}^{-1}$ they are less common than intermediate galaxies and their contribution to the luminosity function drops rapidly. This is likely due to the loss of $W3$ detections moving spheroids into the intermediate set, and we hypothesise that with deeper $W3$

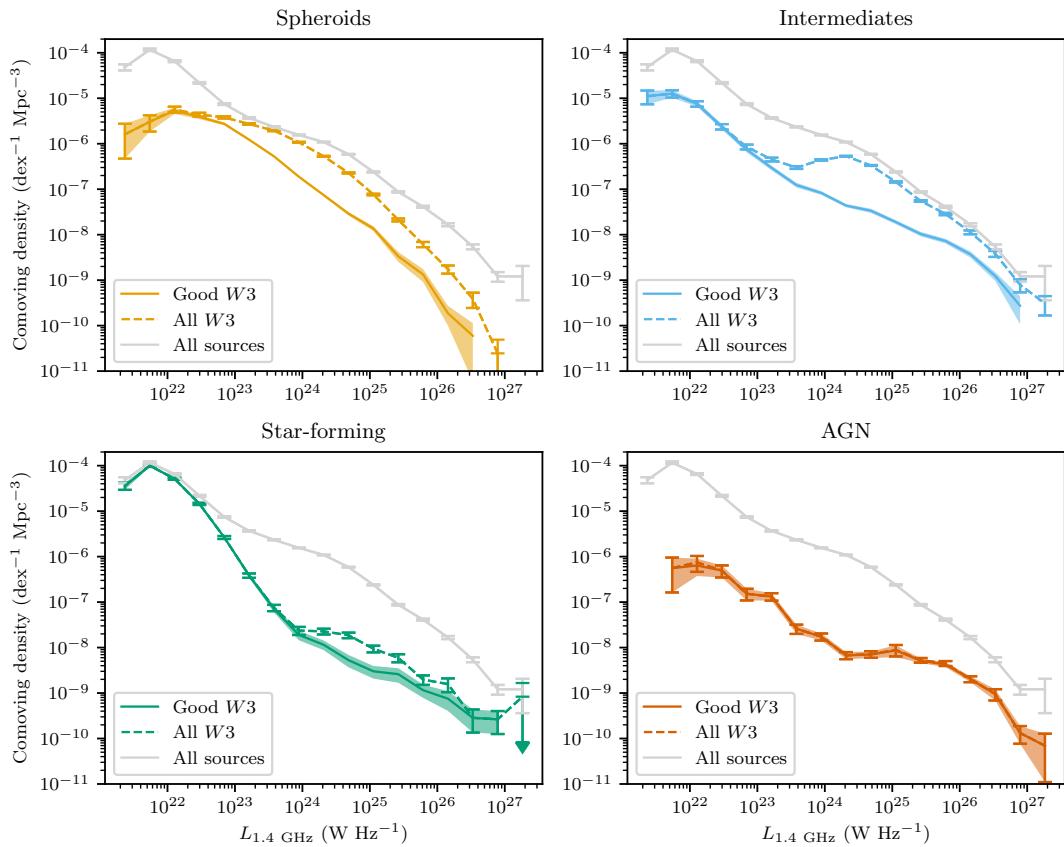


Figure 5.3: RLFs split by host galaxy location in the WISE colour-colour plot (Figure 5.2), using our automated cross-identifications. The grey line is the total RLF for all sources. Solid lines have good W3 detections and dashed lines include W3 with low signal-to-noise.

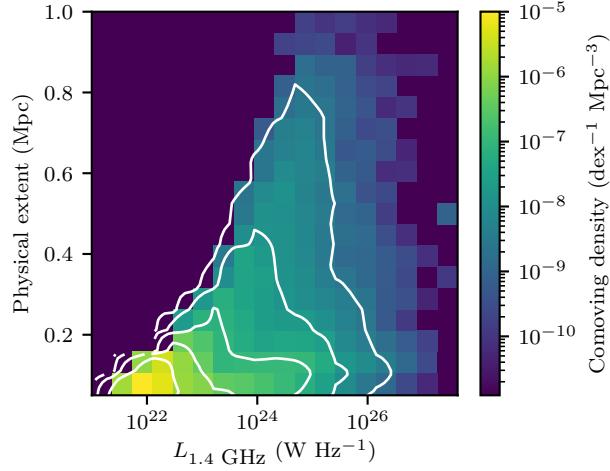


Figure 5.4: Bivariate radio luminosity function showing radio luminosity against projected physical extent. Contours are on a log scale, starting at the median and increasing by 10 per cent per contour.

2611 observations spheroids may dominate above $10^{25} \text{ W Hz}^{-1}$.

2612 Sources with hosts in the mid-infrared AGN region of the colour-colour diagram
 2613 (Figure 5.2) make up the smallest contribution to the radio luminosity function. They
 2614 have a steadily decreasing density from their lowest observed $L_{1.4} \text{ GHz}$ of $10^{22} \text{ W Hz}^{-1}$
 2615 to their highest of $10^{27} \text{ W Hz}^{-1}$, but are present in all luminosity bins except for the
 2616 very lowest. This is a set with a very low fraction of spectroscopic SDSS matches for
 2617 the *WISE* host galaxies. 26 per cent of hosts outside the *WISE* AGN region have an
 2618 SDSS match, compared to just 12 per cent of hosts inside the *WISE* AGN region. This
 2619 is likely due to the incomplete sampling of QSOs in the SDSS spectroscopic survey or
 2620 redshift evolution effects (Strauss et al., 2002). The fraction of the RLF contributed
 2621 by galaxies classed as mid-infrared AGN increases above $10^{25} \text{ W Hz}^{-1}$, meaning that
 2622 high-luminosity radio AGN are also more likely to be infrared AGN than at lower radio
 2623 luminosities. Note that the AGN set is unaffected by missing $W3$ detections, as the
 2624 AGN set is based only on $W1 - W2$.

2625 Galaxies residing in the intermediate mid-infrared colour region can be populated
 2626 by both early- and late-type galaxies, which have a mix of processes occurring within
 2627 them. These ‘intermediate sources’ dominate in most luminosity ranges, and above
 2628 $10^{24} \text{ W Hz}^{-1}$ they comprise the vast majority of our sample. As intermediate-type
 2629 galaxies fall between star-forming galaxies and passive ellipticals on the mid-infrared
 2630 colour-colour plane, they do not have a clear morphological class and are composed
 2631 of overlapping subsets of sources. The most luminous radio AGN are almost entirely
 2632 found in this set of galaxies. In fact, as radio luminosity increases the density fraction
 2633 shifts from spheroids toward intermediate galaxies, likely due to missing $W3$ moving
 2634 objects from the spheroid set into the intermediate set.

2635 In Figure 5.4 we show the radio luminosity function for different ranges of projected

physical extent of their radio emission. We estimate the angular extent as the angular distance between the most separated components in a multi-component source. This result is complementary to other Radio Galaxy Zoo studies on the effect of the environment on the size and asymmetry of the observed extended radio emission (Garon et al., 2019; Rodman et al., 2019).

5 Discussion

5.1 Biases and uncertainties

Biases enter our work due to our chosen samples and methods. Our training set, RGZ, is biased toward sources smaller than $1.5'$ and limited above by $\sqrt{2} \times 3'$ due to the $3' \times 3'$ cutout size of the RGZ user interface. RGZ volunteers preferentially select host galaxies that are brighter in $W1$, so we expect RGZ to overrepresent the number of sources with $W1$ -bright host galaxies.

These biases may affect our trained algorithm: for example, the overabundance of $W1$ -bright host galaxies in RGZ may cause our algorithm to be less accurate when unassociated bright galaxies are in the field of view. Without knowing the true distribution of host galaxies, however, it is difficult to quantify the effect of such biases on our trained method.

FIRST itself is also biased. Helfand et al. (2015) describe several reasons why FIRST flux may be systematically underestimated. Most of these effects are insignificant for extended objects in our sample or are corrected in the FIRST catalogue from which we draw our flux information. The exception is the ‘resolving out’ of diffuse and low surface brightness radio emission by the Very Large Array in its B configuration. This means that we lose flux on most nearby radio galaxies (especially those with very diffuse components) and may miss diffuse or dim radio galaxies entirely. More diffuse radio galaxies such as Fanaroff-Riley type I (FRI; Fanaroff & Riley, 1974) galaxies tend to be toward the low end of the radio-loud luminosity distribution, about $10^{23} \text{ W Hz}^{-1}$ (Best, 2009), so we expect that losing diffuse sources would lower our estimates of density around this luminosity. Large, extended lobes such as those associated with Fanaroff-Riley type II (FRII; Fanaroff & Riley, 1974) galaxies may also be resolved out, so by the same mechanism we expect to lose an increasing amount of flux with increasing source angular size. This effect is compounded by flux loss at 1.4 GHz associated with synchrotron losses and adiabatic expansion losses (Blundell et al., 1999).

Our host galaxy redshifts may be biased. Incorrectly identifying the host galaxy may introduce sources with incorrect redshifts into the RLF, an effect which will be dominated by misidentifying galaxies as hosts where the true host is not detected. Since we are matching to optical spectra in SDSS to find redshifts, we are biased toward brighter host galaxies that are more likely to have such spectra. Without an optically-complete sample — currently impossible on such scales—this effect is unavoidable. Brighter optical sources appear at lower redshifts, so we likely undersample higher-redshift (and hence higher-luminosity) galaxies.

Our requirement for radio emission to be extended will miss radio galaxies that

would be resolved and extended if they were not aligned with the line of sight. We therefore must be underestimating the population of extended sources (though assuming a random distribution of orientations, the majority of galaxies will not be aligned close to the line of sight). The requirement for extended radio emission will also impose a lower limit on linear size, which will vary with redshift: at $z = 0.6$ the effect will be strongest and we will see no sources with linear size under 33.5 kpc. This will cause us to underestimate the population of radio galaxies with linear sizes between 10–30 kpc. On the other hand, we have likely avoided significant overestimation of radio luminosity due to relativistic beaming, since we filter out sources aligned along the line of sight.

We have estimated uncertainties in our RLF from Poisson noise in the histogram bins. We have likely underestimated these uncertainties as it is difficult to estimate uncertainty in our algorithm, though in future we anticipate that we can employ an ensemble of classifiers to estimate this (e.g. Lakshminarayanan et al., 2017).

5.2 Extended radio galaxies in the low- z Universe

Our total RLFs are consistent with the idea that large, extended radio sources are typically hosted by massive ellipticals (Best et al., 2005). These RLFs match existing RLFs such as that of Mauch and Sadler (2007), except at radio luminosities below $10^{22} \text{ W Hz}^{-1}$. This is unsurprising since we employ a requirement for extended emission, and, besides very nearby star-forming galaxies, FRII comprise most of the population of extended radio objects. The fractional RLF split by mid-infrared colour, Figure 5.3, shows that spheroids reach peak density at a radio luminosity associated with a drop in density of intermediates, and intermediates begin to dominate the RLF as the spheroid density drops. Together, these mid-infrared classes of galaxy form the bulk of the extended radio galaxy RLF.

We see a significant star-forming population in our extended sample, which means that we are likely resolving some discs in radio. While the $1/V_{\max}$ method ensures that our RLFs account for similar galaxies throughout the Universe, even though we only resolve very nearby discs, some of the star-forming population is not included. The difference between our RLF and existing RLFs must be due to the latter containing low-luminosity sources that are compact even when very nearby.

Can we use our RLFs to estimate the kinetic energy contribution of AGN to the galaxy halo and beyond? The extended population of AGN will be the population that contributes most mechanical energy: the major part of the energy in the jet expands the radio lobes, drives shocks or is stored in the jet magnetic field, rather than being emitted as radiation (Godfrey & Shabala, 2016; Hardcastle & Krause, 2014). Extended radio sources should therefore represent the bulk of AGN feedback: radio galaxies with extended jets will inject mechanical energy out to larger distances from the core of the host galaxies than those with smaller jets. This is supported by e.g. Turner and Shabala (2015), who found that extended sources comprised the bulk of the mechanical energy contribution. By assuming a relationship between radio luminosity and radio jet mechanical energy, we can use our extended source RLFs to estimate the con-

tribution of extended AGN to energy in the intergalactic/circumgalactic medium (IGM/CGM). But assuming such a relationship is not without problems: the radio lobe luminosity experiences significant evolution (e.g. Bicknell et al., 1997), the surrounding IGM/CGM may interact with the radio lobe expansion in non-trivial ways (e.g. Hardcastle & Krause, 2013) and the relationship between the mechanical energy and radio luminosity has high scatter on individual radio sources (Hardcastle & Krause, 2013). With our sample size, these effects should be diminished, and with these caveats in mind we will estimate the energy contribution of extended sources to the IGM. We assume a scaling relation of $\ln Q = \beta \ln L_\nu + Q_0$, where Q is the jet power and L_ν is the monochromatic radio luminosity at frequency ν . The values for β and Q_0 vary significantly across the literature, based on different physical assumptions and samples. Willott et al. (1999) presented a widely-used relationship

$$\ln Q = \ln(f^{3/2} 3 \times 10^{38}) + \frac{6}{7} \ln \left[\frac{L_{151 \text{ MHz}}}{10^{28} \text{ W Hz}^{-1}} \right], \quad (5.4)$$

with a scaling constant $1 \leq f \leq 20$ and Q in watts. Other models exist with different slopes, e.g. Bîrzan et al. (2008) suggest that $\beta \approx 0.5 - 0.7$ and Cavagnolo et al. (2010) find $\beta \approx 0.7$. [Godfrey and Shabala \(2016\)](#) [Shabala and Godfrey \(2013\)](#) show that the scalings presented by Willott et al. (1999) are consistent with independent theoretical modelling for high-power radio galaxies. [Godfrey and Shabala \(2016\)](#) on the other hand provide a summary of the literature in this field and suggest that these correlations are from mutual distance dependence rather than intrinsic relationships. They find that there is no strong empirical evidence for such a correlation in either FRI or FRII. However, their theoretical models suggest $\beta \approx 0.5, 0.8$ for FRI and FRII respectively, which is consistent with Willott et al. (1999). [Shabala and Godfrey \(2013\)](#) on the other hand show that the scalings presented by Willott et al. (1999) are consistent with independent theoretical modelling for high-power radio galaxies. The relationship between luminosity and kinetic energy is not yet settled, but we can still use this popular scaling method both to explore the consequences of our RLFs and for comparison with previous work.

Scaling the frequency to 1.4 GHz, and assuming a spectral index of $\alpha = -0.7$, $\beta = 6/7$, and $Q_0 = \ln(f^{3/2}) + 89.9$, we can write the Willott et al. (1999) relation as

$$\ln(Q) = \ln(f^{3/2}) + 89.9 + \frac{6}{7} \ln \left[\frac{L_{1400 \text{ MHz}}}{10^{28} \text{ W Hz}^{-1}} \right]. \quad (5.5)$$

Assuming $f \in [1, 20]$ gives $Q_0 \in [89.9, 94.4]$. Integrating over our RLF we find $Q \in [1.3 \times 10^{30}, 1.2 \times 10^{32}] \text{ W Mpc}^{-3}$. This is likely a lower limit as we are missing extended radio sources oriented along our line-of-sight and nearby diffuse extended radio sources (e.g. FRI), and Shabala (2018) argues that many ‘compact’ AGN may in fact be extended but below the sensitivity of surveys such as FIRST. Our results are consistent with other literature (e.g Hardcastle et al., 2019, who estimated the energy contribution as $7 \times 10^{31} \text{ W Mpc}^{-3}$).

2730 **5.3 Future work**

2731 With such a large sample size, further partitioning of the RLF into subsamples is pos-
2732 sible. Any combination of the features investigated here, plus further host galaxy and
2733 radio properties, could be used to generate fractional RLFs. Automated classifiers such
2734 as ClaRAN (Wu et al., 2019) or feature extractors such as PINK (Galvin et al., 2019; Pol-
2735 sterer et al., 2015; Ralph et al., 2019) could provide a way to divide the RLF by radio
2736 morphology. These methods provide a way of dividing galaxy classes based directly
2737 on the radio image, rather than the host galaxy like we have done here, and so should
2738 not be affected by extinction or redshift in the same way as our sample. Such sub-
2739 samples would lend insight into how radio power is connected to radio morphology
2740 and generation mechanisms. Cross-matching with other surveys such as NVSS or the
2741 150 MHz TIFR GMRT Sky Survey would provide properties such as the spectral index
2742 and observations of diffuse emission missed by FIRST (as used by Kimball & Ivezić,
2743 2008). Such properties could also be used to create interesting and insightful fractional
2744 RLFs.

2745 While we have not investigated the link between extended sources and their local
2746 environments, this will be the focus of future work. Environment will heavily factor
2747 into the source sizes, morphologies, and so on, following work such as Rodman et al.
2748 (2019) and Garon et al. (2019).

2749 Ongoing radio surveys such as EMU, VLASS (Lacy et al., 2020), and LoTSS
2750 (Shimwell et al., 2019) will greatly increase the number of extended sources. How-
2751 ever, our sample size limitations in this chapter are not from FIRST, but from SDSS:
2752 until next-generation spectroscopic surveys are available, redshifts will be the limiting
2753 factor. To significantly increase our sample size would require much greater numbers
2754 of redshifts.

2755 **6 Summary**

2756 Extended radio sources provide an opportunity to study the interaction between AGN
2757 and their large-scale environments. We trained the binary cross-identification method
2758 on the Radio Galaxy Zoo to generate the largest sample of reliably cross-identified,
2759 extended radio sources, and this large sample allows us to investigate their bulk dis-
2760 tributions in new, detailed ways. We estimated radio luminosity functions split by
2761 mid-infrared colour, physical extent and redshift. Despite our extendedness criterion,
2762 we found a significant star-forming population. We estimated that extended AGN
2763 contribute between 1.3×10^{30} and 1.2×10^{32} W Mpc $^{-3}$ of mechanical energy to their
2764 environment. Ongoing and future surveys such as EMU will provide even greater
2765 numbers of extended radio sources, and our combination of machine learning and as-
2766 tronomy methodology will allow these samples to be cross-identified and investigated
2767 efficiently and reliably.

2768 7 Acknowledgements

2769 This publication has been made possible by the participation of more than 11 000 volunteers
 2770 in the Radio Galaxy Zoo project. Their contributions are individually acknowledged at <http://rgzauthors.galaxyzoo.org>. Radio Galaxy Zoo makes use of data products
 2771 from the *Wide-field Infrared Survey Explorer* and the Very Large Array. The *Wide-field Infrared Survey Explorer* is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by
 2772 the National Aeronautics and Space Administration. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative
 2773 agreement by Associated Universities, Inc. The figures in this work made use of Astropy, a community-developed core Python package for Astronomy (The Astropy
 2774 Collaboration et al., 2018). Partial support for this work for L.R. and A.F.G. comes from National Science Foundation grant AST-1714205 to the University of Minnesota.
 2775 H.A. benefited from grant CIIC 218/2019 of Universidad de Guanajuato. M.A. and A.T. were supported by the Australian Government Research Training Program. We thank
 2776 J. Nabaglo and A. Gruen for their help with efficient data preprocessing. We thank
 2777 H. Zovaro, T. J. Galvin, and H. Tang for their comments on the draft of this chapter.
 2778

2785 F Sankey diagrams

2786 This section presents Sankey diagrams showing the filtering of components and sources
 2787 from the full FIRST sample in this chapter. A Sankey diagram shows the order and
 2788 number of objects removed from a sample. Figure 5.5 shows the filtering of components
 2789 and Figure 5.6 shows the filtering of sources. The component filters are ‘Bad FIRST’ for
 2790 components on the edge of FIRST with incomplete images, ‘Sidelobe’ for components
 2791 with high sidelobe probability, ‘Low score’ for components with only low-scoring candidate
 2792 hosts, ‘Faint’ for components with less than 10 signal-to-noise according to the FIRST
 2793 catalogue, and ‘Compact’ for components that do not have extended radio emission
 2794 according to Equation 5.1. Sources were removed after each component filter if they
 2795 no longer contained any components.

2796 G Visual verification results

2797 In Section 3.1 we described our visual verification of the BXID method. We list the radio
 2798 components in the verification set in Table 5.2. Each row of the table contains the FIRST
 2799 component, its AllWISE host galaxy according to BXID, and whether the association
 2800 is correct according to our visual verification. If an author was particularly unsure
 2801 about an object, they were able to skip this object, and so are not accounted for in the
 2802 verification for that object. Verification was weighted by the Dawid and Skene (1979)
 2803 maximum likelihood model.

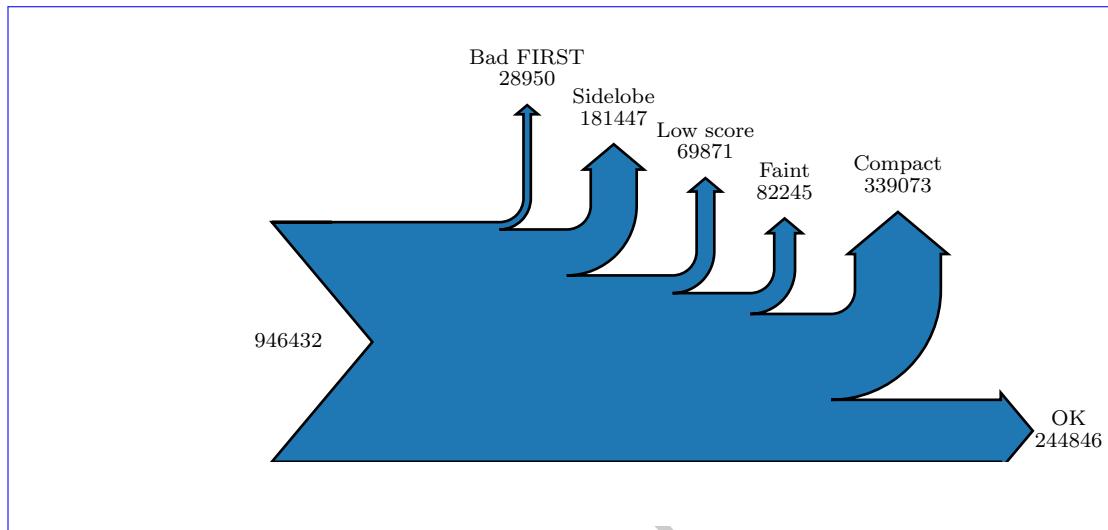


Figure 5.5: Number of components removed from FIRST by each filter.

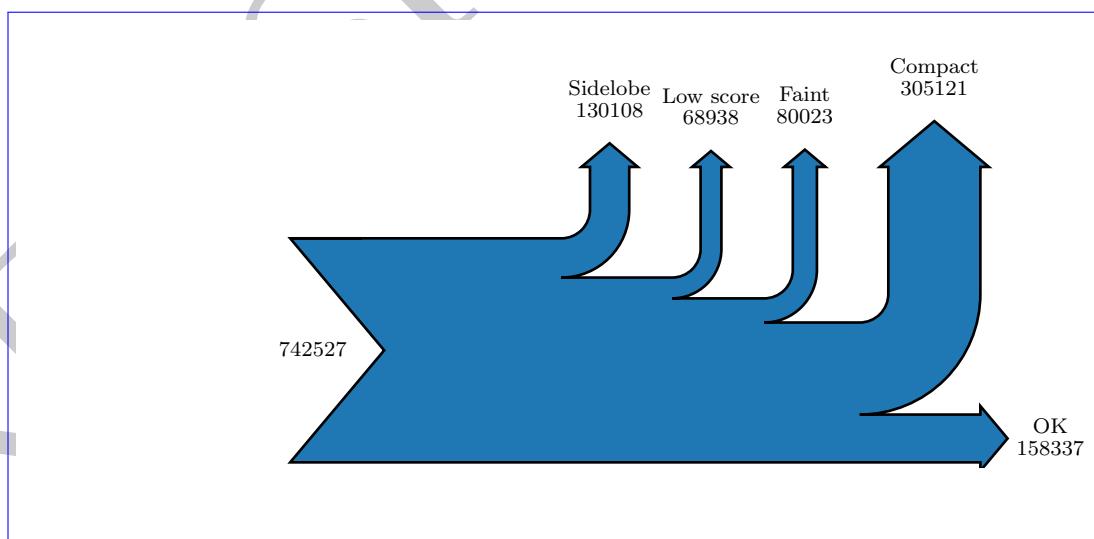


Figure 5.6: Number of sources removed by each filter.

Table 5.2: Validation objects. ‘Agree’ is whether or not the authors of Alger et al. (in prep.) agreed with BXID associating the given FIRST object with the given AllWISE object.

FIRST	AllWISE	Agree	FIRST	AllWISE	Agree
J000234.9-001421	J000242.35-001320.5	n	J094009.5+600403	J094011.55+600357.6	n
J002841.1+141654	J002840.37+141652.7	y	J094023.7+135123	J094023.73+135125.2	y
J003731.4+000156	J003731.26+000146.7	y	J094324.5+435341	J094324.61+435342.0	y
J005407.5-011158	J005407.61-011158.9	y	J094650.8+382015	J094650.44+382010.9	y
J011210.3+002203	J011210.41+002201.9	y	J095011.8+455319	J095011.82+455320.0	y
J012342.4+015849	J012342.24+015850.4	x	J095113.5+180211	J095113.82+180204.2	n
J013015.1+110653	J013015.16+110653.4	y	J095242.4+222638	J095242.45+222638.0	y
J013107.7+070343	J013102.02+070332.0	y	J095538.7+013546	J095539.20+013546.1	y
J014247.9-000039	J014247.81-000040.3	y	J095609.9+363441	J095609.30+363445.4	y
J014250.0-000032	J014247.81-000040.3	n	J095811.8+225056	J095811.90+225055.5	y
J020222.3+030138	J020223.20+030150.4	y	J100019.2+263516	J100018.84+263527.5	y
J020333.8+000853	J020336.94+000759.3	y	J101315.9+064520	J101316.51+064519.0	y
J021840.1-032311	J021840.13-032306.0	y	J101455.2-004716	J101455.30-004718.3	y
J023022.0+010834	J023022.11+010840.0	y	J102153.5+260429	J102153.52+260429.6	y
J024245.3-022535	J024245.35-022534.6	y	J102354.7+390653	J102354.88+390654.0	y
J025901.0+005350	J025901.50+005346.1	y	J102620.4+303600	J102620.46+303550.4	y
J033204.1-004757	J033204.15-004757.1	y	J102710.4+460254	J102714.81+460256.4	n
J073033.2+390413	J073033.21+390412.9	y	J102955.9+424906	J102955.96+424906.7	y
J073954.1+481810	J073954.87+481759.5	y	J103503.9+102404	J103503.92+102403.6	y
J074504.9+331247	J074504.81+331256.2	y	J103839.9+331200	J103839.94+331201.1	y
J074640.4+421709	J074640.45+421709.1	x	J104030.5+211624	J104031.09+211620.6	n
J074707.9+171719	J074708.35+171726.5	y	J104533.8+430025	J104535.22+430020.8	y
J075043.6+274838	J075043.35+274844.8	n	J104907.5+322903	J104907.91+322906.6	x
J075050.3+331937	J075051.25+331905.0	x	J105146.9+552257	J105147.40+552308.4	x
J075422.2+311253	J075422.35+311252.5	x	J105257.5+105418	J105257.53+105421.5	x
J075637.0+212006	J075636.65+212001.4	x	J105521.6+372641	J105521.24+372652.4	x
J082326.1+141438	J082326.34+141435.9	y	J105758.8+321605	J105758.84+321605.3	y
J082422.5+351121	J082422.65+351114.6	y	J110104.9+151618	J110104.90+151618.2	y
J082925.9+462618	J082926.02+462618.5	y	J110353.2+352320	J110353.37+352319.9	y
J083512.4+175441	J083512.45+175441.1	y	J110414.4+481345	J110423.08+481311.0	n
J084133.5+402035	J084133.40+402042.8	x	J111057.7+220756	J111057.18+220758.3	x
J084238.4+405305	J084238.38+405306.6	n	J111208.5+275207	J111201.79+275053.8	n
J084417.3+315845	J084417.92+315845.9	y	J111225.2+233159	J111225.30+233157.9	y
J084728.5+360700	J084728.24+360714.6	x	J111726.3+375336	J111726.35+375337.0	x
J084905.5+111448	J084905.51+111447.8	y	J111746.1+261151	J111746.18+261150.9	y
J085236.8+262006	J085236.11+262013.4	x	J111854.3+424708	J111854.45+424652.8	x
J085415.6+524930	J085415.62+524936.7	x	J112124.4+640417	J112125.02+640408.6	x
J090623.2+300746	J090622.87+300743.9	x	J112135.3+352330	J112135.44+352324.9	x
J091745.1+275049	J091745.89+275103.8	y	J112550.9+200631	J112558.75+200554.3	y
J091752.0+431614	J091752.14+431612.7	y	J112859.7+260923	J112859.86+260911.3	y
J092014.4+302907	J092013.95+302859.3	y	J113201.1+442639	J113201.23+442639.4	y
J092140.5+540118	J092140.24+540121.1	y	J113302.5+355408	J113301.80+355415.3	y
J092213.0+542157	J092213.03+542157.2	x	J113712.7+263301	J113711.86+263335.1	x
J092406.9+562703	J092406.47+562656.2	x	J113756.3+471314	J113756.31+471314.1	x
J092713.1+105841	J092713.14+105839.8	y	J113906.6+230602	J113906.68+230602.1	y
J093108.6+613447	J093108.63+613447.2	x	J114325.0+600721	J114323.90+600737.1	x
J093239.6+052308	J093237.71+052240.7	n	J114759.7+370305	J114759.22+370311.2	x
J093627.8+103610	J093627.87+103609.7	y	J114916.7+083022	J114916.33+083040.5	n
J093645.2+561435	J093645.89+561434.2	y	J115010.9+063340	J115010.93+063340.5	y
J094006.8+482651	J094006.92+482649.2	y	J115308.6+374851	J115316.96+374850.0	y

FIRST	AllWISE	Agree	FIRST	AllWISE	Agree
J115448.7+472222	J115448.67+472223.7	y	J142829.5+070836	J142829.60+070836.3	y
J115603.7+584704	J115603.48+584706.1	y	J143411.0+170036	J143411.18+170035.7	y
J115605.9+343230	J115605.64+343229.4	y	J143624.0-001057	J143623.89-001100.8	y
J115653.0+572338	J115645.38+572151.7	y	J143742.6+104412	J143742.69+104412.8	y
J120138.0+230922	J120137.97+230922.2	y	J143840.8+475355	J143841.08+475356.1	y
J120752.8+533808	J120752.85+533807.3	y	J143909.1+430847	J143909.08+430847.8	y
J120943.3-021934	J120942.89-021943.0	y	J144135.8+102246	J144135.91+102245.1	y
J121045.6+190225	J121045.68+190227.0	y	J144333.6+275229	J144333.02+275250.2	y
J121207.6+115412	J121207.72+115413.8	y	J145012.3+471739	J145012.33+471738.7	y
J121211.3+485951	J121211.86+485952.0	y	J145103.7+452459	J145102.66+452520.5	n
J121406.7+002634	J121406.73+002635.0	y	J145401.6+141009	J145401.70+141009.6	y
J122518.0+350258	J122517.85+350301.9	y	J150158.7+191413	J150158.87+191405.3	y
J122525.1+451530	J122524.71+451508.5	y	J150743.9+352720	J150743.62+352724.1	y
J122640.9+430508	J122640.82+430509.2	y	J151141.6-003209	J151142.01-003213.0	y
J123429.8+260107	J123434.79+260134.3	n	J151315.5+403107	J151315.56+403107.7	y
J123633.1+100928	J123633.12+100928.7	y	J151518.7+230256	J151518.67+230257.3	y
J124839.3+411522	J124839.42+411522.3	n	J151703.6+105947	J151703.68+105947.6	y
J125129.2+551012	J125128.76+551009.3	y	J151736.8+610856	J151736.83+610857.7	y
J130005.8+524801	J130006.14+524803.0	y	J152121.6+281635	J152120.68+281626.2	y
J130132.1+511351	J130132.32+511352.5	y	J152714.8+310425	J152714.88+310424.7	y
J131104.4+464936	J131104.45+464934.0	y	J153428.9+272134	J153429.68+272120.8	y
J131452.2+252811	J131446.81+252820.8	n	J154245.3+100919	J154245.71+100917.8	y
J132033.8+332639	J132033.59+332639.0	n	J154901.6+103159	J154901.40+103152.6	y
J132257.5+191134	J132257.53+191133.9	y	J154925.2+395316	J154926.17+395303.7	y
J132529.3+230734	J132529.35+230733.8	y	J155206.3-005348	J155206.58-005339.3	y
J132546.8+052453	J132546.86+052454.1	y	J155457.3+344637	J155458.45+344644.7	y
J132637.7+112110	J132637.92+112108.8	y	J155743.5+272752	J155743.52+272752.8	y
J132831.8+104339	J132831.88+104338.8	y	J160130.0+083848	J160130.07+083850.7	y
J132932.3+131839	J132932.32+131839.6	y	J160534.8+441220	J160535.55+441221.5	y
J133022.8+311904	J133022.83+311902.8	y	J160859.2+400135	J160901.32+400230.7	n
J133453.3+405653	J133454.13+405650.6	y	J161545.4+231617	J161545.14+231617.2	y
J133741.1+124302	J133741.13+124303.1	y	J161930.4+085533	J161930.51+085532.6	y
J133823.6+103337	J133823.67+103341.9	y	J162228.0+264743	J162228.70+264736.7	y
J134651.2+415154	J134651.06+415156.1	y	J162750.4+473624	J162750.55+473623.5	y
J134704.3+110622	J134704.35+110622.7	y	J162904.2+470852	J162904.34+470853.0	y
J134752.7+555046	J134752.71+555048.6	y	J163038.7+214740	J163037.43+214748.9	n
J134831.7+164325	J134831.57+164328.2	y	J163323.6+424051	J163323.61+424051.9	y
J134949.8+385539	J134949.93+385542.8	y	J163327.5+242426	J163327.87+242427.4	y
J135106.5+074534	J135106.50+074534.2	y	J163533.8+454557	J163534.00+454554.3	y
J135107.7+615502	J135107.75+615502.1	y	J164211.2+512029	J164211.27+512029.3	y
J135658.5+134028	J135659.15+134017.0	y	J165549.1+375923	J165549.01+375923.6	y
J135833.9+180021	J135834.03+180020.4	y	J165620.0+363402	J165619.89+363403.9	y
J140630.7+554017	J140629.32+554009.9	y	J165700.5+474820	J165659.58+474809.0	y
J140804.2+503019	J140804.10+503021.1	y	J171406.2+292712	J171404.16+292704.0	n
J141226.7+454125	J141226.54+454125.5	y	J172126.4+374446	J172126.46+374446.6	y
J141245.0+495213	J141243.84+495206.4	y	J222627.7-005010	J222627.77-005010.8	y
J141317.4+325306	J141317.50+325306.8	y	J223636.4-013827	J223636.48-013827.2	y
J141723.8+543639	J141724.33+543629.5	y	J225619.0+143257	J225621.96+143351.4	y
J141938.8+312146	J141940.16+312138.8	y	J232410.1+001315	J232410.15+001314.5	y
J142515.3+175526	J142513.89+175525.7	y	J234727.9-000919	J234727.65-000912.9	y

2804 H Radio luminosity function

2805 We computed the radio luminosity function following the $1/V_{\max}$ method (Schmidt, 1968).
 2806 We performed the following steps:

- 2807 1. Remove all radio sources that do not fit the selection criteria. This applies for
 2808 both radio and infrared properties, so we choose a minimum radio flux density
 2809 f_{\min} and a maximum infrared magnitude $m_{\max, \text{ir}}$, as well as redshift limits z_{lower}
 2810 and z_{upper} .
2. For each source, compute the maximum redshift that the source could have been observed within the selection criteria. We find this redshift by first numerically solving Equation 5.3 for z with L as the luminosity of each radio source and $f = f_{\min}$ to obtain the maximum redshift z_{radio} at which the source could be observed in radio. We similarly find the maximum redshift that the infrared host galaxy could be observed at within the selection criteria, z_{ir} , by numerically solving Equation 5.6. $d(z_{\text{ir}})$ is the luminosity distance at a redshift z_{ir} , d is the luminosity distance of the host galaxy, and m is the apparent magnitude of the host galaxy, all in the infrared.

$$2811 5 \log_{10} \left(\frac{d(z_{\text{ir}})}{d} \right) + m = m_{\max, \text{ir}} \quad (5.6)$$

2812 The maximum redshift that the source could have been observed within the selection criteria is then $z_{\max} = \min(z_{\text{ir}}, z_{\text{radio}}, z_{\text{upper}})$.

- 2813 3. For each source, compute the comoving volume V_{\max} at redshift z_{\max} .
- 2814 4. The count for each luminosity bin is the sum over $1/V_{\max}$ for each source in the
 2815 bin. We divided these counts by the estimated completeness (Appendix I) to
 2816 account for redshift incompleteness. We account for the fact FIRST does not cover
 2817 the whole sky by multiplying by the total area of the sky divided by the area of
 2818 our selection.

2819 After computing the luminosity function, we estimate the uncertainty in each bin
 2820 using Poisson statistics, \sqrt{N} for a bin count N .

2821 I Redshift completeness estimate

2822 Figure 5.7 shows the estimated completeness of our RLF sample in Chapter 5 as a
 2823 function of W_1 and $W_1 - W_2$. We followed the same method as Pracy et al. (2016)
 2824 for this estimation, averaging completeness over circles centred on each source. Each
 2825 source is associated with a circle of radius equal to the distance to its 50th nearest
 2826 neighbour in the W_1 and $W_1 - W_2$ plane. This appendix was originally part of Alger et al. (in prep.).
 2827

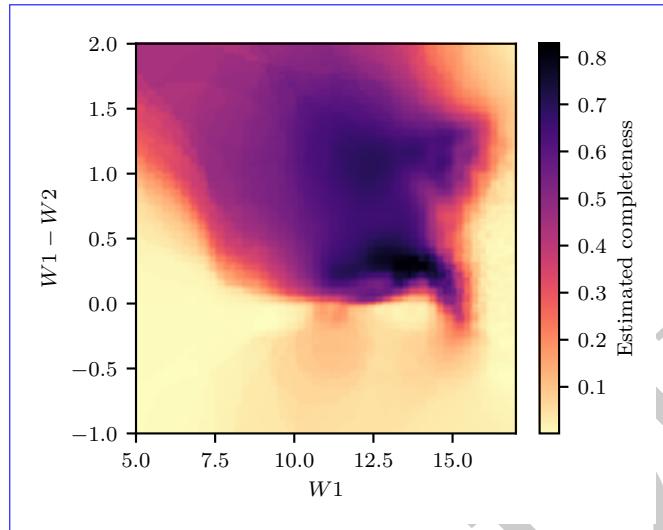


Figure 5.7: Estimated completeness as a function of mid-infrared colour and magnitude.

2828 J Giant radio galaxies

2829 This appendix describes our search for giant radio galaxies in RGZ-Ex, and the results
 2830 of this search. To identify radio sources we assumed that if any two components had
 2831 the same host galaxy then they were part of the same source. This is a reasonable
 2832 assumption if all host galaxies are correctly identified, which was not the case. This
 2833 assumption therefore introduced spurious sources due to galaxies incorrectly identified
 2834 as host galaxies: not all sources used in this chapter are real sources, and in particular
 2835 sources of large angular size are likely to be incorrect. Nevertheless RGZ-Ex provides a
 2836 useful catalogue of *candidate* radio sources, and visual follow-up can confirm whether
 2837 sources of interest are real.

2838 H.A. and M.J.A. examined all 296 candidate sources in the RGZ-Ex catalogue with
 2839 an estimated physical extent larger than 1 Mpc. Of these, 40 were real giant radio
 2840 galaxies, which we show in Table 5.3. We defined ‘giant radio galaxy’ as a radio galaxy
 2841 with emission extended to physical sizes ≥ 1.0 Mpc. Other thresholds, such as 0.7 Mpc,
 2842 also exist in literature. The physical extents of the remaining 256 candidate sources
 2843 were overestimated mostly due to sidelobes/artefacts (103), incorrect source grouping
 2844 (82), or incorrect SDSS matches (21). The citizen scientists who identified giants are:
 2845 WizardHowl, DolorousEdd, antikodon, csunjoto, sisifolibre, JeanTate, JKDF, PADV, and
 2846 firejuggler. H.A., together with his summer students, had previously identified 29 of
 2847 these giants.

2848 Note that this is a particularly challenging set: sources that are misidentified will
 2849 often have unusually large estimated extents due to the inclusion of spurious components.
 2850 The error rate in this set therefore does not reflect the rest of the catalogue.

Table 5.3: Giant radio galaxies found in RGZ-Ex. ‘LLS’ is the projected linear size of the source as measured by the maximum angular distance between radio components. The RA/Dec are the coordinates of the host galaxy. s/p indicates spectroscopic/photometric redshift. ^LExisting in literature. ^RAlso found by RGZ citizen scientists. [†]Misidentified SDSS host, manually corrected to obtain redshift.

AllWISE host (WISEA)	RA (J2000)	Dec (J2000)	z	LLS (Mpc)	
J004210.18-080011.3	10.54	-8.00	0.65 ± 0.14	1.6	p
J021008.48+011839.6 ^L	32.54	1.31	0.86524 ± 0.00001	1.2	s
J075858.29+355643.6 ^R	119.74	35.95	0.74748 ± 0.00013	1.0	s
J080831.68+473523.9 ^R	122.13	47.59	0.58854 ± 0.00016	1.1	s
J083034.78+231124.6	127.64	23.19	0.94 ± 0.13	1.1	p
J090604.03+011114.2	136.52	1.19	0.7975 ± 0.0004	1.6	s
J093256.81+074212.2	143.24	7.70	1.0032 ± 0.0003	1.1	s
J093526.80+051729.8 ^R	143.86	5.29	0.84 ± 0.04	1.2	p
J094238.72+114337.9	145.66	11.73	0.49 ± 0.05	1.2	p
J094835.60+535946.4 ^R	147.15	54.00	0.64 ± 0.10	1.2	p
J095706.12+292439.2	149.28	29.41	0.71 ± 0.12	1.5	p
J102335.25+433208.0	155.90	43.54	0.75 ± 0.09	1.5	p
J102933.99+210345.8 ^R	157.39	21.06	0.82407 ± 0.00008	1.1	s
J103043.98+355451.2 ^R	157.68	35.91	0.64074 ± 0.00008	1.2	s
J104449.92+234525.6 [†]	161.20	23.76	0.57712 ± 0.00009	1.6	s
J110655.98+624759.8 ^R	166.73	62.80	0.84379 ± 0.00004	1.1	s
J112900.68+635543.2	172.25	63.93	0.71 ± 0.06	1.1	p
J112948.20+243922.6	172.45	24.66	0.79 ± 0.07	1.1	p
J114553.67-003304.7	176.47	-0.55	2.0522 ± 0.0006	1.3	s
J121111.26+534840.4	182.80	53.81	0.74 ± 0.14	1.1	p
J121152.04+304232.4 ^R	182.97	30.71	0.47102 ± 0.00012	1.3	s
J121944.73+174121.3	184.94	17.69	1.5129 ± 0.0009	1.0	s
J123735.89+544814.4 ^R	189.40	54.80	1.0271 ± 0.0006	1.2	s
J123819.16+113444.8	189.58	11.58	0.80 ± 0.08	1.2	p
J123846.84-032857.5 [†]	189.70	-3.48	0.67 ± 0.07	1.5	p
J131625.00+272042.8	199.10	27.35	0.69092 ± 0.00004	1.0	s
J133307.00+045048.6 ^R	203.28	4.85	1.40534 ± 0.00016	1.1	s
J141933.36+104706.4 ^R	214.89	10.79	0.33973 ± 0.00003	1.0	s
J142008.45+185422.7 ^R	215.04	18.91	0.63 ± 0.04	1.4	p
J145057.28+530007.7 ^L	222.74	53.00	0.91662 ± 0.00009	1.3	s
J150012.18+604941.3	225.05	60.83	1.6626 ± 0.0007	1.2	s
J153547.13+432245.0 ^R	233.95	43.38	0.63891 ± 0.00007	1.3	s
J154631.18+194819.9	236.63	19.81	0.5917 ± 0.0002	1.4	s
J160852.10+561110.2 ^R	242.22	56.19	1.3196 ± 0.0003	1.3	s
J162200.48+364044.0	245.50	36.68	1.9994 ± 0.0002	1.1	s
J163004.35+103321.9 ^R	247.52	10.56	0.85 ± 0.09	1.2	p
J163125.75+200224.1 ^R	247.86	20.04	0.62662 ± 0.00013	1.0	s
J165055.46+394446.6	252.73	39.75	0.58829 ± 0.00013	1.1	s
J232410.33+045309.6	351.04	4.89	0.76 ± 0.06	1.4	p
J234440.02-003231.6	356.17	-0.54	0.5014 ± 0.0001	1.0	s

2852

Faraday Complexity

2853 This chapter is based on my paper *Interpretable Faraday Complexity Classification*, by M.
2854 J. Alger, J. D. Livingston, N. M. McClure-Griffiths, J. L. Nabaglo, O. I. Wong, and C.
2855 S. Ong; accepted and to be published by the *Publications of the Astronomical Society of
2856 Australia*.

2857 In the last two chapters we developed and applied a way of automatically cross-
2858 identifying radio sources, which will allow us to quickly extract useful information
2859 from wide-area radio continuum surveys like EMU. These are not the only radio sur-
2860 veys that will be conducted in the lead-up to the SKA, however. Polarisation sky sur-
2861 veys like POSSUM can provide very different information about the radio sky. One
2862 piece of information that could be determined from polarised observations is the Far-
2863 day complexity of a radio source. A complex source may be separated from us as ob-
2864 servers by some magnetised structure, or it might have extended structure across the
2865 sky, even if we cannot resolve the source spatially. This tells us something about the
2866 intervening Universe or the physical structure of the source, respectively. This chapter
2867 develops a new machine learning method of classifying sources as Faraday complex
2868 or simple.

2869 Faraday complexity describes whether a spectropolarimetric observation has sim-
2870 ple or complex magnetic structure. Quickly determining the Faraday complexity of a
2871 spectropolarimetric observation is important for processing large, polarised radio sur-
2872 veys. Finding simple sources lets us build rotation measure grids, and finding complex
2873 sources lets us follow these sources up with slower analysis techniques or further ob-
2874 servations. We introduce five features that can be used to train simple, interpretable
2875 machine learning classifiers for estimating Faraday complexity. We train logistic re-
2876 gression and extreme gradient boosted tree classifiers on simulated polarised spectra
2877 using our features, analyse their behaviour, and demonstrate our features are effective
2878 for both simulated and real data. This is the first application of machine learning meth-
2879 ods to real spectropolarimetry data. With 95 per cent accuracy on simulated ASKAP
2880 data and 90 per cent accuracy on simulated ATCA data, our method performs compa-
2881 rably to state-of-the-art convolutional neural networks while being simpler and easier
2882 to interpret. Logistic regression trained with our features behaves sensibly on real data
2883 and its outputs are useful for sorting polarised sources by apparent Faraday comple-
2884 xity.

2885 1 Introduction

2886 As polarised radiation from distant galaxies makes its way to us, magnetised plasma
 2887 along the way can cause the polarisation angle to change due to the Faraday effect.
 2888 The amount of rotation depends on the squared wavelength of the radiation, and the
 2889 rotation per squared wavelength is called the Faraday depth. Multiple Faraday depths
 2890 may exist along one line-of-sight, and if a polarised source is observed at multiple
 2891 wavelengths then these multiple depths can be disentangled. This can provide insight
 2892 into the polarised structure of the source or the intervening medium.

2893 Faraday rotation measure synthesis (RM synthesis) is a technique for decomposing
 2894 a spectropolarimetric observation into flux at its Faraday depths ϕ , the resulting distri-
 2895 bution of depths being called a *Faraday dispersion function* (FDF) or a *Faraday spectrum*.
 2896 It was introduced by Brentjens and de Bruyn (2005) as a way to rapidly and reliably
 2897 analyse the polarisation structure of complex and high-Faraday depth polarised obser-
 2898 vations.

2899 A *Faraday simple* observation is one for which there is only one Faraday depth, and
 2900 in this simple case the Faraday depth is also known as a *rotation measure* (RM). All Far-
 2901 aday simple observations can be modelled as a polarised source with a thermal plasma
 2902 of constant electron density and magnetic field (a ‘Faraday screen’; Anderson et al.,
 2903 2015; Brentjens & de Bruyn, 2005) between the observer and the source. A *Faraday*
 2904 *complex* observation is one which is not Faraday simple, and may differ from a Faraday
 2905 simple source due to plasma emission or composition of multiple screens (Brentjens
 2906 & de Bruyn, 2005). The complexity of a source tells us important details about the
 2907 polarised structure of the source and along the line-of-sight, such as whether the in-
 2908 tervening medium emits polarised radiation, or whether there are turbulent magnetic
 2909 fields or different electron densities in the neighbourhood. The complexity of nearby
 2910 sources taken together can tell us about the magneto-ionic structure of the galactic and
 2911 intergalactic medium between the sources and us as observers. O’Sullivan et al. (2017)
 2912 show examples of simple and complex sources, and Figure 6.1 and Figure 6.2 show an
 2913 example of a simulated simple and complex FDF respectively.

2914 Identifying when an observation is Faraday complex is an important problem in po-
 2915 larised surveys (Sun et al., 2015), and with current surveys such as the Polarised Sky
 2916 Survey of the Universe’s Magnetism (POSSUM) larger than ever before, methods that
 2917 can quickly characterise Faraday complexity en masse are increasingly useful. Being
 2918 able to identify which sources are simple lets us produce a reliable rotation measure
 2919 grid from background sources, and being able to identify which sources might be com-
 2920 plex allows us to find sources to follow-up with slower polarisation analysis methods
 2921 that may require manual oversight, such as QU fitting (as seen in e.g. Miyashita et al.,
 2922 2019; O’Sullivan et al., 2017). In this [paper](#)[chapter](#), we introduce five simple, inter-
 2923 pretable features representing polarised spectra, use these features to train machine
 2924 learning classifiers to identify Faraday complexity, and demonstrate their effective-
 2925 ness on real and simulated data. We construct our features by comparing observed
 2926 polarised sources to idealised polarised sources. The features are intuitive and can be
 2927 estimated from real FDFs.

2928 Section 2 provides a background to our work, including a summary of prior work
2929 and our assumptions on FDFs. Section 3 describes our approach to the Faraday com-
2930 plexity problem. Section 4 explains how we trained and evaluated our method. Finally,
2931 Section 5 discusses these results.

2932 2 Faraday Complexity

2933 Faraday complexity is an observational property of a source: if multiple Faraday depths
2934 are observed within the same apparent source (e.g. due to multiple lines-of-sight being
2935 combined within a beam), then the source is complex. A source composed of multiple
2936 Faraday screens may produce observations consistent with many models (Sun et al.,
2937 2015), including simple sources, so there is some overlap between simple and complex
2938 sources. Faraday thickness is also a source of Faraday complexity: when the interven-
2939 ing medium between a polarised source and the observer also emits polarised light, the
2940 FDF cannot be characterised by a simple Faraday screen. As discussed in Section 2.2
2941 we defer Faraday thick sources to future work. In this section we summarise existing
2942 methods of Faraday complexity estimation and explain our assumptions and model of
2943 simple and complex polarised FDFs.

2944 2.1 Prior work

2945 There are multiple ways to estimate Faraday complexity, including detecting non-
2946 linearity in $\chi(\lambda^2)$ (Goldstein & Reed, 1984), change in fractional polarisation as a func-
2947 tion of frequency (Farnes et al., 2014), non-sinusoidal variation in fractional polari-
2948 sation in Stokes Q and U (O’Sullivan et al., 2012), counting components in the FDF
2949 (Law et al., 2011), minimising the Bayesian information criterion (BIC) over a range of
2950 simple and complex models (called *QU fitting*; O’Sullivan et al., 2017), the method of
2951 Faraday moments (Anderson et al., 2015; Brown, 2011), and deep convolutional neural
2952 network classifiers (CNNs; Brown et al., 2018). See Sun et al. (2015) for a comparison
2953 of these methods.

2954 The most common approaches to estimating complexity are QU fitting
2955 (e.g. O’Sullivan et al., 2017) and Faraday moments (e.g. Anderson et al., 2015). To our
2956 knowledge there is currently no literature examining the accuracy of QU fitting when
2957 applied to complexity classification specifically, though Miyashita et al. (2019) analyse
2958 its effectiveness on identifying the structure of two-component sources. Brown (2011)
2959 suggested Faraday moments as a method to identify complexity, a method later used
2960 by Farnes et al. (2014) and Anderson et al. (2015), but again no literature examines
2961 the accuracy. CNNs are the current state-of-the-art with an accuracy of 94.9 per cent
2962 (Brown et al., 2018) on simulated ASKAP Band 1 and 3 data, and we will compare our
2963 results to this method.

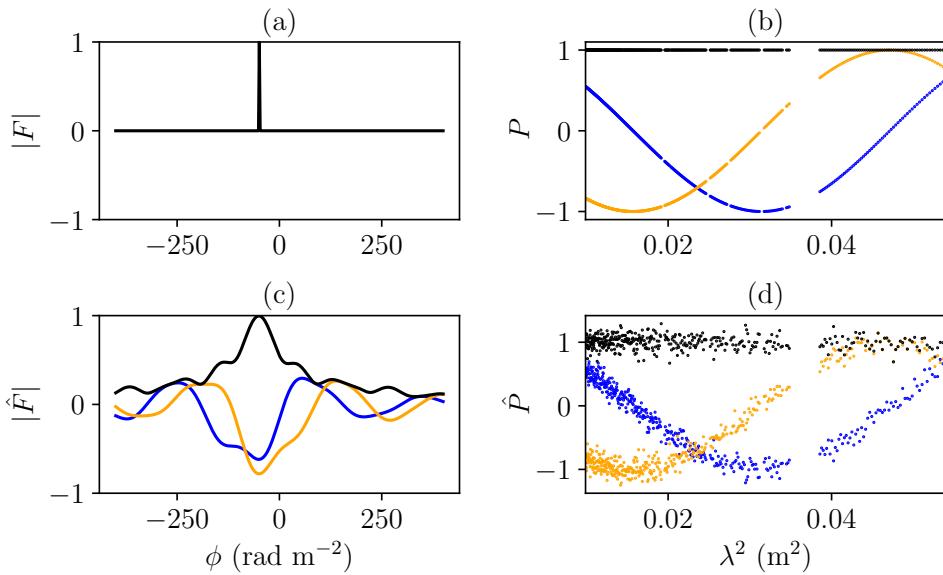


Figure 6.1: A simple FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.

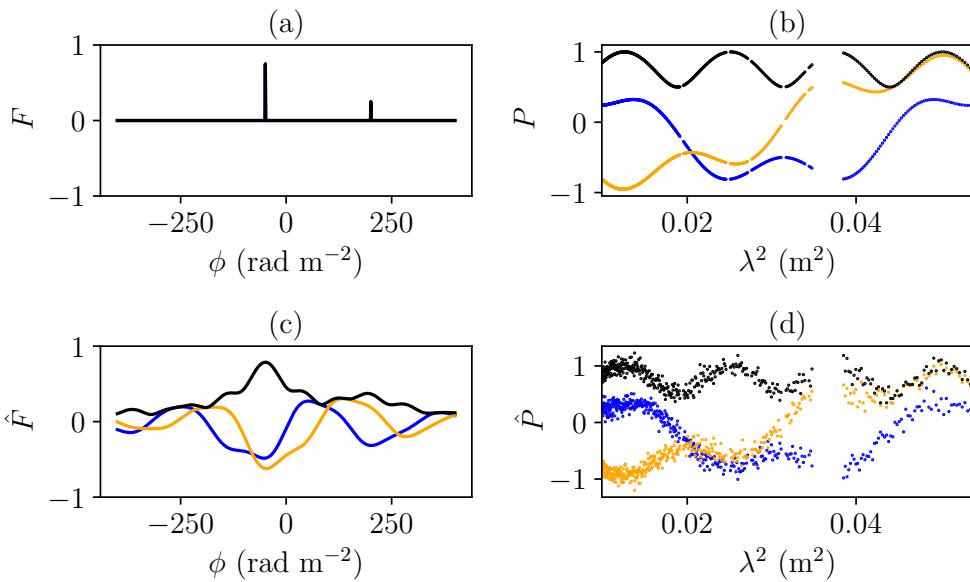


Figure 6.2: A complex FDF and its corresponding polarised spectra: (a) groundtruth FDF F , (b) noise-free polarised spectrum P , (c) noisy observed FDF \hat{F} , (d) noisy polarised spectrum \hat{P} . Blue and orange mark real and imaginary components respectively.

2964 **2.2 Assumptions on Faraday dispersion functions**

Before we can classify FDFs as Faraday complex or Faraday simple, we need to define FDFs and any assumptions we make about them. An FDF is a function that maps Faraday depth ϕ to complex polarisation. It is the distribution of Faraday depths in an observed polarisation spectrum. For a given observation, we assume there is a true, noise-free FDF F composed of at most two Faraday screens. This accounts for most actual sources (Anderson et al., 2015) and extension to three screens would cover most of the remainder—O’Sullivan et al. (2017) found that 89 per cent of their sources were best explained by two or less screens, while the remainder were best explained by three screens. We model the screens by Dirac delta distributions:

$$F(\phi) = A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1). \quad (6.1)$$

A_0 and A_1 are the polarised flux of each Faraday screen, and ϕ_0 and ϕ_1 are the Faraday depths of the respective screens. With this model, a Faraday simple source is one which has $A_0 = 0$, $A_1 = 0$, or $\phi_0 = \phi_1$. By using delta distributions to model each screen, we are assuming that there is no internal Faraday dispersion (which is typically associated with diffuse emission rather than the mostly-compact sources we expect to find in wide-area polarised surveys). F generates a polarised spectrum of the form shown in Equation 6.2:

$$P(\lambda^2) = A_0 e^{2i\phi_0 \lambda^2} + A_1 e^{2i\phi_1 \lambda^2}. \quad (6.2)$$

Such a spectrum would be observed as noisy samples from a number of squared wavelengths $\lambda_j^2, j \in [1, \dots, D]$. We model this noise as a complex Gaussian with standard deviation σ and call the noisy observed spectrum \hat{P} :

$$\hat{P}(\lambda_j^2) \sim \mathcal{N}(P(\lambda_j^2), \sigma^2). \quad (6.3)$$

The constant variance of the noise is a simplifying assumption which may not hold for real data, and exploring this is a topic for future work. By performing RM synthesis (Brentjens & de Bruyn, 2005) on \hat{P} with uniform weighting we arrive at an observed FDF:

$$\hat{F}(\phi) = \frac{1}{D} \sum_{j=1}^D \hat{P}(\lambda_j^2) e^{-2i\phi\lambda_j^2}. \quad (6.4)$$

2965 Examples of F , \hat{F} , P , and \hat{P} for simple and complex observations are shown in Figure 6.1
2966 and Figure 6.2 respectively. Note that there are two reasons that the observed FDF \hat{F}
2967 does not match the groundtruth FDF F . The first is the noise in \hat{P} . The second arises
2968 from the incomplete sampling of \hat{P} .

2969 We do not consider external or internal Faraday dispersion in this work. External
2970 Faraday dispersion would broaden the delta functions of Equation 6.1 into peaks, and
2971 internal Faraday dispersion would broaden them into top-hat functions. All sources
2972 have at least a small amount of dispersion as the Faraday depth is a bulk property of
2973 the intervening medium and is subject to noise, but the assumption we make is that
2974 this dispersion is sufficiently small that the groundtruth FDFs are well-modelled with

2975 delta functions. Faraday thick sources would also invalidate our assumptions, and we
 2976 assume that there are none in our data as Faraday thickness can be consistent with a
 2977 two-component model depending on the wavelength sampling (e.g. Brentjens & de
 2978 Bruyn, 2005; Ma et al., 2019a). Nevertheless some external Faraday dispersion would
 2979 be covered by our model, as depending on observing parameters Faraday thick sources
 2980 may appear as two screens (Van Eck et al., 2017).

2981 To simulate observed FDFs we follow the method of Brown et al. (2018), which we
 2982 describe in Appendix K.

2983 3 Classification approach

2984 The Faraday complexity classification problem is as follows: Given an FDF \hat{F} , is it Far-
 2985 day complex or Faraday simple? In this section we describe the features that we have
 2986 developed to address this problem, which can be used in any standard machine learn-
 2987 ing classifier. We trained two classifiers on these features, which we describe here also.

2988 3.1 Features

Our features are based on a simple idea: all simple FDFs look essentially the same, up to scaling and translation, while complex FDFs may deviate. A noise-free peak-normalised simple FDF \hat{F}_{simple} has the form

$$\hat{F}_{\text{simple}}(\phi; \phi_s) = R(\phi - \phi_s). \quad (6.5)$$

where R is the rotation measure spread function (RMSF), the Fourier transform of the wavelength sampling function which is 1 at all observed wavelengths and 0 otherwise. ϕ_s traces out a curve in the space of all possible FDFs. In other words, \hat{F}_{simple} is a manifold parametrised by ϕ_s . Our features are derived from relating an observed FDF to the manifold of simple FDFs (the ‘simple manifold’). We measure the distance of an observed FDF to the simple manifold using distance measure D_f , that take all values of the FDF into account:

$$\xi_f(\hat{F}) = \min_{\phi_s \in \mathbb{R}} D_f(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_s)). \quad (6.6)$$

2989 We propose two distances that have nice properties:

- 2990 • invariant over changes in complex phase,
- 2991 • translationally invariant in Faraday depth,
- 2992 • zero for Faraday simple sources (i.e. when $A_0 = 0, A_1 = 0$, or $\phi_0 = \phi_1$) when
 2993 there is no noise,
- 2994 • symmetric in components (i.e. swapping $A_0 \leftrightarrow A_1$ and $\phi_0 \leftrightarrow \phi_1$ should not
 2995 change the distance),

- 2996 • increasing as A_0 and A_1 become closer to each other, and
 2997 • increasing as screen separation $|\phi_0 - \phi_1|$ increases over a large range.

2998 Our features are constructed from this distance and its minimiser. In other words we
 2999 look for the simple FDF \hat{F}_{simple} that is “closest” to the observed FDF \hat{F} . The minimiser
 3000 ϕ_s is the Faraday depth of the simple FDF.

3001 While we could choose any distance that operates on functions, we used the 2-
 3002 Wasserstein (W_2) distance (Equation 6.7) and the Euclidean distance (Equation 6.9).
 3003 The W_2 distance operates on probability distributions and can be thought of as the min-
 3004 imum cost to ‘move’ one probability distribution to the other, where the cost of moving
 3005 one unit of probability mass is the squared distance it is moved. Under W_2 distance,
 3006 the minimiser ϕ_w in Equation 6.6 can be interpreted as the Faraday depth that the FDF
 3007 \hat{F} would be observed to have if its complexity was unresolved (i.e. the weighted mean
 3008 of its components). The Euclidean distance is the square root of the least-squares loss
 3009 which is often used for fitting \hat{F}_{simple} to the FDF \hat{F} . Under Euclidean distance, the
 3010 minimiser ϕ_s is equivalent to the depth of the best-fitting single component under as-
 3011 sumption of Gaussian noise in \hat{F} . We calculated the W_2 distance using Python Optimal
 3012 Transport (Flamary & Courty, 2017), and we calculated the Euclidean distance using
 3013 `scipy.spatial.distance.euclidean` (Virtanen et al., 2020). Further intuition about
 3014 the two distances is provided in Section 3.2.

We denote by ϕ_w and ϕ_e , the Faraday depth of the simple FDF that minimises the
 respective distances (2-Wasserstein and Euclidean).

$$\begin{aligned}\phi_w &= \underset{\phi_w}{\operatorname{argmin}} D_{W_2}(\hat{F}(\phi) \| \hat{F}_{\text{simple}}(\phi; \phi_w)), \\ \phi_e &= \underset{\phi_e}{\operatorname{argmin}} D_E(\hat{F}(\phi) \| \hat{F}_{\text{simple}}(\phi; \phi_e)).\end{aligned}$$

These features are depicted on an example FDF in Figure 6.3. For simple observed
 FDFs, the fitted Faraday depths ϕ_w and ϕ_e both tend to be close to the peak of the
 observed FDF. However for complex observed FDFs, ϕ_w tends to be at the average
 depth between the two major peaks of the observed FDF, being closer to the higher
 peak. For notation convenience, we denote the Faraday depth of the observed FDF
 that has largest magnitude as ϕ_a , i.e.

$$\phi_a = \underset{\phi_a}{\operatorname{argmax}} |\hat{F}(\phi_a)|,$$

3015 Note that in practice $\phi_a \approx \phi_e$. For complex observed FDFs, the values of Faraday
 3016 depths ϕ_w and ϕ_a tend to differ (essentially by a proportion of the location of the sec-
 3017 ond screen). The difference between ϕ_w and ϕ_a therefore provides useful information
 3018 to identify complex FDFs. When the observed FDF is simple, the 2-Wasserstein fit will
 3019 overlap significantly, hence the observed magnitudes $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$ will be similar.
 3020 However, for complex FDFs ϕ_w and ϕ_a are at different depths, leading to different val-
 3021 ues of $\hat{F}(\phi_w)$ and $\hat{F}(\phi_a)$. Therefore the magnitudes of the observed FDFs at the depths
 3022 ϕ_w and ϕ_a indicate how different the observed FDF is from a simple FDF.

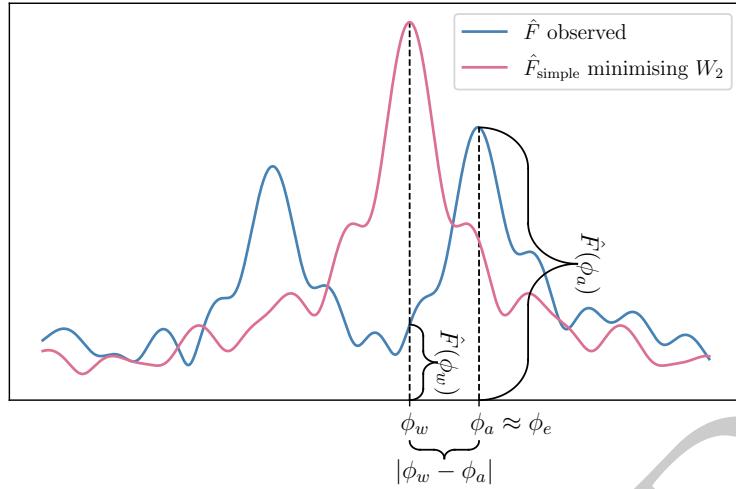


Figure 6.3: An example of how an observed FDF \hat{F} relates to our features. ϕ_w is the W_2 -minimising Faraday depth, and ϕ_a is the \hat{F} -maximising Faraday depth (approximately equal to the Euclidean-minimising Faraday depth). The remaining two features are the W_2 and Euclidean distances between the depicted FDFs.

3023 In summary, we provide the following features to the classifier:

- 3024 • $\log |\phi_w - \phi_a|$,
- 3025 • $\log \hat{F}(\phi_w)$,
- 3026 • $\log \hat{F}(\phi_a)$,
- 3027 • $\log D_{W_2}(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_w))$,
- 3028 • $\log D_E(\hat{F}(\phi) \parallel \hat{F}_{\text{simple}}(\phi; \phi_e))$,

3029 where D_E is the Euclidean distance, D_{W_2} is the W_2 distance, ϕ_a is the Faraday depth of
 3030 the FDF peak, ϕ_w is the minimiser for W_2 distance, and ϕ_e is the minimiser for Euclidean
 3031 distance.

3032 3.2 Interpreting distances

Interestingly, in the case where there is no RMSF, Equation 6.6 with W_2 distance reduces to the Faraday moment already in common use:

$$D_{W_2}(F) = \min_{\phi_w \in \mathbb{R}} D_{W_2}(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_w)) \quad (6.7)$$

$$= \left(\frac{A_0 A_1}{(A_0 + A_1)^2} (\phi_0 - \phi_1)^2 \right)^{1/2}. \quad (6.8)$$

See Appendix L for the corresponding calculation. In this sense, the W_2 distance can be thought of as a generalised Faraday moment, and conversely an interpretation of

Faraday moments as a distance from the simple manifold in the case where there is no RMSF. Euclidean distance behaves quite differently in this case, and the resulting distance measure is totally independent of Faraday depth:

$$D_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \quad (6.9)$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (6.10)$$

3033 See Appendix M for the corresponding calculation.

3034 3.3 Classifiers

3035 We trained two classifiers on simulated observations using these features: logistic re-
3036 gression (LR) and extreme gradient boosted trees (XGB). These classifiers are useful
3037 together for understanding Faraday complexity classification. LR is a linear classi-
3038 fier that is readily interpretable by examining the weights it applies to each feature,
3039 and is one of the simplest possible classifiers. XGB is a powerful off-the-shelf non-
3040 linear ensemble classifier, and is an example of a decision tree ensemble which are
3041 widely used in astronomy (e.g. Hložek et al., 2020; Machado Poletti Valle et al., 2020).
3042 We used the `scikit-learn` implementation of LR and we use the `XGBoost` library
3043 for XGB. We optimised hyperparameters for XGB using a fork of `xgboost-tuner`¹
3044 as utilised by Zhu et al. (2020). We used 1 000 iterations of randomised parameter
3045 tuning and the hyperparameters we found are tabulated in Table 6.2. We optimised
3046 hyperparameters for LR using a 5-fold cross-validation grid search implemented in
3047 `sklearn.model_selection.GridSearchCV`. The resulting hyperparameters are tabu-
3048 lated in Table 6.3 in the Appendix Appendix N.

3049 4 Experimental method and results

3050 We applied our classifiers to classify simulated (Section 4.2 and 4.3) and real (Sec-
3051 tion 4.4) FDFs. We replicated the experimental setup of Brown et al. (2018) for com-
3052 parison with the state-of-the-art CNN classification method, and we also applied our
3053 method to 142 real FDFs observed with the Australia Telescope Compact Array (ATCA)
3054 from Livingston et al. (2021) and O’Sullivan et al. (2017).

3055 4.1 Data

3056 4.1.1 Simulated training and validation data

3057 Our classifiers were trained and validated on simulated FDFs. We produced two sets
3058 of simulated FDFs, one for comparison with the state-of-the-art method in the litera-
3059 ture and one for application to our observed FDFs (described in Section 4.1.2). We
3060 refer to the former as the ‘ASKAP’ dataset as it uses frequencies from the Australian

¹<https://github.com/chengsoonong/xgboost-tuner>

3061 Square Kilometre Array Pathfinder 12-antenna early science configuration. These fre-
 3062 quencies included 900 channels from 700–1300 and 1500–1800 MHz and were used to
 3063 generate simulated training and validation data by Brown et al. (2018). We refer to
 3064 the latter as the ‘ATCA’ dataset as it uses frequencies from the 1–3 GHz configuration
 3065 of the ATCA. These frequencies included 394 channels from 1.29–3.02 GHz and match
 3066 our real data. We simulated Faraday depths from -50 to 50 rad m^{-2} for the ‘ASKAP’
 3067 dataset (matching Brown) and -500 to 500 for the ‘ATCA’ dataset.

For each dataset, we simulated 100 000 FDFs, approximately half simple and half complex. We randomly allocated half of these FDFs to a training set and reserved the remaining half for validation. Each FDF had complex Gaussian noise added to the corresponding polarisation spectrum. For the ‘ASKAP’ dataset, we sampled the standard deviation of the noise uniformly between 0 and $\sigma_{\max} = 0.333$, matching the dataset of Brown et al. (2018). For the ‘ATCA’ dataset, we fit a log-normal distribution to the standard deviations of O’Sullivan’s data (O’Sullivan et al., 2017) from which we sampled our values of σ :

$$\sigma \sim \frac{1}{0.63\sqrt{2\pi}\sigma} \exp\left(-\frac{\log(50\sigma - 0.5)^2}{2 \times 0.63^2}\right) \quad (6.11)$$

3068 4.1.2 Observational data

3069 We used two real datasets containing a total of 142 sources: 42 polarised spectra from
 3070 Livingston et al. (2020, submitted) and 100 polarised spectra from O’Sullivan et al.
 3071 (2017). These datasets were observed in similar frequency ranges on the same tele-
 3072 scope (with different binning), but are in different parts of the sky. The Livingston
 3073 data were taken near the Galactic Centre, and the O’Sullivan data were taken away
 3074 from the plane of the Galaxy. There are more Faraday complex sources near the Galac-
 3075 tic Centre compared to more Faraday simple sources away from the plane of the Galaxy
 3076 (Livingston et al.). The similar frequency channels used in the two datasets result in
 3077 almost identical RMSFs over the Faraday depth range we considered (-500 to 500 rad
 m^{-2}), so we expected that the classifiers would work equally well on both datasets
 3078 with no need to re-train. We discarded the 26 Livingston sources with modelled Far-
 3079 day depths outside of this Faraday depth range, which we do not expect to affect the
 3080 applicability of our methods to wide-area surveys because these fairly high depths are
 3081 not common.

3082 Livingston et al. (2021) used RM-CLEAN (Heald, 2008) to identify significant com-
 3083 ponents in their FDFs. Some of these components had very high Faraday depths up to
 3084 2000 rad m^{-2} , but we chose to ignore these components in this [paper](#)-[chapter](#) as they
 3085 are much larger than might be expected in a wide-area survey like POSSUM. They used
 3086 the second Faraday moment (Brown, 2011) to estimate Faraday complexity, with Far-
 3087 day depths determined using `scipy.signal.find_peaks` on the cleaned FDFs, with
 3088 a cutoff of 7 times the noise of the polarised spectrum. Using this method, they es-
 3089 timated that 89 per cent of their sources were Faraday complex i.e. had a Faraday
 3090 moment greater than 0.

3092 O’Sullivan et al. (2017) used the QU-fitting and model selection technique described
 3093 in O’Sullivan et al. (2012). The QU-fitting models contained up to three Faraday screen
 3094 components as well as a term for internal and external Faraday dispersion. We ignore
 3095 the Faraday thickness and dispersion for the purposes of this paperchapter, as most
 3096 sources were not found to have Faraday thickness and dispersion is beyond the scope
 3097 of our current work. 37 sources had just one component, 52 had two, and the remaining
 3098 11 had three.

3099 4.2 Results on ‘ASKAP’ dataset

Table 6.1: Confusion matrix entries for LR and XGB on ‘ASKAP’ and ‘ATCA’ simulated datasets, and the CNN confusion matrix entries adapted from Brown et al. (2018).

	‘ASKAP’			‘ATCA’	
	LR	XGB	CNN	LR	XGB
True negative rate	0.99	0.99	0.97	0.92	0.91
False positive rate	0.01	0.01	0.03	0.08	0.09
False negative rate	0.10	0.09	0.07	0.16	0.10
True positive rate	0.90	0.91	0.93	0.84	0.90

3100 The accuracy of the LR and XGB classifiers on the ‘ASKAP’ testing set was 94.4 and
 3101 95.1 per cent respectively. The rates of true and false identifications are summarised
 3102 in Table 6.1. These results are very close to the CNN presented by Brown et al. (2018),
 3103 with a slightly higher true negative rate and a slightly lower true positive rate (recall-
 3104 ing that positive sources are complex, and negative sources are simple). The accuracy
 3105 of the CNN was 94.9, slightly lower than our XGB classifier and slightly higher than
 3106 our LR classifier. Both of our classifiers therefore produce similar classification perfor-
 3107 mance to the CNN, with faster training time and easier interpretation.

3108 4.3 Results on ‘ATCA’ dataset

3109 The accuracy of the LR and XGB classifiers on the ‘ATCA’ dataset was 89.2 and 90.5
 3110 per cent respectively. The major differences between the ‘ATCA’ and the ‘ASKAP’ ex-
 3111 periments are the range of the simulated Faraday depths and the distribution of noise
 3112 levels. The ‘ASKAP’ dataset, to match past CNN work, only included depths from
 3113 -50 to 50 rad m^{-2} , while the ‘ATCA’ dataset includes depths from -500 to 500 rad
 3114 m^{-2} . The rates of true and false identifications are again shown in Table 6.1.

3115 As we know the true Faraday depths of the components in our simulation, we can
 3116 investigate the behaviour of these classifiers as a function of physical properties. Fig-
 3117 ure 6.4 shows the mean classifier prediction as a function of component depth separa-
 3118 tion and minimum component amplitude. This is tightly related to the mean accuracy,
 3119 as the entire plot domain contains complex spectra besides the left and bottom edge:
 3120 by thresholding the classifier prediction to a certain value, the accuracy will be one
 3121 hundred per cent on the non-edge for all sources with higher prediction values.

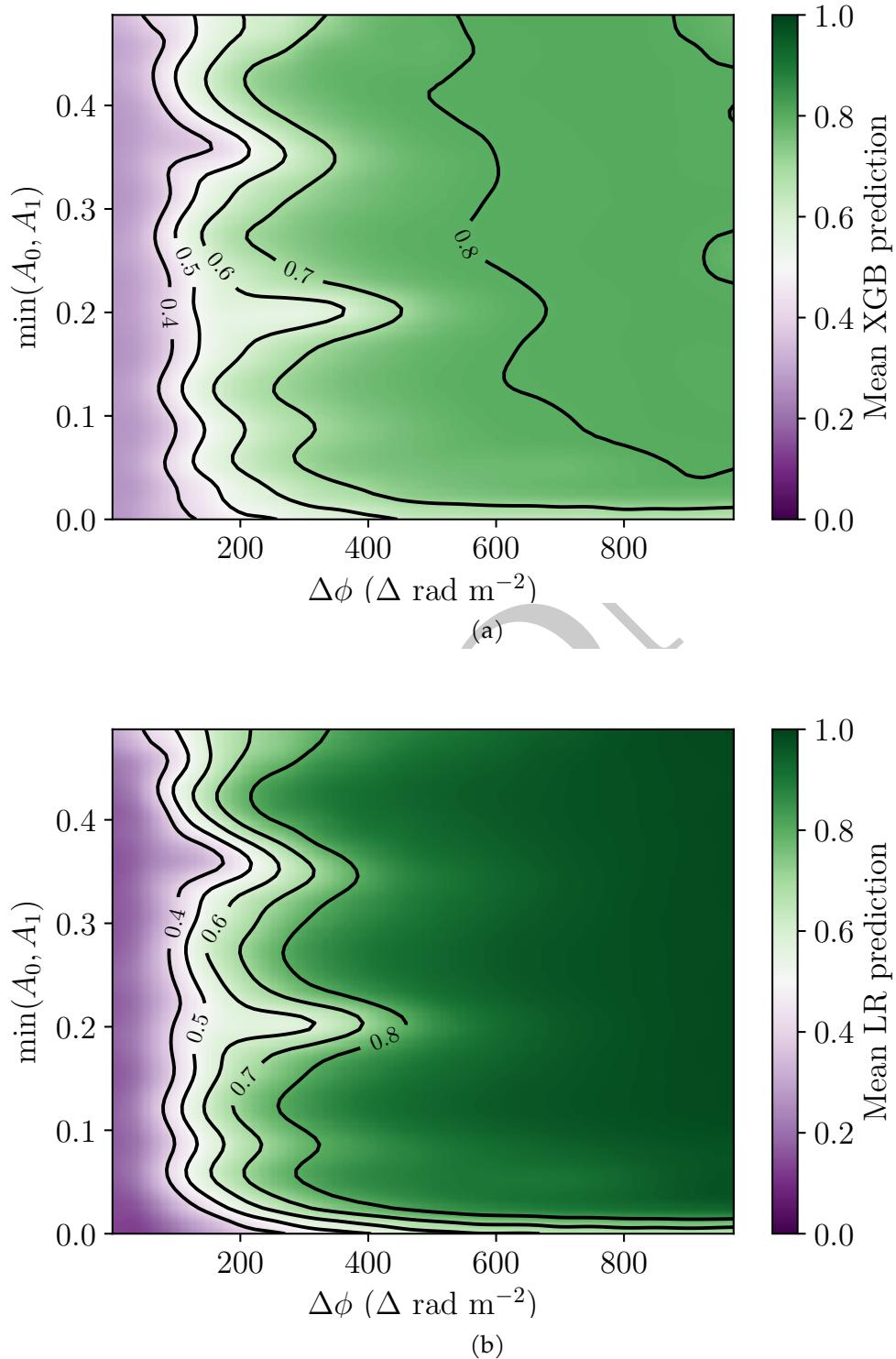


Figure 6.4: Mean prediction as a function of component depth separation and minimum component amplitude for (a) XGB and (b) LR.

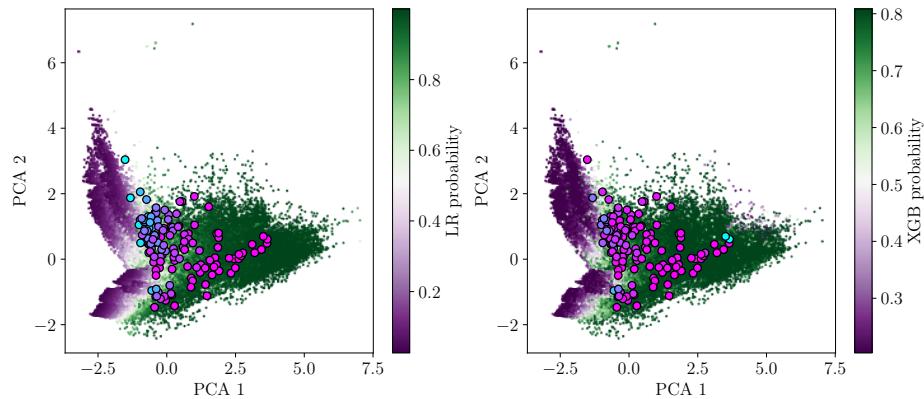


Figure 6.5: Principal component analysis for simulated data (coloured dots) with observations overlaid (black-edged circles). Observations are coloured by their XGB or LR estimated probability of being complex, with blue indicating ‘most simple’ and pink indicating ‘most complex’.

3122 4.4 Results on observed FDFs

3123 We used the LR and XGB classifiers which were trained on the ‘ATCA’ dataset to esti-
 3124 mate the probability that our 142 observed FDFs (Section 4.1.2) were Faraday complex.
 3125 As these classifiers were trained on simulated data, they face the issue of the ‘domain
 3126 gap’: the distribution of samples from a simulation differs from the distribution of real
 3127 sources, and this affects performance on real data. Solving this issue is called ‘domain
 3128 adaptation’ and how to do this is an open research question in machine learning (Pan
 3129 & Yang, 2010; Zhang, 2020). Nevertheless, the features of our observations mostly fall
 3130 in the same region of feature space as the simulations (Figure 6.5) and so we expect
 3131 reasonably good domain transfer.

3132 Two apparently complex sources in the Livingston sample are classified as simple
 3133 with high probability by XGB. These outliers are on the very edge of the training sam-
 3134 ple (Figure 6.5) and the underdensity of training data here is likely the cause of this
 3135 issue. LR does not suffer the same issue, producing plausible predictions for the entire
 3136 dataset, and these sources are instead classified as complex with high probability.

3137 With a threshold of 0.5, LR predicted that 96 and 83 per cent of the Livingston and
 3138 O’Sullivan sources were complex respectively. This is in line with expectations that the
 3139 Livingston data should have more Faraday complex sources than the O’Sullivan data
 3140 due to their location near the Galactic Centre. XGB predicted that 93 and 100 per cent
 3141 of the Livingston and O’Sullivan sources were complex respectively. Livingston et al.
 3142 (2021) found that 90 per cent of their sources were complex, and O’Sullivan et al. (2017)
 3143 found that 64 per cent of their sources were complex. This suggests that our classifiers
 3144 are overestimating complexity, though it could also be the case that the methods used
 3145 by Livingston and O’Sullivan underestimate complexity. Modifying the prediction
 3146 threshold from 0.5 changes the estimated rate of Faraday complexity, and we show

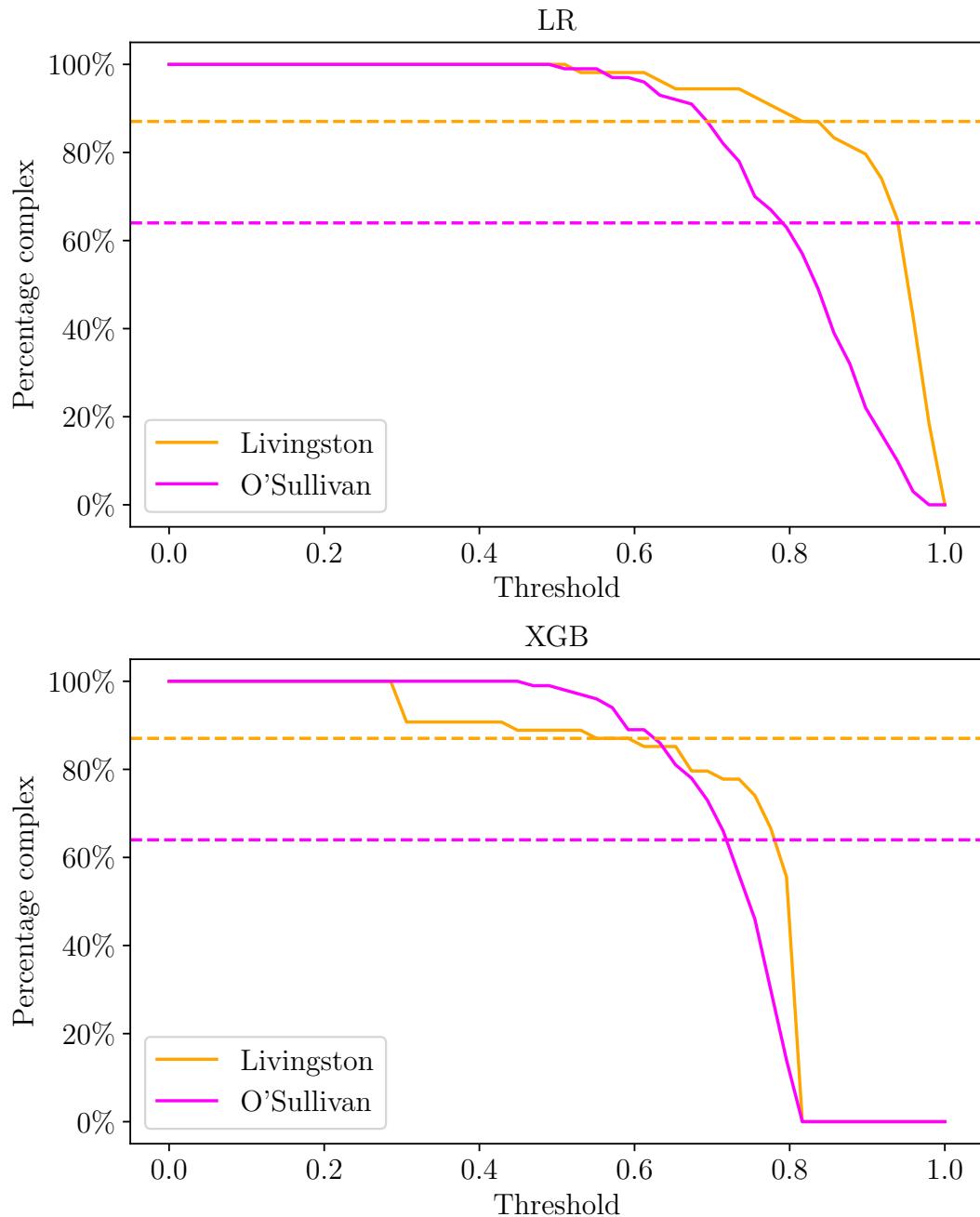


Figure 6.6: Estimated rates of Faraday complexity for the Livingston and O'Sullivan datasets as functions of threshold. The horizontal lines indicate the rates of Faraday complexity estimated by Livingston and O'Sullivan respectively.

3147 the estimated rates against threshold for both classifiers in Figure 6.6. We suggest that
3148 this result is indicative of our probabilities being uncalibrated, and a higher threshold
3149 should be chosen in practice. We chose to keep the threshold at 0.5 as this had the
3150 highest accuracy on the simulated validation data. The very high complexity rates of
3151 XGB and two outlying classifications indicate that the XGB classifier may be overfitting
3152 to the simulation and that it is unable to generalise across the domain gap.

3153 Figure 6.7 and Figure 6.8 in Appendix O show every observed FDF ordered by esti-
3154 mated Faraday complexity, alongside the models predicted by Livingston and O’Sullivan
3155 et al. (2017), for LR and XGB respectively. There is a clear visual trend of increasingly
3156 complex sources with increasing predicted probability of being complex.

3157 5 Discussion

3158 On simulated data (Section 4.3) we achieve state-of-the-art accuracy. Our results on
3159 observed FDFs show that our classifiers produce plausible results, with Figure 6.7 and
3160 Figure 6.8 showing a clear trend of apparent complexity. Some issues remain: we
3161 discuss the intrinsic overlap between simple and complex FDFs in Section 5.1 and the
3162 limitations of our method in Section 5.2.

3163 5.1 Complexity and seeming ‘not simple’

3164 Through this work we found our methods limited by the significant overlap between
3165 complex and simple FDFs. Complex FDFs can be consistent with simple FDFs due to
3166 close Faraday components or very small amplitudes on the secondary component, and
3167 vice versa due to noise.

3168 The main failure mode of our classifiers is misclassifying a complex source as sim-
3169 ple (Table 6.1). Whether sources with close components or small amplitudes should
3170 be considered complex is not clear, since for practical purposes they can be treated as
3171 simple: assuming the source is simple yields a very similar RM to the RM of the pri-
3172 mary component, and thus would not negatively impact further data products such as
3173 an RM grid. The scenarios where we would want a Faraday complexity classifier rather
3174 than a polarisation structure model – large-scale analysis and wide-area surveys – do
3175 not seem to be disadvantaged by considering such sources simple. Additional sources
3176 similar to these are likely hidden in presumably ‘simple’ FDFs by the frequency range
3177 and spacing of the observations, just as how these complex sources would be hidden
3178 in lower-resolution observations. Note also that misidentification of complex sources
3179 as simple is intrinsically a problem with complexity estimation even for models not
3180 well-represented by a simple FDF, as complex sources may conspire to appear as a
3181 wide range of viable models including simple (Sun et al., 2015).

3182 Conversely, high-noise simple FDFs may be consistent with complex FDFs. One
3183 key question is how Faraday complexity estimators should behave as the noise in-
3184 creases: should high noise result in a complex prediction or a simple prediction, given
3185 that a complex or simple FDF would both be consistent with a noisy FDF? Occam’s
3186 razor suggests that we should choose the simplest suitable model, and so increasing

3187 noise should lead to predictions of less complexity. This is not how our classifiers op-
 3188 erate, however: high-noise FDFs are different to the model simple FDFs and so are
 3189 predicted to be ‘not simple’. In some sense our classifiers are not looking for complex
 3190 sources, but are rather looking for ‘not simple’ sources.

3191 5.2 Limitations

3192 Our main limitations are our simplifying assumptions on FDFs and the domain gap be-
 3193 tween simulated and real observations. However, our proposed features (Section Sec-
 3194 tion 3.1) can be applied to future improved simulations.

3195 It is unclear what the effect of our simplifying assumptions are on the effective-
 3196 ness of our simulation. The three main simplifications that may negatively affect our
 3197 simulations are 1) limiting to two components, 2) assuming no external Faraday dis-
 3198 persions, and 3) assuming no internal Faraday dispersion (Faraday thickness). Future
 3199 work will explore removing these simplifying assumptions, but will need to account
 3200 for the increased difficulty in characterising the simulation with more components and
 3201 no longer having Faraday screens as components. Additionally, more work will be re-
 3202 quired to make sure that the rates of internal and external Faraday dispersion match
 3203 what might be expected from real sources, or risk making a simulation that has too
 3204 large a range of consistent models for a given source: for example, a two-component
 3205 source could also be explained as a sufficiently wide or resolved-out Faraday thick
 3206 source or a three-component source with a small third component. This greatly com-
 3207 plifies the classification task.

3208 Previous machine learning work (e.g. Brown et al., 2018) has not been run before
 3209 on real FDF data, so this [paper chapter](#) is the first example of the domain gap arising
 3210 in Faraday complexity classification. This is a problem that requires further research
 3211 to solve. We have no good way to ensure that our simulation matches reality, so some
 3212 amount of domain adaptation will always be necessary to train classifiers on simulated
 3213 data and then apply these classifiers to real data. But with the low source counts in po-
 3214 larisation science (high-resolution spectropolarimetric data currently numbers in the
 3215 few hundreds) any machine learning method will need to be trained on simulations.
 3216 This is not just a problem in Faraday complexity estimation, and domain adaptation
 3217 is also an issue faced in the wider astroinformatics community: large quantities of la-
 3218 belled data are hard to come by, and some sources are very rare (e.g. gravitational
 3219 wave detections or fast radio bursts; Agarwal et al., 2020; Gebhard et al., 2019; Zevin
 3220 et al., 2017). LR seems to handle the domain adaptation better than XGB, with only a
 3221 slightly lower accuracy on simulated data. Our results are plausible and the distribu-
 3222 tion of our simulation well overlaps the distribution of our real data (Figure 6.5).

3223 6 Conclusion

3224 We developed a simple, interpretable machine learning method for estimating Faraday
 3225 complexity. Our interpretable features were derived by comparing observed FDFs to

3226 idealised simple FDFs, which we could determine both for simulated and real obser-
 3227 vations. We demonstrated the effectiveness of our method on both simulated and real
 3228 data. Using simulated data, we found that our classifiers were 95 per cent accurate,
 3229 with near perfect recall (specificity) of Faraday simple sources. On simulated data
 3230 that matched existing observations, our classifiers obtained an accuracy of 90 per cent.
 3231 Evaluating our classifiers on real data gave the plausible results shown in Figure 6.7,
 3232 and marks the first application of machine learning to observed FDFs. Future work
 3233 will need to narrow the domain gap to improve transfer of classifiers trained on simu-
 3234 lations to real, observed data.

3235 **7 Acknowledgements**

3236 This research was conducted in Canberra, on land for which the Ngunnawal and Ngam-
 3237 bri people are the traditional and ongoing custodians. M.J.A. and J.D.L. were sup-
 3238 ported by the Australian Government Research Training Program. M.J.A. was sup-
 3239 ported by the Astronomical Society of Australia. The Australia Telescope Compact
 3240 Array is part of the Australia Telescope National Facility which is funded by the Austra-
 3241 lian Government for operation as a National Facility managed by CSIRO. We acknowl-
 3242 edge the Gomeroi people as the traditional ~~owners~~custodians of the Observatory site.
 3243 We thank the anonymous referee for their comments on this work.

3244 **K Simulating observed FDFs**

This appendix describes how we simulated FDFs in Chapter 6. We simulated FDFs by approximating them by arrays of complex numbers. An FDF F is approximated on the domain $[-\phi_{\max}, \phi_{\max}]$ by a vector $\vec{F} \in \mathbb{R}^d$:

$$\vec{F}_j = \sum_{k=0}^1 A_k \delta(-\phi_{\max} + j\delta\phi - \phi_k) \quad (6.12)$$

where $\delta\phi = (\phi_{\max} - \phi_{\min})/d$ and d is the number of Faraday depth samples in the FDF. \vec{F} is sampled by uniformly sampling its parameters:

$$\phi_k \in [\phi_{\min}, \phi_{\min} + \delta\phi, \dots, \phi_{\max}] \quad (6.13)$$

$$A_k \sim \mathcal{U}(0, 1). \quad (6.14)$$

We then generate a vector polarisation spectrum $\vec{P} \in \mathbb{R}^m$ from \vec{F} using a Equation 6.15:

$$\vec{P}_\ell = \sum_{j=0}^j F_j e^{2i(\phi_{\min} + j\delta\phi)\lambda_\ell^2} d\phi. \quad (6.15)$$

λ_ℓ^2 is the discretised value of λ^2 at the ℓ th index of \vec{P} . This requires a set of λ^2 values, which depends on the dataset being simulated. These values can be treated as the channel wavelengths at which the polarisation spectrum was observed. We then add Gaussian noise with variance σ^2 to each element of \vec{P} to obtain a discretised noisy observation $\hat{\vec{P}}$. Finally, we perform RM synthesis using the Canadian Initiative for Radio Astronomy Data Analysis RM package², which is a Python module that implements a discrete version of RM synthesis:

$$\hat{F}_j = m^{-1} \sum_{\ell=1}^m \vec{P}_\ell e^{-2i(\phi_{\min} + j\delta_\phi)\lambda_\ell^2}. \quad (6.16)$$

3245 L 2-Wasserstein begets Faraday moments

Minimising the 2-Wasserstein distance between a model FDF and the simple manifold gives the second Faraday moment of that FDF. This appendix demonstrates that fact. Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the sum-normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1} \quad (6.17)$$

$$\tilde{S}(\phi; \phi_w) = \delta(\phi - \phi_w). \quad (6.18)$$

The W_2 distance, usually defined on probability distributions, can be extended to one-dimensional complex functions A and B by normalising them:

$$D_{W_2}(A \parallel B)^2 = \inf_{\gamma \in \Gamma(A, B)} \iint_{\phi_{\min}}^{\phi_{\max}} |x - y|^2 d\gamma(x, y) \quad (6.19)$$

$$\tilde{A}(\phi) = \frac{|A(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |A(\theta)| d\theta} \quad (6.20)$$

$$\tilde{B}(\phi) = \frac{|B(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |B(\theta)| d\theta} \quad (6.21)$$

3246 where $\Gamma(A, B)$ is the set of couplings of A and B , i.e. the set of joint probability distributions
 3247 that marginalise to A and B ; and $\inf_{\gamma \in \Gamma(A, B)}$ is the infimum over $\Gamma(A, B)$. This can be
 3248 interpreted as the minimum cost to ‘move’ one probability distribution to the other,
 3249 where the cost of moving one unit of probability mass is the squared distance it is
 3250 moved.

²<https://github.com/CIRADA-Tools/RM>

The set of couplings $\Gamma(\tilde{F}, \tilde{S})$ is the set of all joint probability distributions γ such that

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\phi = \tilde{S}(\varphi; \phi_w), \quad (6.22)$$

$$\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\varphi = \tilde{F}(\phi). \quad (6.23)$$

The coupling that minimises the integral in Equation 6.19 will be the optimal transport plan between \tilde{F} and \tilde{S} . Since \tilde{F} and \tilde{S} are defined in terms of delta functions, the optimal transport problem reduces to a discrete optimal transport problem and the optimal transport plan is:

$$\gamma(\phi, \varphi) = \frac{A_0 \delta(\phi - \phi_0) + A_1 \delta(\phi - \phi_1)}{A_0 + A_1} \delta(\varphi - \phi_w). \quad (6.24)$$

In other words, to move the probability mass of \tilde{S} to \tilde{F} , a fraction $A_0/(A_0 + A_1)$ is moved from ϕ_w to ϕ_0 and the complementary fraction $A_1/(A_0 + A_1)$ is moved from ϕ_w to ϕ_1 . Then:

$$D_{W_2}(\tilde{F} \parallel \tilde{S})^2 = \iint_{\phi_{\min}}^{\phi_{\max}} |\phi - \varphi|^2 d\gamma(\phi, \varphi) \quad (6.25)$$

$$= \frac{A_0(\phi_0 - \phi_w)^2 + A_1(\phi_1 - \phi_w)^2}{A_0 + A_1}. \quad (6.26)$$

To obtain the W_2 distance to the simple manifold, we need to minimise this over ϕ_w . Differentiate with respect to ϕ_w and set equal to zero to find

$$\phi_w = \frac{A_0 \phi_0 + A_1 \phi_1}{A_0 + A_1}. \quad (6.27)$$

Substituting this back in, we find

$$\zeta_{W_2}(F)^2 = \frac{A_0 A_1}{A_0 + A_1} (\phi_0 - \phi_1)^2 \quad (6.28)$$

³²⁵¹ which is the Faraday moment.

3252 M Euclidean distance in the no-RMSF case

In this appendix we calculate the minimised Euclidean distance evaluated on a model FDF (Equation 6.1). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the normalised

simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1} \quad (6.29)$$

$$\tilde{S}(\phi; \phi_e) = \delta(\phi - \phi_e). \quad (6.30)$$

The Euclidean distance between \tilde{F} and \tilde{S} is then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (6.31)$$

$$= \int_{\phi_{\min}}^{\phi_{\max}} |\tilde{F}(\phi) - \delta(\phi - \phi_e)|^2 d\phi. \quad (6.32)$$

Assume $\phi_0 \neq \phi_1$ (otherwise, D_E will always be either 0 or 2). If $\phi_e = \phi_0$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \quad (6.33)$$

$$= \frac{1}{(A_0 + A_1)^2} \int_{\phi_{\min}}^{\phi_{\max}} A_1^2 |\delta(\phi - \phi_1) - \delta(\phi - \phi_0)|^2 d\phi \quad (6.34)$$

$$= \frac{2A_1^2}{(A_0 + A_1)^2} \quad (6.35)$$

and similarly for $\phi_e = \phi_1$. If $\phi_e \neq \phi_0$ and $\phi_e \neq \phi_1$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 = \frac{A_0^2 + A_1^2 + 1}{(A_0 + A_1)^2}. \quad (6.36)$$

The minimised Euclidean distance when $\phi_0 \neq \phi_1$ is therefore

$$D_E(F) = \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \quad (6.37)$$

$$= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \quad (6.38)$$

3253 If $\phi_0 = \phi_1$, then the minimised Euclidean distance is 0.

3254 N Hyperparameters for LR and XGB

3255 This section contains tables of the hyperparameters that we used for our classifiers
 3256 in Chapter 6. Table 6.2 and Table 6.3 tabulate the hyperparameters for XGB and LR
 3257 respectively for the 'ATCA' dataset. Table 6.4 and Table 6.5 tabulate the hyperparameters
 3258 for XGB and LR respectively for the 'ASKAP' dataset.

Table 6.2: XGB hyperparameters for the ‘ATCA’ dataset.

Parameter	Value
<code>colsample_bytree</code>	0.912
<code>gamma</code>	0.532
<code>learning_rate</code>	0.1
<code>max_depth</code>	7
<code>min_child_weight</code>	2
<code>scale_pos_weight</code>	1
<code>subsample</code>	0.557
<code>n_estimators</code>	135
<code>reg_alpha</code>	0.968
<code>reg_lambda</code>	1.420

Table 6.3: LR hyperparameters for the ‘ATCA’ dataset.

Parameter	Value
<code>penalty</code>	L1
<code>C</code>	1.668

Table 6.4: XGB hyperparameters for the ‘ASKAP’ dataset.

Parameter	Value
<code>colsample_bytree</code>	0.865
<code>gamma</code>	0.256
<code>learning_rate</code>	0.1
<code>max_depth</code>	6
<code>min_child_weight</code>	1
<code>scale_pos_weight</code>	1
<code>subsample</code>	0.819
<code>n_estimators</code>	108
<code>reg_alpha</code>	0.049
<code>reg_lambda</code>	0.454

Table 6.5: LR hyperparameters for the ‘ASKAP’ dataset.

Parameter	Value
<code>penalty</code>	L2
<code>C</code>	0.464

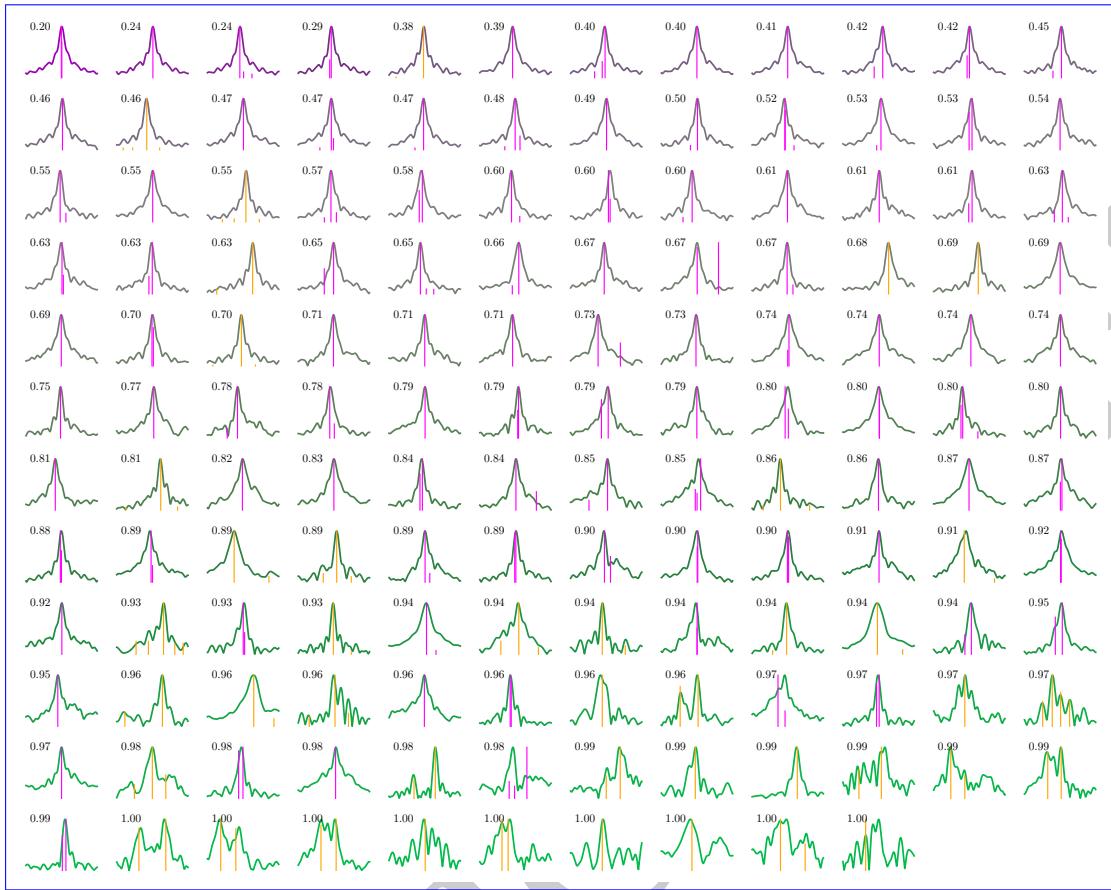


Figure 6.7: The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the LR estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.

3259 O Predictions on real data

3260 This appendix contains Figure 6.7 and Figure 6.8. These show the predicted probability
 3261 of being Faraday complex for all real data used in Chapter 6, drawn from Livingston et al. (2021)
 3262 and O'Sullivan et al. (2017).

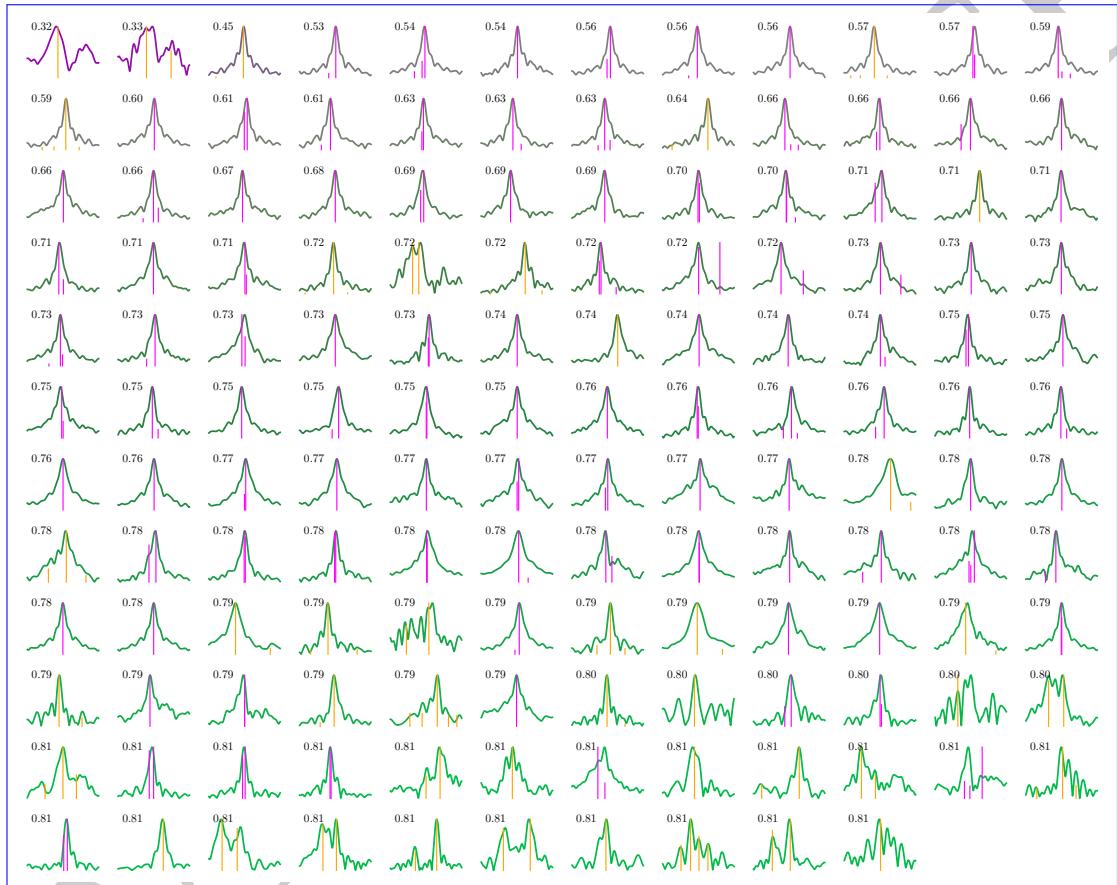


Figure 6.8: The 142 observed FDFs ordered by XGB-estimated probability of being Faraday complex. Livingston-identified components are shown in orange while O'Sullivan-identified components are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple while more complex FDFs are shown in green, and the numbers overlaid indicate the XGB estimate. A lower number indicates a lower probability that the corresponding source is complex, i.e. lower numbers correspond to simpler spectra.

REVISION DIFF

3264

Conclusion

3265 The future of radio astronomy is immensely exciting, with upcoming radio surveys
3266 like EMU and POSSUM sure to revolutionise our understanding of the violent Uni-
3267 verse occupied by radio active galactic nuclei. Without innovative new methods for
3268 processing astronomical data at scale, however, we will be limited in what interesting
3269 physics we can learn about with these surveys. In this thesis we have explored meth-
3270 ods for identifying extragalactic radio sources in these future wide-area radio surveys.
3271 We developed a new automated cross-identification approach and a new way of clas-
3272 sifying radio sources as Faraday complex or Faraday simple. We applied our methods
3273 to real data, and used our cross-identification method to directly probe the mechani-
3274 cal energy contribution of active galactic nuclei (AGN) to the intergalactic medium
3275 (IGM). Our contributions here are all aimed at extracting more information out of the
3276 very large radio surveys we have begun to face.

3277 Chapter 4 developed a new approach to cross-identifying radio components with
3278 their infrared host galaxies and applied this method to the CDFS and ELAIS-S1 AT-
3279 LAS survey, a pilot survey for the upcoming ASKAP key survey EMU. This was the
3280 first application of machine learning to radio cross-identification. In the process, we
3281 demonstrated that our methods worked on ELAIS-S1 with similar performance to how
3282 they worked on CDFS, despite these being different parts of the sky. We also eval-
3283 uated simple positional matching cross-identification on the ATLAS fields, showing
3284 better performance than either machine learning or Radio Galaxy Zoo volunteers on
3285 this dataset, which we suggest is because of how most objects in ATLAS are very com-
3286 pact. Our machine learning methods performed similarly whether they were trained
3287 on expert-provided or citizen science-provided labelled datasets, clearly demonstra-
3288 ting the benefit of citizen science labels to radio astroinformatics even if they may be
3289 less accurate than labels provided by experts. We showed that the algorithm we had
3290 developed could in principle perform much better than positional matching if only its
3291 binary accuracy could be raised, perhaps with a more complex or better-trained classi-
3292 fier. Finally, we concluded that ATLAS was an insufficiently complex dataset to train
3293 machine learning algorithms for future radio surveys. ATLAS contains many compact
3294 sources, and while there will be many such sources in EMU and other future surveys
3295 there will also be a zoo of partially resolved, extended, strange, or unusual morpholo-
3296 gies. Methods designed or trained on a dataset like ATLAS, which lacks diversity of
3297 non-compact sources, will not work on the wide range of extended radio sources that

3298 will appear. More complex and larger training sets, perhaps real or perhaps simulated,
 3299 will be necessary for producing algorithms intended for use in future surveys.

3300 To further investigate our classification-based cross-identification algorithm we
 3301 needed a more complex dataset, and in Chapter 5 we turned to FIRST, a legacy sur-
 3302 vey containing around 900 000 radio components. Of these, 250 000 or so were not
 3303 compact, compared to the vast majority of ATLAS. We increased the complexity of
 3304 our machine learning model and trained it on this dataset along with 75 000 Radio
 3305 Galaxy Zoo labels, far more than we had available for ATLAS. The resulting classifier
 3306 could then be used to cross-identify every extended object in FIRST. Using the fact
 3307 that any two radio components in the same radio source should also have the same
 3308 host galaxy, we identified not only the host galaxies of our radio components but also
 3309 their association to other radio components. This allowed us to produce the RGZ-Ex
 3310 catalogue, containing over 150 000 extended radio sources—the largest existing cata-
 3311 logue of extended radio sources at the time of writing. In this catalogue we identified
 3312 40 giant radio galaxies, most of which were new to literature.

3313 Such a large catalogue enabled us to estimate a radio luminosity function (RLF)
 3314 for extended radio sources, the first time a RLF has been produced for just extended
 3315 sources. We could also subdivide this RLF into a fractional RLF, and we were able
 3316 to produce RLFs divided by the mid-infrared colour of the host galaxy, their physi-
 3317 cal extent, and their redshift. Extended radio sources ought to be the sources which
 3318 contribute the most mechanical energy to their surrounding IGM, and with an RLF
 3319 dedicated solely to extended sources we were able to estimate this energy contribu-
 3320 tion as 1.3×10^{30} and $1.2 \times 10^{32} \text{ W Mpc}^{-3}$. Perhaps most importantly, we showed our
 3321 cross-identification approach works and used it to obtain a physically meaningful re-
 3322 sult. Our methods can easily be extended to new surveys, as long as sufficient host
 3323 galaxy/radio component pairs are known so that the algorithm can be trained.

3324 Our work on cross-identification allowed us to extract more information from wide-
 3325 area continuum radio surveys. To help gain more use out of large polarisation surveys
 3326 as well, we developed an interpretable Faraday complexity classifier for Faraday dis-
 3327 persions functions (FDFs) in Chapter 6. We constructed features that were easy to un-
 3328 derstand by measuring the distance of observations from a simple model of Faraday
 3329 simple sources. Our resulting features could be calculated from both simulated and
 3330 real observations. We demonstrated the effectiveness of our method on both simulated
 3331 and real data and showed that on simulated data our simple methods matched the
 3332 state-of-the-art convolutional neural network (CNN) classifier despite having far less
 3333 parameters. This was the first application of machine learning to real spectropolarimet-
 3334 ric data. We highlighted the domain gap between simulated and observed FDFs and
 3335 suggested reframing the Faraday complexity classification task as finding simple and
 3336 non-simple sources. This method will be applicable to future surveys like POSSUM.

3337 1 Future Work

3338 There are many ways that the research in this thesis can be extended in future. We will
3339 summarise some of these here.

3340 Our methods can be extended in a number of ways. These fall into two categories:
3341 further applications and extensions to the algorithms. An obvious target for future
3342 work is to apply our binary cross-identification algorithm (BXID; Section 4.3.1 and Sec-
3343 tion 3) to new and upcoming surveys like LotSS and EMU. These promise a tremen-
3344 dous amount of data with new discoveries certain to be lying in wait within, and cross-
3345 identifying the radio emission with its corresponding infrared or optical host galaxy
3346 will be vital for uncovering those secrets. Similarly, we would like to apply our Faraday
3347 complexity classifier to future spectropolarimetric surveys like POSSUM. Our method-
3348 ology can also be improved. Better models almost certainly exist than the CNN we
3349 applied to BXID. As we demonstrated in Chapter 6, a well thought-out model and fea-
3350 tures may best a complex model like the CNN. How would our cross-identification
3351 approach worked if, say, we applied it to hand-selected features such as those chosen
3352 by Proctor (2006)? Would a search over more CNN architectures, like that described
3353 by Lukic et al. (2019), result in better classifiers and hence better cross-identifications?
3354 Perhaps we could even improve performance by pre-training on some unlabelled but
3355 larger dataset? A less obvious improvement to our BXID approach would be to change
3356 how the classification scores are aggregated. Currently this is a weighted maximisa-
3357 tion over candidate host galaxies, but other methods are possible. The weights could
3358 be something other than a Gaussian function of distance, from other functions of dis-
3359 tance to an entire separate classification model. Maybe we could aggregate the scores
3360 in bulk, using some kind of algorithm that assigns radio-host relationships based on
3361 not just the radio source itself, but also on the other radio sources around it and how
3362 they have been paired to their own host galaxies.

3363 The way that our labels were generated for BXID could be improved. Our algo-
3364 rithms in Chapter 4 and Chapter 5 were trained on labels generated by Radio Galaxy
3365 Zoo. These labels were aggregated from multiple different labellers (usually 20) by
3366 majority vote, with the most common label for any given radio object being assigned
3367 as the true label in Radio Galaxy Zoo. This is not the only possible aggregation strategy,
3368 though. We employed the Dawid-Skene method (Dawid & Skene, 1979) ourselves in
3369 Section 3.1 to help assess the performance of our cross-identification algorithm, and
3370 this model for example may also be applied to Radio Galaxy Zoo. There are in fact ag-
3371 gregation strategies that work in tandem with a machine learning model to get better
3372 labels, such as Raykar et al. (2010). These methods simultaneously take into account
3373 the labels and the labellers, and can accommodate for different levels of ability in the
3374 labellers, or different levels of difficulty in the examples being labelled.

3375 Our RLFs could be improved. The RLF calculations in Chapter 5 are severely lim-
3376 ited by the availability of redshifts. We limit our analyses to host galaxies that do have
3377 available spectroscopic redshifts in SDSS, but we could also employ the less-reliable
3378 but considerably more prolific photometric redshifts. These are derived from regres-
3379 sion models rather than direct observations of redshifted spectral lines, and so can

be produced from photometric surveys without dedicated spectroscopy. However, without methods to handle the uncertainty introduced by photometric redshifts, the resulting RLFs would be unreliable. Developing ways to not only decrease the uncertainty in photometric redshifts but also to understand and incorporate the uncertainty into downstream calculations like those of the RLFs, for example using probabilistic programming (e.g. Bingham et al., 2019), will allow these photometric redshifts to be used and tremendously increase sample sizes. Besides existing photometric redshifts, future surveys will also produce many more redshifts, both spectroscopic and photometric. This will be very important for surveys like EMU, which are both deep and wide with low redshift availability.

The RGZ-Ex dataset (Chapter 5) also lets us pose many other interesting science questions. We demonstrated in Appendix J that rare galaxy classes can be identified from within this dataset, including examples that have never before been identified in the literature. Our dataset may be augmented with other features and used to identify unusual objects in a similar way. Besides this, our fractional RLFs could also be extended with any number of galaxy properties. One particularly interesting property could be morphology, as other algorithms in radio astroinformatics are developed which can automatically identify morphologies (e.g. Wu et al., 2019): such a classifier could be used to segment RGZ-Ex and a fractional morphological RLF could be obtained. These morphologies may even be classes that are not easily separated, such as those found by self-organising maps (e.g. Polsterer et al., 2015). Of course, there are other properties that are more easily extracted, such as optical lines and colours which could be taken from SDSS using our SDSS cross-identifications.

While creating features for FDFs in Chapter 6 we demonstrated that W2 distance was a sensible distance measure between FDFs. This is useful for more than just feature construction, as it implies a geometry on the space of FDFs. This distance could be used to help gain insight on the behaviours of future algorithms that work on FDFs. A particularly exciting idea is to improve QU fitting by modifying the distance function to match ours. Our features could also be used to develop other methods for FDF analysis, like outlier detection or data visualisation.

Further research is needed to close the domain gap for FDFs. This is an interesting case study as it is such a simple case, where we know essentially all the physics behind the observations and the observations are one-dimensional. Even this is not enough, and whether through unmodelled physics (e.g. more than two screens, depolarisation) or unmodelled observational properties (e.g. radio frequency interference) simulation and observation do not fully line up. This is critical if we want to train machine learning algorithms on simulations in the future, and we very much want to do this to augment our limited observational training data. Similarly, the domain gap must be reduced for radio continuum observations. Our results in Chapter 4 show that pilot surveys like ATLAS may not contain enough complex sources to train machine learning models, and while larger surveys like FIRST exist, transferring models from a survey undertaken with one set of observing parameters (telescope, frequency, depth, resolution...) is both non-trivial and as-yet relatively unexplored.

3423 2 Implications for radio citizen science

3424 Our results in Chapter 4 demonstrated that machine learning methods trained on citizen
3425 science labels perform comparably to those trained on expert labels, even when
3426 those labels are lower quality than expert labels. We applied this insight in Chap-
3427 ter 5 to obtain scientific results, using machine learning to extrapolate the labels to a
3428 larger dataset. While Radio Galaxy Zoo alone was not enough to fully label FIRST, Ra-
3429 dio Galaxy Zoo working in tandem with machine learning was. This is a pattern that
3430 may hold true for future surveys and applications, too. Future astronomical research
3431 at-scale may leverage the idea of people working with machine learning, sometimes
3432 called *human-in-the-loop* (e.g. Holzinger, 2016) learning, to pore through data-at-scale.

3433 3 Implications for wide-area radio surveys

3434 As we move toward larger and larger datasets, an important question is how applica-
3435 ble our models will be across the sky. Our results in Chapter 4 showed that we can
3436 expect some generalisation, as our model trained on one patch of sky was applicable
3437 to another without a great loss of performance. Similarly, our classifier trained on part
3438 of FIRST seemed to work well on the rest of the dataset. This is promising as it implies
3439 that limited area surveys may help develop training sets that generalise to the whole
3440 sky, potentially making the process of generating training sets considerably cheaper.

3441 With our work in Chapter 4 and 5 we demonstrated that a large set of good quality,
3442 complex data is required for training good astroinformatics algorithms. Pilot datasets
3443 like ATLAS will not work by themselves: the sources they contain are too simple and
3444 their complex sources are too few. A sensible question to ask is, could we simulate
3445 data for training purposes? We trained our classifier in Chapter 6 on simulated data
3446 and found that it was difficult to bridge the domain gap between simulation and ob-
3447 servation, even in a well-understood, one-dimensional case—let alone the complex
3448 three-dimensional projected morphologies we observe in imagery. Getting across this
3449 domain gap will be difficult and will necessarily be a major topic of research in the
3450 astroinformatics field in the near future.

3451 Tying observations, simulations, and models together are the representation of the
3452 data: features. Our results in Chapter 6 show that judicious choice of features can out-
3453 perform even complex and powerful models. This is good for two reasons. The first
3454 reason is that these features may be more easily interpreted. The meaning of the fea-
3455 tures may be understood to be representative of some physical property, or at least the
3456 relationship between physical reality and predictions may be more easily found. The
3457 second reason is that features may be selected which can generalise well to datasets
3458 beyond just the training set. In other words, features that are less overfit to the train-
3459 ing set. This is of particular concern when developing predictive models on simulated
3460 training data, as features being less suited for real data than for simulations is one as-
3461 pect that may contribute to the domain gap. Choosing good features in astronomy may
3462 be more important than in many other fields for which machine learning is applied, as
3463 while in most fields it is possible to conduct experiments, in astronomy we only have

3464 one Universe to look at. We need to make the best use we can of the limited radio sky.

3465 4 Final remarks

3466 Radio astronomy faces a deluge of data, with current and upcoming surveys delivering
 3467 incredible amounts of data for science use. While we can get a lot out of these data—
 3468 from investigations into new physics, to finding rare and unusual objects—doing so
 3469 is dependent on the development of new methods for astronomy at scale. These fu-
 3470 ture methods will necessarily be computational, and so the challenge lies in encoding
 3471 the abstract concepts of astronomy and astrophysics into a rigorously defined set of
 3472 rules that a computer can interpret and execute at scale. This is decidedly non-trivial,
 3473 and the nuance and unique skills required to do so motivates the burgeoning field of
 3474 astroinformatics. By combining concepts from deep within the often disparate fields
 3475 of astronomy and machine learning, we hope that we will be able to make the absolute
 3476 most of the incredible new technologies and instruments that will arise in the future
 3477 of radio astronomy.

3478 5 Classification models

3479 This appendix describes the three different models we used for binary classification in
 3480 (logistic regression, convolutional neural networks, and random forests) and was part
 3481 of Alger et al. (2018).

3482 4.1 Logistic Regression

3483 Logistic regression is linear in the feature space and outputs the probability that the
 3484 input has a positive label. The model is (Bishop, 2006):

$$f(\vec{x}) = \sigma(\vec{w}^T \vec{x} + b) ,$$

where $\vec{w} \in \mathbb{R}^D$ is a vector of parameters, $b \in \mathbb{R}$ is a bias term, $\vec{x} \in \mathbb{R}^D$ is the feature
 vector representation of a candidate host, and $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ is the logistic sigmoid function:

$$\sigma(a) = (1 + \exp(-a))^{-1} .$$

3485 The logistic regression model is fully differentiable, and the parameters \vec{w} can therefore
 3486 be learned using gradient-based optimisation methods. We used the scikit-learn
 3487 (Pedregosa et al., 2011) implementation of logistic regression with balanced classes.

3488 4.1 Convolutional neural networks

3489 Convolutional neural networks (CNN) are a biologically-inspired prediction model
 3490 for prediction with image inputs. The input image is convolved with a number of
 3491 filters to produce output images called feature maps. These feature maps can then

3492 be convolved again with other filters on subsequent layers, producing a network of
3493 convolutions. The whole network is differentiable with respect to the values of the
3494 filters and the filters can be learned using gradient-based optimisation methods. The
3495 final layer of the network is logistic regression, with the convolved outputs as input
3496 features. For more detail, see subsection II.A, LeCun et al. (1998). We used KERAS
3497 (Chollet et al., 2015) to implement our CNN, accounting for class imbalance by reweighting
3498 the classes.

3499 Architecture of our CNN. Parenthesised numbers indicate the size of output layers
3500 as a tuple (width, height, depth). The concatenate layer flattens the output of the previous
3501 layer and adds the 10 features derived from the candidate host in SWIRE, i.e. the flux
3502 ratios, stellarity indices, and distance. The dropout layer randomly sets 25% of its
3503 inputs to zero during training to prevent overfitting. Diagram based on .

3504 CNNs have recently produced good results on large image-based datasets in astronomy
3505 (Dieleman et al., 2015; Lukic et al., 2018, e.g.). We employed only a simple CNN model
3506 in as a proof of concept that CNNs may be used for class probability prediction on radio
3507 images. The model architecture we used is shown in .

3508 4.1 Random Forests

3509 Random forests are an ensemble of decision trees (Breiman, 2001). They consider
3510 multiple subsamples of the training set, where each subsample is sampled with replacement
3511 from the training set. For each subsample a decision tree classifier is constructed
3512 by repeatedly making axis-parallel splits based on individual features. In a random
3513 forest the split decision is taken based on a random subset of features. To classify
3514 a new data point, the random forest takes the weighted average of all classifications
3515 produced by each decision tree. In we used the scikit-learn (Pedregosa et al., 2011)
3516 implementation of random forests with 10 trees, the information entropy split criterion,
3517 a minimum leaf size of 45, and balanced classes.

3518 5 Accuracy tables

3519 This section contains tables of accuracy for our cross-identification method applied to
3520 CDFS and ELAIS-S1, and was originally presented in Alger et al. (2018). In and we
3521 list the balanced accuracies of our classifiers on the cross-identification task for CDFS
3522 and ELAIS-S1 respectively, averaged over each set of training quadrants. In and we
3523 list the balanced accuracies of classifiers on the cross-identification task for CDFS and
3524 ELAIS-S1 respectively, averaged over each set of training quadrants.

3525 Balanced accuracies for different binary classification models trained and tested on
3526 SWIRE objects in CDFS. The ‘Labeller’ column states what set of training labels were
3527 used to train the classifier, and the ‘Classifier’ column states what classification model
3528 was used. ‘CNN’ is a convolutional neural network, ‘LR’ is logistic regression and ‘RF’
3529 is random forests. Accuracies are evaluated against the expert label set derived from
3530 Norris et al. (2006). The standard deviation of balanced accuracies evaluated across
3531 the four quadrants of CDFS () is also shown. The ‘compact’ set refers to SWIRE objects

3532 within 1' of a compact radio component, the 'resolved' set refers to SWIRE objects
 3533 within 1' of a resolved radio component, and 'all' is the union of these sets. Labeller
 3534 Classifier Mean 'Compact' Mean 'Resolved' Mean 'All' accuracy accuracy accuracy (per
 3535 cent) (per cent) (per cent) Norris LR 91.5 ± 1.0 93.2 ± 1.0 93.0 ± 1.2 CNN 92.6 ± 0.7
 3536 91.2 ± 0.5 92.0 ± 0.6 RF 96.7 ± 1.5 91.0 ± 4.5 96.0 ± 2.5 RGZ LR 89.5 ± 0.8 90.5 ± 1.7 90.2 ± 0.8 CNN
 3537 89.4 ± 0.6 89.6 ± 1.3 89.4 ± 0.5 RF 94.5 ± 0.2 95.8 ± 0.4 94.7 ± 0.3

3538 Balanced accuracies for different binary classification models trained on SWIRE
 3539 objects in CDFS and tested on SWIRE objects in ELAIS-S1. Columns and abbreviations
 3540 are as in . Accuracies are evaluated against the expert label set derived from Middelberg et al. (2008)
 3541 . The standard deviations of balanced accuracies of models trained on the four subsets
 3542 of CDFS () are also shown. Labeller Classifier Mean 'Compact' Mean 'Resolved' Mean
 3543 'All' accuracy accuracy accuracy (per cent) (per cent) (per cent) Norris LR 94.6 ± 0.4
 3544 93.3 ± 2.0 95.3 ± 0.1 CNN 94.8 ± 0.2 92.8 ± 0.5 94.4 ± 0.2 RF 85.9 ± 3.8 70.0 ± 2.8 86.6 ± 3.2 RGZ
 3545 LR 91.8 ± 0.3 91.9 ± 0.5 92.0 ± 0.2 CNN 90.1 ± 0.3 91.1 ± 0.9 90.2 ± 0.3 RF 95.1 ± 0.1 95.2 ± 0.0
 3546 95.2 ± 0.3

3547 Cross-identification accuracies for different classification models on CDFS. The 'Labeller'
 3548 column states what set of training labels were used to train the method, and the 'Classifier'
 3549 column states what classification model was used. 'CNN' is a convolutional neural
 3550 network, 'LR' is logistic regression, 'RF' is random forests, and 'Labels' is the accuracy
 3551 of the label set itself. 'Perfect' indicates that the true labels of the test set were used and
 3552 hence represents an upper bound on cross-identification accuracy with our method.
 3553 'NN' is a nearest neighbours approach. Accuracies are evaluated against the expert
 3554 label set, so 'Norris' labels are 100 per cent accurate by definition. The standard deviation
 3555 of accuracies evaluated across the four quadrants of CDFS () is also shown. Labeller
 3556 Classifier Mean 'Compact' Mean 'Resolved' Mean 'All' accuracy accuracy accuracy (per
 3557 cent) (per cent) (per cent) NN 97.2 ± 1.7 75.7 ± 7.9 93.4 ± 0.8 Random 97.9 ± 2.2
 3558 22.3 ± 9.2 83.2 ± 4.7 Norris Labels 100.0 ± 0.0 100.0 ± 0.0 100.0 ± 0.0 Perfect 97.9 ± 2.2
 3559 99.0 ± 1.8 98.1 ± 1.7 LR 97.3 ± 0.5 76.0 ± 3.2 93.7 ± 1.8 CNN 96.6 ± 0.9 74.3 ± 12.3 93.5 ± 0.5 RF
 3560 96.1 ± 1.4 75.8 ± 6.7 93.8 ± 2.0 RGZ Labels 53.1 ± 8.5 56.7 ± 5.9 54.4 ± 5.9 LR 97.3 ± 1.9
 3561 74.5 ± 5.1 93.6 ± 1.7 CNN 85.4 ± 2.6 68.1 ± 9.2 92.4 ± 1.1 RF 97.5 ± 0.9 74.3 ± 7.9 93.7 ± 1.5

3562 Cross-identification accuracies for different classification models on ELAIS-S1. Columns
 3563 and abbreviations are as in . Accuracies are evaluated against the expert label set
 3564 derived from Middelberg et al. (2008) cross-identifications. The standard deviation
 3565 of accuracies evaluated across models trained on the four quadrants of CDFS () is
 3566 also shown. Labeller Classifier Mean 'Compact' Mean 'Resolved' Mean 'All' accuracy
 3567 accuracy accuracy (per cent) (per cent) (per cent) NN 95.5 ± 0.0 92.8 ± 0.0 95.5 ± 0.0 Random
 3568 61.9 ± 1.1 26.6 ± 2.1 61.9 ± 1.1 Middelberg Perfect 99.6 ± 0.0 99.8 ± 0.0 99.6 ± 0.0 Norris
 3569 LR 89.0 ± 1.1 89.7 ± 1.8 94.4 ± 0.9 CNN 89.7 ± 0.3 89.4 ± 1.4 94.3 ± 0.7 RF 83.8 ± 5.6 82.3 ± 4.1
 3570 90.6 ± 2.1 RGZ LR 90.5 ± 1.0 92.7 ± 0.2 95.9 ± 0.1 CNN 84.6 ± 0.6 84.6 ± 0.6 91.8 ± 0.3 RF
 3571 91.3 ± 1.0 90.3 ± 2.4 94.7 ± 1.2

3572 **5 SWIRE object scores**

3573 This appendix is from Alger et al. (2018), and contains scores predicted by our binary
 3574 classifiers for each SWIRE object within 1' of a radio component in CDFS and ELAIS-S1.
 3575 Scores for SWIRE CDFS objects are shown in and scores for SWIRE ELAIS-S1 are
 3576 shown in . For CDFS, the score for an object in a quadrant is predicted by binary
 3577 classifiers trained on all other quadrants. For ELAIS-S1, we show the scores predicted
 3578 by binary classifiers trained on each CDFS quadrant. Note that these scores have *not*
 3579 been weighted by Gaussians. These are partial tables, and the full tables are available
 3580 online at the *Monthly Notices of the Royal Astronomical Society* website¹.

3581 The columns of the score tables are defined as follows:

- 3582 • *SWIRE*—SWIRE designation for candidate host galaxy.
- 3583 • *RA*—Right ascension (J2000).
- 3584 • *Dec*—Declination (J2000).
- 3585 • *Expert host*—Whether the candidate host galaxy is a host galaxy according to
 3586 Norris et al. (2006) or Middelberg et al. (2008) cross-identifications of CDFS and
 3587 ELAIS-S1 respectively.
- 3588 • *RGZ host*—Whether the candidate host galaxy is a host galaxy according to Radio
 3589 Galaxy Zoo cross-identifications (Wong et al., in prep.). This is always ‘no’ for
 3590 ELAIS-S1 objects.
- 3591 • *C/L/D*—Score assigned by binary classifier *C* trained on label set *L* of *D* candidate
 3592 host galaxies. *C* may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or
 3593 random forests respectively. *L* may be ‘Norris’ or ‘RGZ’ for expert and Radio
 3594 Galaxy Zoo labels respectively. *D* may be ‘All’, ‘Compact’ or ‘Resolved’ for each
 3595 respective subset defined in .

3596 Scores output by our trained classifiers for SWIRE CDFS candidate host galaxies.
 3597 Columns are defined in . Full table electronic. SWIRE RA Dec Expert RGZ host host All
 3598 Compact Resolved All Compact Resolved J032603.15-284708.5 51.5132-28.7857 yes no
 3599 0.5838-0.4697 0.4848 0.3754 0.3881 0.3404 J032603.39-284010.1 51.5142-28.6695 no no
 3600 0.0373-0.5814 0.4878 0.7896 0.7616 0.4668 J032603.44-284210.1 51.5144-28.7028 no no
 3601 0.0232-0.4891 0.5101 0.4319 0.4298 0.3474 J032603.44-284222.2 51.5143-28.7062 no no
 3602 0.0006-0.4164 0.5216 0.0400 0.0444 0.0276 J032603.45-284748.4 51.5144-28.7968 no no
 3603 0.0014-0.4914 0.4865 0.1904 0.1895 0.1467 J032603.50-284637.0 51.5146-28.7770 no no
 3604 0.0074-0.4144 0.5382 0.1418 0.1515 0.1166 J032603.60-284627.4 51.5150-28.7743 no no
 3605 0.0012-0.4578 0.5165 0.0850 0.0904 0.0484 J032603.63-283840.5 51.5151-28.6446 no no
 3606 0.0021-0.4153 0.5577 0.1678 0.1746 0.1323 J032603.66-283822.8 51.5153-28.6397 no no
 3607 0.0001-0.4752 0.5009 0.0864 0.0861 0.0613 J032603.75-284014.1 51.5156-28.6706 no no
 3608 0.0547-0.3408 0.5388 0.4889 0.5242 0.7301 All Compact Resolved All Compact Resolved
 3609 All Compact Resolved All Compact Resolved 0.2489-0.0009 0.1557 0.2939 0.0007 0.1174
 3610 0.8922-0.8018 0.8732 0.7167 0.6599 0.7801 0.0183 0.1646 0.1480 0.7637 0.7065 0.6070 0.0000

3611 0.0000 0.0000 0.1629 0.0519 0.1275 0.0155 0.0164 0.0815 0.3714 0.5626 0.2488 0.0000 0.0734
 3612 0.0000 0.1315 0.2116 0.4150 0.0005 0.0006 0.0175 0.0460 0.0810 0.0299 0.2656 0.1418 0.0000
 3613 0.7631 0.8166 0.5378 0.0013 0.0037 0.0160 0.1792 0.0663 0.1821 0.0000 0.0000 0.0000 0.0255
 3614 0.0000 0.0000 0.0047 0.0010 0.0337 0.1284 0.2198 0.0694 0.0720 0.0000 0.0000 0.6240 0.6681
 3615 0.6704 0.0008 0.0006 0.0374 0.1053 0.1424 0.0807 0.1231 0.0876 0.0000 0.8517 0.7532 0.7019
 3616 0.0021 0.0073 0.0386 0.1482 0.0403 0.1210 0.0000 0.0532 0.0000 0.0000 0.0302 0.0000 0.0001
 3617 0.0004 0.0038 0.0854 0.0447 0.0514 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0542 0.2712
 3618 0.2318 0.5026 0.5631 0.5032 0.0595 0.0545 0.0000 0.4289 0.0789 0.1420
 3619 Scores output by our trained classifiers for SWIRE-ELAIS-S1 candidate host galaxies.
 3620 Columns are defined in . Full table electronic. SWIRE-RA Dec Expert RGZ host host
 3621 All Compact Resolved All Compact Resolved J002925.73-440256.2 7.3572-44.0490 yes
 3622 no 0.9537 0.8638 0.5552 0.9195 0.9037 0.9371 J002926.14-440249.0 7.3590-44.0470 no no
 3623 0.7361 0.8752 0.5640 0.7740 0.7474 0.7952 J002926.52-440247.0 7.3605-44.0464 no no 0.3390
 3624 0.8338 0.5556 0.7275 0.6894 0.7197 J002926.63-440301.1 7.3610-44.0503 no no 0.2108 0.8251
 3625 0.5623 0.3434 0.3306 0.3292 J002927.13-440232.6 7.3631-44.0424 no no 0.0339 0.8479 0.5669
 3626 0.5853 0.5148 0.5159 J002927.28-440245.3 7.3637-44.0459 no no 0.0406 0.8345 0.5540 0.2702
 3627 0.2340 0.2133 J002927.44-440238.5 7.3644-44.0440 no no 0.0116 0.8267 0.5746 0.2228 0.2182
 3628 0.2028 J002928.08-440230.37 7.3670-44.0418 no no 0.0024 0.8626 0.5791 0.2297 0.1963 0.1549 J002928.11-440312.7
 3629 7.3671-44.0535 no no 0.0011 0.8159 0.5514 0.0377 0.0384 0.0271 J002928.80-440306.87 7.3700
 3630 -44.0519 no no 0.0003 0.8405 0.5668 0.0236 0.0226 0.0136 All Compact Resolved All Compact
 3631 Resolved All Compact Resolved All Compact Resolved 0.9722 0.9955 0.8769 0.9933
 3632 0.9934 0.9658 0.8824 0.9664 0.7950 0.8078 0.9227 0.7677 0.4669 0.0111 0.4249 0.3926 0.2220
 3633 0.5947 0.2077 0.0000 0.1613 0.1876 0.0852 0.4546 0.2264 0.0254 0.2389 0.6275 0.3033 0.6812
 3634 0.1347 0.0857 0.0399 0.3582 0.4854 0.5347 0.0603 0.0007 0.0734 0.0688 0.0141 0.1581 0.0917
 3635 0.0000 0.0399 0.2846 0.1245 0.2833 0.0248 0.0334 0.0301 0.5735 0.5065 0.5265 0.1977 0.1507
 3636 0.0000 0.3334 0.6593 0.3995 0.0173 0.0016 0.0359 0.1056 0.0492 0.1456 0.0000 0.0000 0.0000
 3637 0.0000 0.0000 0.0287 0.0064 0.0049 0.0187 0.1981 0.1534 0.1493 0.0000 0.0000 0.0000 0.1565
 3638 0.1634 0.1284 0.0020 0.0005 0.0239 0.1337 0.1001 0.1310 0.0000 0.0000 0.0358 0.0000 0.0000
 3639 0.0190 0.0008 0.0013 0.0119 0.0280 0.0361 0.0205 0.1171 0.0000 0.0000 0.0873 0.0383 0.0000
 3640 0.0004 0.0014 0.0095 0.0339 0.0408 0.0136 0.0000 0.0000 0.0000 0.1114 0.1480 0.1584

3641 5 ATLAS component cross-identifications

3642 This section contains cross-identifications predicted by our cross-identification method
 3643 for each ATLAS radio component in CDFS and ELAIS-S1. Cross-identifications for
 3644 ATLAS-CDFS components are shown in and cross-identifications for ATLAS-ELAIS-S1
 3645 are shown in . For CDFS, the cross-identification for a component in a quadrant is
 3646 predicted using our method with binary classifiers trained on all other quadrants. For
 3647 ELAIS-S1, we show the cross-identifications predicted by our method using binary
 3648 classifiers trained on each CDFS quadrant. For CDFS, we also show the Radio-Galaxy
 3649 Zoo consensus, which is a proxy for the difficulty of cross-identifying a component
 3650 (Wong et al., in prep.). These are partial tables, and the full tables are available online

3651 at the *Monthly Notices of the Royal Astronomical Society* website¹.

3652 The columns of the cross-identification tables are defined as follows:

- 3653 • *ATLAS*—ATLAS designation for radio component.

- 3654 • *RA*—Right ascension of radio component (J2000).

- 3655 • *Dec*—Declination of radio component (J2000).

- 3656 • *CID*—Radio Galaxy Zoo component ID.

- 3657 • *Zooniverse ID*—Radio Galaxy Zoo Zooniverse ID.

- 3658 • *Norris/Middleberg*—Designation of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.

- 3660 • *Norris/Middleberg RA*—Right ascension of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.

- 3662 • *Norris/Middleberg Dec*—Right ascension of SWIRE cross-identification from Norris et al. (2006) or Middleberg et al. (2008) for CDFS and ELAIS-S1 respectively.

- 3664 • *RGZ*—Designation of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.)

- 3665 • *RGZ RA*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.).

- 3668 • *RGZ Dec*—Right ascension of SWIRE cross-identification from Radio Galaxy Zoo (Wong et al., in prep.).

- 3670 • *RGZ radio consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on the radio component configuration.

- 3672 • *RGZ IR consensus*—Percentage agreement of Radio Galaxy Zoo volunteers on the host galaxy of this radio component.

- 3674 • *C / L / D*—Designation of SWIRE cross-identification made by our method using classification model *C* trained on label set *L* of *D* candidate host galaxies. *C* may be ‘CNN’, ‘LR’ or ‘RF’ for CNN, logistic regression or random forests respectively. *L* may be ‘Norris’ or ‘RGZ’ for expert and Radio Galaxy Zoo labels respectively. *D* may be ‘All’, ‘Compact’ or ‘Resolved’ for each respective subset defined in .

- 3679 • *C / L / D RA*—Right ascension (J2000) of SWIRE cross-identification made by our method using classification model *C* trained on label set *L* of *D* candidate host galaxies. *C*, *L* and *D* are defined as for designation.

- 3682 • $C / L / D$ Dec—Declination (J2000) of SWIRE cross-identification made by our
 3683 method using classification model C trained on label set L of D candidate host
 3684 galaxies. C , L and D are defined as for designation.

3685 Cross-identifications for ATLAS CDFS components. Columns are defined in . Full
 3686 table electronic. ATLAS RA Dec CID Zooniverse ID RA Dec RA Dec radio IR RA Dec
 3687 RA Dec consensus consensus J032602.82-284708.1C 51.5117-28.7856 CI0412 ARG0003rb2
 3688 J032603.15-284708.551.5132-28.7857 0.4516 0.3214 J032615.49-284629.4C 51.5646-28.7749
 3689 CI0614 ARG0003rfr J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642
 3690 -28.7752 0.2941 0.8000 J032615.55-280559.8C 51.5648-28.1000 CI0320 ARG0003r8s J032615.52-280559.8
 3691 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 0.5625 0.8333 J032617.35-280710.2C
 3692 51.5723-28.1195 CI0059C1 ARG0003r2j J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2
 3693 51.5746-28.1187 0.4146 1.0000 J032625.13-280909.8C 51.6047-28.1527 CI0409 ARG0003raz
 3694 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 0.3158 0.6667
 3695 J032629.10-280650.1C 51.6213-28.1139 CI0963 ARG0003ro4 J032629.13-280650.7 51.6214
 3696 -28.1141 J032626.74-280636.7 51.6114-28.1102 0.3333 1.0000 J032629.61-284052.7C 51.6234
 3697 -28.6813 CI0304 ARG0003r8e J032629.54-284055.8 51.6231-28.6822 J032629.54-284055.8
 3698 51.6231-28.6822 0.2676 1.0000 J032629.92-284753.5C 51.6247-28.7982 CI0120 ARG0003r3w
 3699 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 1.0000 0.8571
 3700 J032630.66-283657.3C 51.6278-28.6159 CI0172C1 ARG0003r55 J032630.64-283658.0 51.6277
 3701 -28.6161 J032628.56-283744.8 51.619-28.6291 0.3611 0.7308 J032634.59-282022.8C 51.6441
 3702 -28.3397 CI0757 ARG0003rj2 J032634.58-282022.8 51.6441-28.3397 J032631.96-281941.0
 3703 51.6332-28.3281 0.5781 0.5405 RA Dec RA Dec RA Dec RA Dec J032602.36-284711.5
 3704 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865
 3705 J032602.36-284711.5 51.5098-28.7865 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7
 3706 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.52-280559.8
 3707 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000
 3708 J032615.52-280559.8 51.5647-28.1000 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2
 3709 51.5746-28.1187 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2 51.5746-28.1187 J032625.19-280910.1
 3710 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032624.50-280905.9 51.6021-28.1517
 3711 J032625.19-280910.1 51.6050-28.1528 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7
 3712 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.54-284051.9
 3713 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811
 3714 J032629.54-284051.9 51.6231-28.6811 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4
 3715 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032630.64-283658.0
 3716 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161
 3717 J032630.64-283658.0 51.6277-28.6161 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8
 3718 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397 RA
 3719 Dec RA Dec RA Dec J032604.58-284650.9 51.5191-28.7808 J032602.08-284713.1
 3720 51.5087-28.787 J032602.36-284711.5 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865 J032615.41-284630.7
 3721 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752
 3722 J032615.41-284630.7 51.5642-28.7752 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8
 3723 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.86-280628.8
 3724 51.5661-28.1080 J032615.16-280742.2 51.5632-28.1284 J032615.86-280628.8 51.5661-28.1080

3725 J032618.84-280722.6 51.5785-28.1230 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1
 3726 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032629.13-
 3727 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141
 3728 J032629.13-280650.7 51.6214-28.1141 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9
 3729 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.81-
 3730 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985
 3731 J032629.81-284754.4 51.6242-28.7985 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0
 3732 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032634.58-
 3733 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397
 3734 J032634.58-282022.8 51.6441-28.3397 RA Dec RA Dec RA Dec J032603.15-284708.5
 3735 51.5132-28.7857 J032602.36-284711.5 51.5098-28.7865 J032602.36-284711.5 51.5098-28.7865
 3736 J032602.36-284711.5 51.5098-28.7865 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7
 3737 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752 J032615.41-284630.7 51.5642-28.7752
 3738 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032615.52-280559.8
 3739 51.5647-28.1000 J032615.52-280559.8 51.5647-28.1000 J032617.89-280707.2 51.5746-28.1187
 3740 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2 51.5746-28.1187 J032617.89-280707.2
 3741 51.5746-28.1187 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528
 3742 J032625.19-280910.1 51.6050-28.1528 J032625.19-280910.1 51.6050-28.1528 J032629.13-280650.7
 3743 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141 J032629.13-280650.7 51.6214-28.1141
 3744 J032629.13-280650.7 51.6214-28.1141 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9
 3745 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811 J032629.54-284051.9 51.6231-28.6811
 3746 J032630.12-284751.2 51.6255-28.7976 J032629.81-284754.4 51.6242-28.7985 J032629.81-284754.4
 3747 51.6242-28.7985 J032629.81-284754.4 51.6242-28.7985 J032630.64-283658.0 51.6277-28.6161
 3748 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0 51.6277-28.6161 J032630.64-283658.0
 3749 51.6277-28.6161 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397
 3750 J032634.58-282022.8 51.6441-28.3397 J032634.58-282022.8 51.6441-28.3397
 3751 Cross-identifications for ATLAS ELAIS-S1 components. Columns are defined in
 3752 . Full table electronic. ATLAS RA Dec CID Zooniverse ID RA Dec RA Dec radio
 3753 IR RA Dec RA Dec consensus consensus J002925.68-440256.8 7.3570-44.0491-C0375
 3754 J002925.73-440256.2 7.3572-44.0490 J002938.19-432946.77.4092-43.4963 C0832 J002938.07-432947.9
 3755 7.4087-43.4967 J002940.13-440309.2 7.4172-44.0526 C0374 J002940.19-440309.6 7.4175
 3756 -44.0527 J002943.14-440812.3 7.4298-44.1368 C0302 J002943.15-440813.6 7.4298-44.1371
 3757 J002944.51-433627.8 7.4355-43.6077 C0727 J002944.36-433630.2 7.4348-43.6084 J002945.31-432148.5
 3758 7.4388-43.3635 C0943.1 J002945.64-432149.3 7.4402-43.3637 J002946.14-432149.1 7.4423
 3759 -43.3637 C0943 J002945.64-432149.3 7.4402-43.3637 J002949.89-440541.4 7.4579-44.0948
 3760 C0345 J002951.13-432354.3 7.4631-43.3984 C0924 J002951.14-432355.3 7.4631-43.3987
 3761 J002951.19-440556.6 7.4633-44.0991 C0342 J002951.26-440556.4 7.4636-44.0990 RA Dec
 3762 RA Dec RA Dec RA Dec J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572
 3763 -44.0490 J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490 J002938.07-432947.9
 3764 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002937.50-432945.4 7.4063-43.4959
 3765 J002937.50-432945.4 7.4063-43.4959 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6
 3766 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002943.15-4408
 3767 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371
 3768 J002943.15-440813.6 7.4298-44.1371 J002944.36-433630.2 7.4348-43.6084 J002944.36-433630.2

3769 7.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002945.64-432149.3
 3770 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637
 3771 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3772 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002951.44-440546.1
 3773 7.4644-44.0962 J002951.44-440546.1 7.4644-44.0962 J002951.44-440546.1 7.4644-44.0962
 3774 J002951.44-440546.1 7.4644-44.0962 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3
 3775 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.26-440556.4
 3776 7.4636-44.0990 J002951.44-440546.1 7.4644-44.0962 J002951.51-440617.1 7.4646-44.1048
 3777 J002951.44-440546.1 7.4644-44.0962 RA Dee RA Dee RA Dee J002925.73-440256.2
 3778 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490
 3779 J002925.73-440256.2 7.3572-44.0490 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9
 3780 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002940.19-440309.6
 3781 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527
 3782 J002940.19-440309.6 7.4175-44.0527 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6
 3783 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002944.36-433630.2
 3784 7.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002944.36-433630.27.4348-43.6084
 3785 J002944.36-433630.27.4348-43.6084 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3786 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3787 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637
 3788 J002945.64-432149.3 7.4402-43.3637 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4
 3789 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.14-432355.3
 3790 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987
 3791 J002951.14-432355.3 7.4631-43.3987 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4
 3792 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 RA
 3793 Dee RA Dee RA Dec J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490
 3794 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490 J002925.73-440256.2 7.3572-44.0490
 3795 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002938.07-432947.9
 3796 7.4087-43.4967 J002938.07-432947.9 7.4087-43.4967 J002940.19-440309.6 7.4175-44.0527
 3797 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6 7.4175-44.0527 J002940.19-440309.6
 3798 7.4175-44.0527 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371
 3799 J002943.15-440813.6 7.4298-44.1371 J002943.15-440813.6 7.4298-44.1371 J002944.36-433630.2
 3800 7.4348-43.6084 J002944.36-433630.27.4348-43.6084 J002944.36-433630.27.4348-43.6084
 3801 J002944.36-433630.27.4348-43.6084 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3802 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637
 3803 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002945.64-432149.3
 3804 7.4402-43.3637 J002945.64-432149.3 7.4402-43.3637 J002951.26-440556.4 7.4636-44.0990
 3805 J002951.26-440556.4 7.4636-44.0990 J002949.13-440536.5 7.4547-44.0935 J002949.13-440536.5
 3806 7.4547-44.0935 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987
 3807 J002951.14-432355.3 7.4631-43.3987 J002951.14-432355.3 7.4631-43.3987 J002951.26-440556.4
 3808 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990 J002951.26-440556.4 7.4636-44.0990
 3809 J002951.26-440556.4 7.4636-44.0990

3810 5 Cross-identification figures

3811 This section contains figures of our cross-identifications of each ATLAS radio component
3812 in CDFS and ELAIS-S1, and was originally an appendix to Alger et al. (2018). These
3813 are just five examples, but all 469 examples are available online at the *Monthly Notices*
3814 of the Royal Astronomical Society website¹.

3815 One resolved component and resolved source. Three resolved components comprising
3816 one resolved source. Three resolved components comprising one resolved source.
3817 Three resolved components comprising one resolved source. Two compact components,
3818 each a compact source. Examples of resolved sources with high disagreement between
3819 cross-identifiers. The contours show ATLAS radio data and start at 4σ , increasing
3820 geometrically by a factor of 2. The background image is the SWIRE image. Binary
3821 classifier model/training set combinations are denoted $C(S)$ where C is the binary
3822 classifier model and S is the training set. ‘LR’ is logistic regression, ‘CNN’ is convolutional
3823 neural networks, and ‘RF’ is random forests. ‘Norris’ refers to the expert labels and
3824 ‘RGZ’ refers to the Radio Galaxy Zoo labels. The cross-identification made by nearest
3825 neighbours is shown by ‘NN’. The complete set of figures for 469 examples is available
3826 in the supplementary information online.

3827 5 Sankey diagrams

3828 This section presents Sankey diagrams showing the filtering of components and sources
3829 from the full FIRST sample in, and was originally an appendix to Alger et al. (in prep.)
3830 . A Sankey diagram shows the order and number of objects removed from a sample.
3831 shows the filtering of components and shows the filtering of sources. The component
3832 filters are ‘Bad FIRST’ for components on the edge of FIRST with incomplete images,
3833 ‘Sidelobe’ for components with high sidelobe probability, ‘Low score’ for components
3834 with only low-scoring candidate hosts, ‘Faint’ for components with less than 10 signal-to-noise
3835 according to the FIRST catalogue, and ‘Compact’ for components that do not have
3836 extended radio emission according to . Sources were removed after each component
3837 filter if they no longer contained any components.

3838 Number of components removed from FIRST by each filter.

3839 Number of sources removed by each filter.

3840 5 Radio luminosity function

3841 We computed the radio luminosity function following the $1/V_{\max}$ method (Schmidt, 1968)
3842 . This appendix explains our implementation in and was originally an appendix to
3843 Alger et al. (in prep.). We performed the following steps:

- 3844 1. Remove all radio sources that do not fit the selection criteria. This applies for
3845 both radio and infrared properties, so we choose a minimum radio flux density

3846 f_{\min} and a maximum infrared magnitude m_{\max} , as well as redshift limits z_{lower}
 3847 and z_{upper} .

2. For each source, compute the maximum redshift that the source could have been observed within the selection criteria. We find this redshift by first numerically solving for z with L as the luminosity of each radio source and $f = f_{\min}$ to obtain the maximum redshift z_{radio} at which the source could be observed in radio. We find the maximum redshift z_{ir} that the host galaxy could be observed within the selection criteria by numerically solving for z , where $d(z)$ is the luminosity distance at a redshift z , d is the luminosity distance of the host galaxy, and m is the apparent magnitude of the host galaxy.

$$5 \log_{10} \left(\frac{d(z)}{d} \right) + m = m_{\max}$$

3848 The maximum redshift that the source could have been observed within the
 3849 selection criteria is then $z_{\max} = \min(z_{\text{ir}}, z_{\text{radio}}, z_{\text{upper}})$.

3. For each source, compute the comoving volume V_{\max} at redshift z_{\max} .
4. The count for each luminosity bin is the sum over $1/V_{\max}$ for each source in the bin. We account for the fact FIRST does not cover the whole sky by multiplying by the total area of the sky divided by the area of our selection.

3854 After computing the luminosity function, we estimate the uncertainty in each bin
 3855 using Poisson statistics, \sqrt{N} for a bin count N .

3856 5 Redshift completeness estimate

3857 shows the estimated completeness of our RLF sample in as a function of $W1$ and
 3858 $W1 - W2$. We followed the same method as Pracy et al. (2016) for this estimation,
 3859 averaging completeness over circles centred on each source. Each source is associated
 3860 with a circle of radius equal to the distance to its 50th nearest neighbour in the $W1$ and
 3861 $W1 - W2$ plane. This appendix was originally part of Alger et al. (in prep.).

3862 Estimated completeness as a function of mid-infrared colour and magnitude.

3863 5 Giant radio galaxies

3864 Giant radio galaxies found in RGZ. Ex. ‘LLS’ is the projected linear size of the source as
 3865 measured by the maximum angular distance between radio components. The RA/Dec
 3866 are the coordinates of the host galaxy. s/p indicates spectroscopic/photometric redshift.
 3867 ^LExisting in literature. ^RAlso found by RGZ citizen scientists. [†]Misidentified SDSS
 3868 host, manually corrected to obtain redshift. AllWISE host (WISEA) RA (J2000) Dec
 3869 ($J2000$) z LLS (Mpc) $J004210.18-080011.3$ $10.54-8.00$ 0.65 ± 0.14 1.6 p $J021008.48+011839.6$
 3870 $32.54-1.31$ 0.86524 ± 0.0001 1.2 s $J075858.29+355643.6$ $119.74-35.95$ 0.74748 ± 0.00013

3871 ~~1.0 sJ080831.68+473523.9^R 122.13 47.59 0.58854 ± 0.00016 1.1 sJ083034.78+231124.6 127.64~~
 3872 ~~23.19 0.94 ± 0.13 1.1 pJ090604.03+011114.2 136.52 1.19 0.7975 ± 0.0004 1.6 sJ093256.81+074212.2~~
 3873 ~~143.24 7.70 1.0032 ± 0.0003 1.1 sJ093526.80+051729.8^R 143.86 5.29 0.84 ± 0.04 1.2 pJ094238.72+114337.9~~
 3874 ~~145.66 11.73 0.49 ± 0.05 1.2 pJ094835.60+535946.4^R 147.15 54.00 0.64 ± 0.10 1.2 pJ095706.12+292439.2~~
 3875 ~~149.28 29.41 0.71 ± 0.12 1.5 pJ102335.25+433208.0 155.90 43.54 0.75 ± 0.09 1.5 pJ102933.99+210345.8^R~~
 3876 ~~157.39 21.06 0.82407 ± 0.00008 1.1 sJ103043.98+355451.2^R 157.68 35.91 0.64074 ± 0.00008~~
 3877 ~~1.2 sJ104449.92+234525.6[†] 161.20 23.76 0.57712 ± 0.00009 1.6 sJ110655.98+624759.8^R~~
 3878 ~~166.73 62.80 0.84379 ± 0.00004 1.1 sJ112900.68+635543.2 172.25 63.93 0.71 ± 0.06 1.1~~
 3879 ~~pJ112948.20+243922.6 172.45 24.66 0.79 ± 0.07 1.1 pJ114553.67-003304.7 176.47 -0.55~~
 3880 ~~2.0522 ± 0.0006 1.3 sJ121111.26+534840.4 182.80 53.81 0.74 ± 0.14 1.1 pJ121152.04+304232.4^R~~
 3881 ~~182.97 30.71 0.47102 ± 0.00012 1.3 sJ121944.73+174121.3 184.94 17.69 1.5129 ± 0.0009~~
 3882 ~~1.0 sJ123735.89+544814.4^R 189.40 54.80 1.0271 ± 0.0006 1.2 sJ123819.16+113444.8 189.58~~
 3883 ~~11.58 0.80 ± 0.08 1.2 pJ123846.84-032857.5[†] 189.70 -3.48 0.67 ± 0.07 1.5 pJ131625.00+272042.8~~
 3884 ~~199.10 27.35 0.69092 ± 0.00004 1.0 sJ133307.00+045048.6^R 203.28 4.85 1.40534 ± 0.00016~~
 3885 ~~1.1 sJ141933.36+104706.4^R 214.89 10.79 0.33973 ± 0.00003 1.0 sJ142008.45+185422.7^R~~
 3886 ~~215.04 18.91 0.63 ± 0.04 1.4 pJ145057.28+530007.7^L 222.74 53.00 0.91662 ± 0.00009 1.3~~
 3887 ~~sJ150012.18+604941.3 225.05 60.83 1.6626 ± 0.0007 1.2 sJ153547.13+432245.0^R 233.95~~
 3888 ~~43.38 0.63891 ± 0.00007 1.3 sJ154631.18+194819.9 236.63 19.81 0.5917 ± 0.0002 1.4 sJ160852.10+561110.7~~
 3889 ~~242.22 56.19 1.3196 ± 0.0003 1.3 sJ162200.48+364044.0 245.50 36.68 1.9994 ± 0.0002 1.1~~
 3890 ~~sJ163004.35+103321.9^R 247.52 10.56 0.85 ± 0.09 1.2 pJ163125.75+200224.1^R 247.86 20.04~~
 3891 ~~0.62662 ± 0.00013 1.0 sJ165055.46+394446.6 252.73 39.75 0.58829 ± 0.00013 1.1 sJ232410.33+045309.6~~
 3892 ~~351.04 4.89 0.76 ± 0.06 1.4 pJ234440.02-003231.6 356.17 -0.54 0.5014 ± 0.0001 1.0 s~~

3893 This appendix describes our search for giant radio galaxies in RGZ-Ex, and the
 3894 results of this search. It was originally an appendix to Alger et al. (in prep.). To identify
 3895 radio sources we assumed that if any two components had the same host galaxy then
 3896 they were part of the same source. This is a reasonable assumption if all host galaxies
 3897 are correctly identified, which was not the case. This assumption therefore introduced
 3898 spurious sources due to galaxies incorrectly identified as host galaxies: not all sources
 3899 used in are real sources, and in particular sources of large angular size are likely to be
 3900 incorrect. Nevertheless RGZ-Ex provides a useful catalogue of *candidate* radio sources,
 3901 and visual follow-up can confirm whether sources of interest are real.

3902 H.A. and M.J.A. examined all 296 candidate sources in the RGZ-Ex catalogue with
 3903 an estimated physical extent larger than 1 Mpc. Of these, 40 were real giant radio
 3904 galaxies, which we show in . We defined ‘giant radio galaxy’ as a radio galaxy with
 3905 emission extended to physical sizes ≥ 1.0 Mpc. Other thresholds, such as 0.7 Mpc,
 3906 also exist in literature. The physical extents of the remaining 256 candidate sources
 3907 were overestimated mostly due to sidelobes/artefacts (103), incorrect source grouping
 3908 (82), or incorrect SDSS matches (21). The citizen scientists who identified giants are:
 3909 WizardHowl, DolorousEdd, antikodon, csunjoto, sisifolibre, JeanTate, JKD, PADV, and
 3910 firejuggler. H.A., together with his summer students, had previously identified 29 of
 3911 these giants.

3912 Note that this is a particularly challenging set: sources that are misidentified will
 3913 often have unusually large estimated extents due to the inclusion of spurious components.
 3914 The error rate in this set therefore does not reflect the rest of the catalogue.

3915 **5 Visual verification results**

3916 In we described our visual verification of the BXID method from . We list the radio
 3917 components in the verification set in . Each row of the table contains the FIRST component,
 3918 its AllWISE host galaxy according to BXID, and whether the association is correct
 3919 according to our visual verification. If an author was particularly unsure about an
 3920 object, they were able to skip this object, and so are not accounted for in the verification
 3921 for that object. Verification was weighted by the Dawid and Skene (1979) maximum
 3922 likelihood model. This appendix was originally part of (Alger et al., in prep.).

3923 Validation objects. ‘Agree’ is whether or not the authors of Alger et al. (in prep.)
 3924 agreed with BXID associating the given FIRST object with the given AllWISE object.
 3925 FIRST AllWISE AgreeJ000234.9-001421 J000242.35-001320.5 nJ002841.1+141654 J002840.37+141652.7
 3926 yJ003731.4+000156 J003731.26+000146.7 yJ005407.5-011158 J005407.61-011158.9 yJ011210.3+002203
 3927 J011210.41+002201.9 yJ012342.4+015849 J012342.24+015850.4 yJ013015.1+110653 J013015.16+110653.4
 3928 yJ013107.7+070343 J013102.02+070332.0 yJ014247.9-000039 J014247.81-000040.3 yJ014250.0-000032
 3929 J014247.81-000040.3 nJ020222.3+030138 J020223.20+030150.4 yJ020333.8+000853 J020336.94+000759.3
 3930 yJ021840.1-032311 J021840.13-032306.0 yJ023022.0+010834 J023022.11+010840.0 yJ024245.3-022535
 3931 J024245.35-022534.6 yJ025901.0+005350 J025901.50+005346.1 yJ033204.1-004757 J033204.15-004757.1
 3932 yJ073033.2+390413 J073033.21+390412.9 yJ073954.1+481810 J073954.87+481759.5 yJ074504.9+331247
 3933 J074504.81+331256.2 yJ074640.4+421709 J074640.45+421709.1 yJ074707.9+171719 J074708.35+171726.5
 3934 yJ075043.6+274838 J075043.35+274844.8 nJ075050.3+331937 J075051.25+331905.0 yJ075422.2+311253
 3935 J075422.35+311252.5 yJ075637.0+212006 J075636.65+212001.4 yJ082326.1+141438 J082326.34+141435.9
 3936 yJ082422.5+351121 J082422.65+351114.6 yJ082925.9+462618 J082926.02+462618.5 yJ083512.4+175441
 3937 J083512.45+175441.1 yJ084133.5+402035 J084133.40+402042.8 yJ084238.4+405305 J084238.38+405306.6
 3938 nJ084417.3+315845 J084417.92+315845.9 yJ084728.5+360700 J084728.24+360714.6 yJ084905.5+111448
 3939 J084905.51+111447.8 yJ085236.8+262006 J085236.11+262013.4 yJ085415.6+524930 J085415.62+524936.7
 3940 yJ090623.2+300746 J090622.87+300743.9 yJ091745.1+275049 J091745.89+275103.8 yJ091752.0+431614
 3941 J091752.14+431612.7 yJ092014.4+302907 J092013.95+302859.3 yJ092140.5+540118 J092140.24+540121.1
 3942 yJ092213.0+542157 J092213.03+542157.2 yJ092406.9+562703 J092406.47+562656.2 yJ092713.1+105841
 3943 J092713.14+105839.8 yJ093108.6+613447 J093108.63+613447.2 yJ093239.6+052308 J093237.71+052240.7
 3944 nJ093627.8+103610 J093627.87+103609.7 yJ093645.2+561435 J093645.89+561434.2 yJ094006.8+482651
 3945 J094006.92+482649.2 yFIRST AllWISE AgreeJ094009.5+600403 J094011.55+600357.6 nJ094023.7+135123
 3946 J094023.73+135125.2 yJ094324.5+435341 J094324.61+435342.0 yJ094650.8+382015 J094650.44+382010.9
 3947 yJ095011.8+455319 J095011.82+455320.0 yJ095113.5+180211 J095113.82+180204.2 nJ095242.4+222638
 3948 J095242.45+222638.0 yJ095538.7+013546 J095539.20+013546.1 yJ095609.9+363441 J095609.30+363445.4
 3949 yJ095811.8+225056 J095811.90+225055.5 yJ100019.2+263516 J100018.84+263527.5 yJ101315.9+064520
 3950 J101316.51+064519.0 yJ101455.2-004716 J101455.30-004718.3 yJ102153.5+260429 J102153.52+260429.6
 3951 yJ102354.7+390653 J102354.88+390654.0 yJ102620.4+303600 J102620.46+303550.4 yJ102710.4+460254
 3952 J102714.81+460256.4 nJ102955.9+424906 J102955.96+424906.7 yJ103503.9+102404 J103503.92+102403.6
 3953 yJ103839.9+331200 J103839.94+331201.1 yJ104030.5+211624 J104031.09+211620.6 nJ104533.8+430025
 3954 J104535.22+430020.8 yJ104907.5+322903 J104907.91+322906.6 yJ105146.9+552257 J105147.40+552308.4
 3955 yJ105257.5+105418 J105257.53+105421.5 yJ105521.6+372641 J105521.24+372652.4 yJ105758.8+321605
 3956 J105758.84+321605.3 yJ110104.9+151618 J110104.90+151618.2 yJ110353.2+352320 J110353.37+352319.9
 3957 yJ110414.4+481345 J110423.08+481311.0 nJ111057.7+220756 J111057.18+220758.3 yJ111208.5+275207

3958 J111201.79+275053.8 nJ111225.2+233159 J111225.30+233157.9 yJ111726.3+375336 J111726.35+375337.
 3959 yJ111746.1+261151 J111746.18+261150.9 yJ111854.3+424708 J111854.45+424652.8 yJ112124.4+640417
 3960 J112125.02+640408.6 yJ112135.3+352330 J112135.44+352324.9 yJ112550.9+200631 J112558.75+200554.
 3961 yJ112859.7+260923 J112859.86+260911.3 yJ113201.1+442639 J113201.23+442639.4 yJ113302.5+355408
 3962 J113301.80+355415.3 yJ113712.7+263301 J113711.86+263335.1 yJ113756.3+471314 J113756.31+471314.
 3963 yJ113906.6+230602 J113906.68+230602.1 yJ114325.0+600721 J114323.90+600737.1 yJ114759.7+370305
 3964 J114759.22+370311.2 yJ114916.7+083022 J114916.33+083040.5 nJ115010.9+063340 J115010.93+063340.
 3965 yJ115308.6+374851 J115316.96+374850.0 yFIRST AllWISE AgreeJ115448.7+472222 J115448.67+472223
 3966 yJ115603.7+584704 J115603.48+584706.1 yJ115605.9+343230 J115605.64+343229.4 yJ115653.0+572338
 3967 J115645.38+572151.7 yJ120138.0+230922 J120137.97+230922.2 yJ120752.8+533808 J120752.85+533807.
 3968 yJ120943.3-021934 J120942.89-021943.0 yJ121045.6+190225 J121045.68+190227.0 yJ121207.6+115412
 3969 J121207.72+115413.8 yJ121211.3+485951 J121211.86+485952.0 yJ121406.7+002634 J121406.73+002635.
 3970 yJ122518.0+350258 J122517.85+350301.9 yJ122525.1+451530 J122524.71+451508.5 yJ122640.9+430508
 3971 J122640.82+430509.2 yJ123429.8+260107 J123434.79+260134.3 nJ123633.1+100928 J123633.12+100928.
 3972 yJ124839.3+411522 J124839.42+411522.3 nJ125129.2+551012 J125128.76+551009.3 yJ130005.8+524801
 3973 J130006.14+524803.0 yJ130132.1+511351 J130132.32+511352.5 yJ131104.4+464936 J131104.45+464934.
 3974 yJ131452.2+252811 J131446.81+252820.8 nJ132033.8+332639 J132033.59+332639.0 nJ132257.5+191134
 3975 J132257.53+191133.9 yJ132529.3+230734 J132529.35+230733.8 yJ132546.8+052453 J132546.86+052454.
 3976 yJ132637.7+112110 J132637.92+112108.8 yJ132831.8+104339 J132831.88+104338.8 yJ132932.3+131839.
 3977 J132932.32+131839.6 yJ133022.8+311904 J133022.83+311902.8 yJ133453.3+405653 J133454.13+405650.
 3978 yJ133741.1+124302 J133741.13+124303.1 yJ133823.6+103337 J133823.67+103341.9 yJ134651.2+415154.
 3979 J134651.06+415156.1 yJ134704.3+110622 J134704.35+110622.7 yJ134752.7+555046 J134752.71+555048.
 3980 yJ134831.7+164325 J134831.57+164328.2 yJ134949.8+385539 J134949.93+385542.8 yJ135106.5+074534.
 3981 J135106.50+074534.2 yJ135107.7+615502 J135107.75+615502.1 yJ135658.5+134028 J135659.15+134017.
 3982 yJ135833.9+180021 J135834.03+180020.4 yJ140630.7+554017 J140629.32+554009.9 yJ140804.2+503019.
 3983 J140804.10+503021.1 yJ141226.7+454125 J141226.54+454125.5 yJ141245.0+495213 J141243.84+495206.
 3984 yJ141317.4+325306 J141317.50+325306.8 yJ141723.8+543639 J141724.33+543629.5 yJ141938.8+312146.
 3985 J141940.16+312138.8 yJ142515.3+175526 J142513.89+175525.7 yFIRST AllWISE AgreeJ142829.5+070835.
 3986 J142829.60+070836.3 yJ143411.0+170036 J143411.18+170035.7 yJ143624.0-001057 J143623.89-001100.8.
 3987 yJ143742.6+104412 J143742.69+104412.8 yJ143840.8+475355 J143841.08+475356.1 yJ143909.1+430847.
 3988 J143909.08+430847.8 yJ144135.8+102246 J144135.91+102245.1 yJ144333.6+275229 J144333.02+275250.
 3989 yJ145012.3+471739 J145012.33+471738.7 yJ145103.7+452459 J145102.66+452520.5 nJ145401.6+141009.
 3990 J145401.70+141009.6 yJ150158.7+191413 J150158.87+191405.3 yJ150743.9+352720 J150743.62+352724.
 3991 yJ151141.6-003209 J151142.01-003213.0 yJ151315.5+403107 J151315.56+403107.7 yJ151518.7+230256.
 3992 J151518.67+230257.3 yJ151703.6+105947 J151703.68+105947.6 yJ151736.8+610856 J151736.83+610857.
 3993 yJ152121.6+281635 J152120.68+281626.2 yJ152714.8+310425 J152714.88+310424.7 yJ153428.9+272134.
 3994 J153429.68+272120.8 yJ154245.3+100919 J154245.71+100917.8 yJ154901.6+103159 J154901.40+103152.
 3995 yJ154925.2+395316 J154926.17+395303.7 yJ155206.3-005348 J155206.58-005339.3 yJ155457.3+344637.
 3996 J155458.45+344644.7 yJ155743.5+272752 J155743.52+272752.8 yJ160130.0+083848 J160130.07+083850.
 3997 yJ160534.8+441220 J160535.55+441221.5 yJ160859.2+400135 J160901.32+400230.7 nJ161545.4+231617.
 3998 J161545.14+231617.2 yJ161930.4+085533 J161930.51+085532.6 yJ162228.0+264743 J162228.70+264736.
 3999 yJ162750.4+473624 J162750.55+473623.5 yJ162904.2+470852 J162904.34+470853.0 yJ163038.7+214740.
 4000 J163037.43+214748.9 nJ163323.6+424051 J163323.61+424051.9 yJ163327.5+242426 J163327.87+242427.
 4001 yJ163533.8+454557 J163534.00+454554.3 yJ164211.2+512029 J164211.27+512029.3 yJ165549.1+375923.

4002 J165549.01+375923.6 yJ165620.0+363402J165619.89+363403.9 yJ165700.5+474820J165659.58+474809.0
 4003 yJ171406.2+292712J171404.16+292704.0 nJ172126.4+374446J172126.46+374446.6 yJ222627.7-005010
 4004 J222627.77-005010.8 yJ223636.4-013827J223636.48-013827.2 yJ225619.0+143257J225621.96+143351.4
 4005 yJ232410.1+001315J232410.15+001314.5 yJ234727.9-000919J234727.65-000912.9 y

4006 5 2-Wasserstein begets Faraday moments

Minimising the 2-Wasserstein distance between a model FDF and the simple manifold gives the second Faraday moment of that FDF. This appendix demonstrates that fact, and was originally part of Alger et al. (2021). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the sum-normalised simple model FDF:

$$\begin{aligned}\tilde{F}(\phi) &= \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1} \\ \tilde{S}(\phi; \phi_w) &= \delta(\phi - \phi_w).\end{aligned}$$

The W_2 distance, usually defined on probability distributions, can be extended to one-dimensional complex functions A and B by normalising them:

$$\begin{aligned}D_{W_2}(A \parallel B)^2 &= \inf_{\gamma \in \Gamma(A, B)} \iint_{\phi_{\min}}^{\phi_{\max}} |x - y|^2 d\gamma(x, y) \\ \tilde{A}(\phi) &= \frac{|A(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |A(\theta)| d\theta} \\ \tilde{B}(\phi) &= \frac{|B(\phi)|}{\int_{\phi_{\min}}^{\phi_{\max}} |B(\theta)| d\theta}\end{aligned}$$

4007 where $\Gamma(A, B)$ is the set of couplings of A and B , i.e. the set of joint probability distributions
 4008 that marginalise to A and B ; and $\inf_{\gamma \in \Gamma(A, B)}$ is the infimum over $\Gamma(A, B)$. This can be
 4009 interpreted as the minimum cost to ‘move’ one probability distribution to the other,
 4010 where the cost of moving one unit of probability mass is the squared distance it is
 4011 moved.

The set of couplings $\Gamma(\tilde{F}, \tilde{S})$ is the set of all joint probability distributions γ such that

$$\begin{aligned}\int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\phi &= \tilde{S}(\varphi; \phi_w), \\ \int_{\phi_{\min}}^{\phi_{\max}} \gamma(\phi, \varphi) d\varphi &= \tilde{F}(\phi).\end{aligned}$$

The coupling that minimises the integral in will be the optimal transport plan between

\tilde{F} and \tilde{S} . Since \tilde{F} and \tilde{S} are defined in terms of delta functions, the optimal transport problem reduces to a discrete optimal transport problem and the optimal transport plan is:

$$\gamma(\phi, \varphi) = \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1}\delta(\varphi - \phi_w).$$

In other words, to move the probability mass of \tilde{S} to \tilde{F} , a fraction $A_0/(A_0 + A_1)$ is moved from ϕ_w to ϕ_0 and the complementary fraction $A_1/(A_0 + A_1)$ is moved from ϕ_w to ϕ_1 . Then:

$$\begin{aligned} D_{W_2}(\tilde{F} \parallel \tilde{S})^2 &= \iint_{\phi_{\min}}^{\phi_{\max}} |\phi - \varphi|^2 d\gamma(\phi, \varphi) \\ &= \frac{A_0(\phi_0 - \phi_w)^2 + A_1(\phi_1 - \phi_w)^2}{A_0 + A_1}. \end{aligned}$$

To obtain the W_2 distance to the simple manifold, we need to minimise this over ϕ_w . Differentiate with respect to ϕ_w and set equal to zero to find

$$\phi_w = \frac{A_0\phi_0 + A_1\phi_1}{A_0 + A_1}.$$

Substituting this back in, we find

$$D_{W_2}(F)^2 = \frac{A_0 A_1}{A_0 + A_1} (\phi_0 - \phi_1)^2$$

4012 which is the Faraday moment.

4013 5 Euclidean distance in the no-RMSF case

In this appendix, originally from Alger et al. (2021), we calculate the minimised Euclidean distance evaluated on a model FDF (). Let \tilde{F} be the sum-normalised model FDF and let \tilde{S} be the normalised simple model FDF:

$$\tilde{F}(\phi) = \frac{A_0\delta(\phi - \phi_0) + A_1\delta(\phi - \phi_1)}{A_0 + A_1}$$

$$\tilde{S}(\phi; \phi_e) = \delta(\phi - \phi_e).$$

The Euclidean distance between \tilde{F} and \tilde{S} is then

$$\begin{aligned} & D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \\ &= \int_{\phi_{\min}}^{\phi_{\max}} |\tilde{F}(\phi) - \delta(\phi - \phi_e)|^2 d\phi. \end{aligned}$$

Assume $\phi_0 \neq \phi_1$ (otherwise, D_E will always be either 0 or 2). If $\phi_e = \phi_0$, then

$$\begin{aligned} & D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 \\ &= \frac{1}{(A_0 + A_1)^2} \int_{\phi_{\min}}^{\phi_{\max}} A_1^2 |\delta(\phi - \phi_1) - \delta(\phi - \phi_0)|^2 d\phi \\ &= \frac{2A_1^2}{(A_0 + A_1)^2} \end{aligned}$$

and similarly for $\phi_e = \phi_1$. If $\phi_e \neq \phi_0$ and $\phi_e \neq \phi_1$, then

$$D_E(\tilde{F}(\phi) \parallel \tilde{S}(\phi; \phi_e))^2 = \frac{A_0^2 + A_1^2 + 1}{(A_0 + A_1)^2}.$$

The minimised Euclidean distance when $\phi_0 \neq \phi_1$ is therefore

$$\begin{aligned} D_E(F) &= \min_{\phi_e \in \mathbb{R}} D_E(F(\phi) \parallel F_{\text{simple}}(\phi; \phi_e)) \\ &= \sqrt{2} \frac{\min(A_0, A_1)}{A_0 + A_1}. \end{aligned}$$

4014 If $\phi_0 = \phi_1$, then the minimised Euclidean distance is 0.

4015 5 Hyperparameters for LR and XGB

4016 This section contains tables of the hyperparameters that we used for our classifiers in
 4017 and was originally an appendix to Alger et al. (2021). and tabulate the hyperparameters
 4018 for XGB and LR respectively for the 'ATCA' dataset. and tabulate the hyperparameters
 4019 for XGB and LR respectively for the 'ASKAP' dataset.

4020 XGB hyperparameters for the 'ATCA' dataset. Parameter Value colsample_bytree
 4021 0.912 gamma 0.532 learning_rate 0.1 max_depth 7 min_child_weight 2 scale_pos_weight
 4022 1 subsample 0.557 n_estimators 135 reg_alpha 0.968 reg_lambda 1.420

4023 LR hyperparameters for the 'ATCA' dataset. Parameter Value penalty L1 C 1.668

4024 XGB hyperparameters for the 'ASKAP' dataset. Parameter Value colsample_bytree
 4025 0.865 gamma 0.256 learning_rate 0.1 max_depth 6 min_child_weight 1 scale_pos_weight
 4026 1 subsample 0.819 n_estimators 108 reg_alpha 0.049 reg_lambda 0.454

4027 LR hyperparameters for the ‘ASKAP’ dataset. Parameter Valuepenalty L2C 0.464

4028 5 Predictions on real data

4029 This appendix, originally part of Alger et al. (2021), contains and . These show the
 4030 predicted probability of being Faraday complex for all real data used in , drawn from
 4031 Livingston et al. (2021) and O’Sullivan et al. (2017).

4032 The 142 observed FDFs ordered by LR-estimated probability of being Faraday complex.
 4033 Livingston-identified components are shown in orange while O’Sullivan-identified components
 4034 are shown in magenta. Simpler FDFs (as deemed by the classifier) are shown in purple
 4035 while more complex FDFs are shown in green, and the numbers overlaid indicate the
 4036 LR estimate. A lower number indicates a lower probability that the corresponding
 4037 source is complex, i.e. lower numbers correspond to simpler spectra.

4038 The 142 observed FDFs ordered by XGB-estimated probability of being Faraday
 4039 complex. Livingston-identified components are shown in orange while O’Sullivan-identified
 4040 components are shown in magenta. Simpler FDFs (as deemed by the classifier) are
 4041 shown in purple while more complex FDFs are shown in green, and the numbers
 4042 overlaid indicate the XGB estimate. A lower number indicates a lower probability
 4043 that the corresponding source is complex, i.e. lower numbers correspond to simpler
 4044 spectra.

4045 5 Simulating observed FDFs

This appendix was originally part of Alger et al. (2021) and describes how we simulated
 FDFs in . We simulated FDFs by approximating them by arrays of complex numbers.
 An FDF F is approximated on the domain $[-\phi_{\max}, \phi_{\max}]$ by a vector $\vec{F} \in \mathbb{R}^d$:

$$\vec{F}_j = \sum_{k=0}^1 A_k \delta(-\phi_{\max} + j\delta\phi - \phi_k)$$

where $\delta\phi = (\phi_{\max} - \phi_{\min})/d$ and d is the number of Faraday depth samples in the FDF.
 \vec{F} is sampled by uniformly sampling its parameters:

$$\underline{\phi_k \in [\phi_{\min}, \phi_{\min} + \delta\phi, \dots, \phi_{\max}]}$$

$$\underline{A_k \sim \mathcal{U}(0, 1)}.$$

We then generate a vector polarisation spectrum $\vec{P} \in \mathbb{R}^m$ from \vec{F} using a :

$$\vec{P}_{\ell} = \sum_{j=0}^j F_j e^{2i(\phi_{\min} + j\delta\phi)\lambda_{\ell}^2} d\phi.$$

λ_ℓ^2 is the discretised value of λ^2 at the ℓ th index of \vec{P} . This requires a set of λ^2 values, which depends on the dataset being simulated. These values can be treated as the channel wavelengths at which the polarisation spectrum was observed. We then add Gaussian noise with variance σ^2 to each element of \vec{P} to obtain a discretised noisy observation $\hat{\vec{P}}$. Finally, we perform RM synthesis using the Canadian Initiative for Radio Astronomy Data Analysis RM package¹, which is a Python module that implements a discrete version of RM synthesis:

$$\hat{F}_j = m^{-1} \sum_{\ell=1}^m \vec{P}_\ell e^{-2i(\phi_{\min} + j\delta_\phi)\lambda_\ell^2}.$$

¹

Bibliography

- 4048 Agarwal, D., Aggarwal, K., Burke-Spolaor, S., Lorimer, D. R., & Garver-Daniels, N.
 4049 (2020). FETCH: A deep-learning based classifier for fast transient classification.
 4050 *MNRAS*, 497, 1661–1674. <https://doi.org/10.1093/mnras/staa1856>
- 4051 Aguado, D. S., Ahumada, R., Almeida, A., Anderson, S. F., Andrews, B. H., Anguiano,
 4052 B., Aquino Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., & Aubert,
 4053 M. e. a. (2019). The fifteenth data release of the Sloan Digital Sky Surveys: First
 4054 release of MaNGA-derived quantities, data visualization tools, and stellar li-
 4055 brary. *ApJS*, 240(2). <https://doi.org/10.3847/1538-4365/aaf651>
- 4056 Alger, M. J., Banfield, J. K., Ong, C. S., Rudnick, L., Wong, O. I., Wolf, C., Andernach, H.,
 4057 Norris, R. P., & Shabala, S. S. (2018). Radio Galaxy Zoo: Machine learning for
 4058 radio source host galaxy cross-identification. *MNRAS*, 478, 5547–5563. <https://doi.org/10.1093/mnras/sty1308>
- 4059 4060 Alger, M. J., Livingston, J. D., McClure-Griffiths, N. M., Nabaglo, J. L., Wong, O. I.,
 4061 & Ong, C. S. (2021). Interpretable Faraday complexity classification. *PASA*, 38.
 4062 <https://doi.org/10.1017/pasa.2021.10>
- 4063 4064 Alger, M. J., Wong, O. I., Ong, C. S., McClure-Griffiths, N. M., Andernach, H., Rudnick,
 4065 L., Shabala, S. S., Garon, A. F., Banfield, J. K., Kapińska, A. D., Norris, R. P., &
 4066 Thomson, A. J. M. (in prep.). Radio Galaxy Zoo: Radio luminosity functions of
 extended sources.
- 4067 4068 Alhassan, W., Taylor, A. R., & Vaccari, M. (2018). The FIRST Classifier: Compact and
 4069 extended radio galaxy classification using deep convolutional neural networks.
MNRAS, 480, 2085–2093. <https://doi.org/10.1093/mnras/sty2038>
- 4070 4071 Anderson, C. S., Gaensler, B. M., Feain, I. J., & Franzen, T. M. O. (2015). Broadband
 4072 radio polarimetry and Faraday rotation of 563 extragalactic radio sources. *ApJ*,
 815(1). <https://doi.org/10.1088/0004-637X/815/1/49>
- 4073 4074 Aniyan, A. K., & Thorat, K. (2017). Classifying radio galaxies with the convolutional
 4075 neural network. *ApJS*, 230. <https://doi.org/10.3847/1538-4365/aa7333>
- 4076 Antonucci, R. (1993). Unified models for active galactic nuclei and quasars. *ARA&A*,
 31(1), 473–521. <https://doi.org/10.1146/annurev.aa.31.090193.002353>
- 4077 4078 Arsioli, B., & Dedin, P. (2020). Machine learning applied to multifrequency data in
 4079 astrophysics: Blazar classification. *MNRAS*, 498, 1750–1764. <https://doi.org/10.1093/mnras/staa2449>

- 4080 Balakrishnan, V., Champion, D., & Barr, E. (2020). Pulsar candidate identification us-
4081 ing semi-supervised generative adversarial networks. *arXiv e-prints*, arXiv:
4082 2010.07457v1 [astro-ph.IM].
- 4083 Ball, N. M., Brunner, R. J., Myers, A. D., & Tcheng, D. (2006). Robust machine learn-
4084 ing applied to astronomical data sets. I. Star-galaxy classification of the Sloan
4085 Digital Sky Survey DR3 using decision trees. *ApJ*, 650(1), 497–509. <https://doi.org/10.1086/507440>
- 4087 Ball, N. M., Loveday, J., Fukugita, M., Nakamura, O., Okamura, S., Brinkmann, J., &
4088 Brunner, R. J. (2004). Galaxy types in the Sloan Digital Sky Survey using su-
4089 pervised artificial neural networks. *MNRAS*, 348, 1038–1046. <https://doi.org/10.1111/j.1365-2966.2004.07429.x>
- 4091 Bañados, E., Venemans, B. P., Mazzucchelli, C., Farina, E. P., Walter, F., Wang, F., Decarli,
4092 R., Stern, D., Fan, X., Davies, F. B., Hennawi, J. F., Simcoe, R. A., Turner, M. L.,
4093 Rix, H.-W., Yang, J., Kelson, D. D., Rudie, G. C., & Winters, J. M. (2018). An 800-
4094 million-solar-mass black hole in a significantly neutral Universe at a redshift of
4095 7.5. *Nature*, 553(7689), 473–476. <https://doi.org/10.1038/nature25180>
- 4096 Banerji, M., Lahav, O., Lintott, C. J., Abdalla, F. B., Schawinski, K., Bamford, S. P., An-
4097 dreescu, D., Murray, P., Raddick, M. J., Slosar, A., Szalay, A., Thomas, D., &
4098 Vandenberg, J. (2010). Galaxy Zoo: Reproducing galaxy morphologies via ma-
4099 chine learning. *MNRAS*, 406, 342–353. <https://doi.org/10.1111/j.1365-2966.2010.16713.x>
- 4101 Banfield, J. K., Andernach, H., Kapińska, A. D., Rudnick, L., Hardcastle, M. J., Cotter,
4102 G., Vaughan, S., Jones, T. W., Heywood, I., Wing, J. D., Wong, O. I., Matorny,
4103 T., Terentev, I. A., López-Sánchez, Á. R., Norris, R. P., Seymour, N., Shabala,
4104 S. S., & Willett, K. W. (2016). Radio Galaxy Zoo: Discovery of a poor cluster
4105 through a giant wide-angle tail radio galaxy. *MNRAS*, 460, 2376–2384. <https://doi.org/10.1093/mnras/stw1067>
- 4107 Banfield, J. K., Wong, O. I., Willett, K. W., Norris, R. P., Rudnick, L., Shabala, S. S., Sim-
4108 mons, B. D., Snyder, C., Garon, A., Seymour, N., Middelberg, E., Andernach,
4109 H., Lintott, C. J., Jacob, K., Kapińska, A. D., Mao, M. Y., Masters, K. L., Jarvis,
4110 M. J., Schawinski, K., ... Whyte, L. (2015). Radio Galaxy Zoo: Host galaxies and
4111 radio morphologies derived from visual inspection. *MNRAS*, 453, 2326–2340.
4112 <https://doi.org/10.1093/mnras/stv1688>
- 4113 Bastien, D., Oozeer, N., & Somanah, R. (2017). Classifying bent radio galaxies from a
4114 mixture of point-like/extended images with machine learning. *IOP Conference
4115 Series: Materials Science and Engineering*, 198. [https://doi.org/10.1088/1757-899X/198/1/012013](https://doi.org/10.1088/1757-
4116 899X/198/1/012013)
- 4117 Baum, W. A. (1962). Photoelectric magnitudes and red-shifts. *Proceedings from the In-
4118 ternational Astronomical Union Symposium 15: Problems of Extra-Galactic Research*,
4119 390.
- 4120 Beaklini, P. P. B., Quadros, A. V. C., de Avellar, M. G. B., Dantas, M. L. L., & Cançado,
4121 A. L. F. (2020). AGN dichotomy beyond radio loudness: A Gaussian mixture
4122 model analysis. *MNRAS*, 497, 1463–1474. <https://doi.org/10.1093/mnras/staa2072>

- 4124 Becker, R. H., White, R. L., & Helfand, D. J. (1995). The FIRST survey: Faint images of
 4125 the radio sky at twenty centimeters. *ApJ*, *450*(2), 559. <https://doi.org/10.1086/176166>
- 4127 Bertin, E., & Arnouts, S. (1996). SExtractor: Software for source extraction. *A&AS*,
 4128 *117*(2), 393–404. <https://doi.org/10.1051/aas:1996164>
- 4129 Best, P. N. (2009). Radio source populations: Results from SDSS. *Astronomische
 4130 Nachrichten*, *330*, 184–189. <https://doi.org/10.1002/asna.200811152>
- 4131 Best, P. N., Kauffmann, G., Heckman, T. M., Brinchmann, J., Charlot, S., Ivezić, Ž., &
 4132 White, S. D. M. (2005). The host galaxies of radio-loud active galactic nuclei:
 4133 Mass dependences, gas cooling and active galactic nuclei feedback. *MNRAS*,
 4134 *362*, 25–40. <https://doi.org/10.1111/j.1365-2966.2005.09192.x>
- 4135 Bicknell, G. V., Dopita, M. A., & O'Dea, C. P. O. (1997). Unification of the radio and opti-
 4136 cal properties of gigahertz peak spectrum and compact steep-spectrum radio
 4137 sources. *ApJ*, *485*(1), 112–124. <https://doi.org/10.1086/304400>
- 4138 Bicknell, G. V. (1995). Relativistic Jets and the Fanaroff-Riley classification of radio
 4139 galaxies. *ApJS*, *101*, 29–39. <https://doi.org/10.1086/192232>
- 4140 Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T.,
 4141 Singh, R., Szerlip, P., Horsfall, P., & Goodman, N. D. (2019). Pyro: Deep Uni-
 4142 versal Probabilistic Programming. *Journal of Machine Learning Research*, *20*.
- 4143 Bîrzan, L., McNamara, B. R., Nulsen, P. E. J., Carilli, C. L., & Wise, M. W. (2008). Ra-
 4144 diative efficiency and content of extragalactic radio sources: Toward a univer-
 4145 sal scaling relation between jet power and radio power. *ApJ*, *686*(2), 859–880.
 4146 <https://doi.org/10.1086/591416>
- 4147 Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- 4148 Black, A. R. S., Baum, S. A., Leahy, J. P., Perley, R. A., Riley, J. M., & Scheuer, P. A. G.
 4149 (1992). A study of FRII radio galaxies with z less than 0.15 — I. high-resolution
 4150 maps of eight sources at 3.6 cm. *MNRAS*, *256*, 186–208. <https://doi.org/10.1093/mnras/256.2.186>
- 4152 Blundell, K. M., Rawlings, S., & Willott, C. J. (1999). The nature and evolution of clas-
 4153 sical double radio sources from complete samples. *AJ*, *117*(2), 677–706. <https://doi.org/10.1086/300721>
- 4155 Bonaldi, A., An, T., Bruggen, M., Burkutean, S., Coelho, B., Goodarzi, H., Hartley, P.,
 4156 Sandhu, P. K., Wu, C., Yu, L., Haghghi, M. H. Z., Anton, S., Bagheri, Z., Barbosa,
 4157 D., Barraca, J. P., Bartashevich, D., Bergano, M., Bonato, M., Brand, J., ... Wong,
 4158 O. I. (2020). Square Kilometre Array Science Data Challenge 1: Analysis and
 4159 results. *MNRAS*, *500*(3), 3821–3837. <https://doi.org/10.1093/mnras/staa3023>
- 4160 Bosch-Ramon, V. (2018). The role of AGN jets in the reionization epoch. *A&A*, *617*.
 4161 <https://doi.org/10.1051/0004-6361/201833952>
- 4162 Bowles, M., Scaife, A. M. M., Porter, F., Tang, H., & Bastien, D. J. (2020). Attention-
 4163 gating for improved radio galaxy classification. *MNRAS*, *501*, 4579–4595. <https://doi.org/10.1093/mnras/staa3946>
- 4165 Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. <https://doi.org/10.1023/A:1018054314350>

- 4167 Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- 4168 Brentjens, M. A., & de Bruyn, A. G. (2005). Faraday rotation measure synthesis. *A&A*, 441(3), 1217–1228. <https://doi.org/10.1051/0004-6361:20052990>
- 4171 Brown, S. (2011). *POSSUM report 9: Assess the complexity of an RM synthesis spectrum*.
- 4172 Brown, S., Bergerud, B., Costa, A., Gaensler, B. M., Isbell, J., LaRocca, D., Norris, R.,
4173 Purcell, C., Rudnick, L., & Sun, X. (2018). Classifying complex Faraday spectra
4174 with convolutional neural networks. *MNRAS*, 483, 964–970. <https://doi.org/10.1093/mnras/sty2908>
- 4176 Capetti, A., Brienza, M., Baldi, R. D., Giovannini, G., Morganti, R., Hardcastle, M. J.,
4177 Rottgering, H. J. A., Brunetti, G. F., Best, P. N., & Miley, G. (2020). The LOFAR
4178 view of FR 0 radio galaxies. *A&A*, 642. <https://doi.org/10.1051/0004-6361/202038671>
- 4180 Carballo, R., Cofiño, A. S., & González-Serrano, J. I. (2004). Selection of quasar can-
4181 didates from combined radio and optical surveys using neural networks. *MNRAS*, 353, 211–220. <https://doi.org/10.1111/j.1365-2966.2004.08056.x>
- 4183 Cattaneo, A., Faber, S. M., Binney, J., Dekel, A., Kormendy, J., Mushotzky, R., Babul,
4184 A., Best, P. N., Brüggen, M., Fabian, A. C., Frenk, C. S., Khalatyan, A., Netzer,
4185 H., Mahdavi, A., Silk, J., Steinmetz, M., & Wisotzki, L. (2009). The role of black
4186 holes in galaxy formation and evolution. *Nature*, 460(7252), 213–219. <https://doi.org/10.1038/nature08135>
- 4188 Cavagnolo, K. W., McNamara, B. R., Nulsen, P. E. J., Carilli, C. L., Jones, C., & Bîrzan, L.
4189 (2010). A relationship between AGN jet power and radio power. *ApJ*, 720(2),
4190 1066–1072. <https://doi.org/10.1088/0004-637X/720/2/1066>
- 4191 Cheung, C. C. (2007). FIRST “winged” and X-shaped radio source candidates. *AJ*,
4192 133(5), 2097–2121. <https://doi.org/10.1086/513095>
- 4193 Chollet, F. et al. (2015). Keras. GitHub. <https://github.com/fchollet/keras>
- 4194 Collier, J. D., Banfield, J. K., Norris, R. P., Schnitzeler, D. H. F. M., Kimball, A. E., Fil-
4195 ipović, M. D., Jarrett, T. H., Lonsdale, C. J., & Tothill, N. F. H. (2014). Infrared-
4196 faint radio sources: A new population of high-redshift radio galaxies. *MNRAS*,
4197 439, 545–565. <https://doi.org/10.1093/mnras/stt2485>
- 4198 Condon, J. J. (1991). Radio luminosity functions (N. Duric & P. C. Crane, Eds.). In N.
4199 Duric & P. C. Crane (Eds.), *The interpretation of modern synthesis observations of*
4200 *spiral galaxies*.
- 4201 Condon, J. J. (1992). Radio emission from normal galaxies. *ARA&A*, 30, 575–611. <https://doi.org/10.1146/annurev.aa.30.090192.003043>
- 4203 Condon, J. J., Cotton, W. D., & Broderick, J. J. (2002). Radio sources and star formation
4204 in the local universe. *AJ*, 124(2), 675–689. <https://doi.org/10.1086/341650>
- 4205 Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., &
4206 Broderick, J. J. (1998). The NRAO VLA Sky Survey. *AJ*, 115(5), 1693–1716. <https://doi.org/10.1086/300337>
- 4208 Condon, J. J., Matthews, A. M., & Broderick, J. J. (2019). Radio sources in the nearby
4209 universe. *ApJ*, 872(2). <https://doi.org/10.3847/1538-4357/ab0301>

- 4210 Condon, J. J., & Ransom, S. M. (2016). *Essential radio astronomy*. Princeton University
4211 Press.
- 4212 Connor, L., & van Leeuwen, J. (2018). Applying deep learning to fast radio burst clas-
4213 sification. AJ, 156(6). <https://doi.org/10.3847/1538-3881/aae649>
- 4214 Contigiani, O., de Gasperin, F., Miley, G. K., Rudnick, L., Andernach, H., Banfield, J. K.,
4215 Kapińska, A. D., Shabala, S. S., & Wong, O. I. (2017). Radio Galaxy Zoo: Cos-
4216 mological alignment of radio sources. MNRAS, 472, 636–646. <https://doi.org/10.1093/mnras/stx1977>
- 4217 Cutri, R., Wright, E., Conrow, T., Fowler, J., Eisenhardt, P., Grillmair, C., Kirkpatrick, J.,
4218 Masci, F., McCallon, H., Wheelock, S., et al. (2013). Explanatory supplement to
4219 the allwise data release products. *Explanatory Supplement to the AllWISE Data*
4220 *Release Products*.
- 4221 Dabhade, P., Mahato, M., Bagchi, J., Saikia, D. J., Combes, F., Sankhyayan, S., Röttgering,
4222 H. J. A., Ho, L. C., Gaikwad, M., Raychaudhury, S., Vaidya, B., & Guiderdoni,
4223 B. (2020). Search and analysis of giant radio galaxies with associated nuclei
4224 (SAGAN). I. new sample and multi-wavelength studies. A&A, 642. <https://doi.org/10.1051/0004-6361/202038344>
- 4225 Dabhade, P., Gaikwad, M., Bagchi, J., Pandey-Pommier, M., Sankhyayan, S., & Ray-
4226 chaudhury, S. (2017). Discovery of giant radio galaxies from NVSS: Radio and
4227 infrared properties. MNRAS, 469, 2886–2906. <https://doi.org/10.1093/mnras/stx860>
- 4228 Dawid, A. P., & Skene, A. M. (1979). Maximum likelihood estimation of observer error-
4229 rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C (Ap-
4230 plied Statistics)*, 28(1), 20–28.
- 4231 Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*
4232 (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781108679930>
- 4233 Diamond, P. (2017). SKA community briefing. Retrieved February 1, 2021, from <https://www.skatelescope.org/ska-community-briefing-18jan2017/>
- 4234 Dieleman, S., Willett, K. W., & Dambre, J. (2015). Rotation-invariant convolutional neu-
4235 ral networks for galaxy morphology prediction. MNRAS, 450, 1441–1459. <https://doi.org/10.1093/mnras/stv632>
- 4236 Fan, D., Budavári, T., Norris, R. P., & Basu, A. (2020). Optimal probabilistic catalogue
4237 matching for radio sources. MNRAS, 498, 565–573. <https://doi.org/10.1093/mnras/staa2447>
- 4238 Fan, D., Budavári, T., Norris, R. P., & Hopkins, A. M. (2015). Matching radio catalogues
4239 with realistic geometry: Application to SWIRE and ATLAS. MNRAS, 451(2),
4240 1299–1305. <https://doi.org/10.1093/mnras/stv994>
- 4241 Fanaroff, B. L., & Riley, J. M. (1974). The morphology of extragalactic radio sources
4242 of high and low luminosity. MNRAS, 167, 31P–36P. <https://doi.org/10.1093/mnras/167.1.31P>
- 4243 Farnes, J. S., Gaensler, B. M., & Carretti, E. (2014). A broadband polarization catalog
4244 of extragalactic radio sources. ApJS, 212(1). <https://doi.org/10.1088/0067-0049/212/1/15>

- 4253 Flamary, R., & Courty, N. (2017). POT: Python optimal transport library. <https://github.com/rflamary/POT>
- 4254
- 4255 Franzen, T. M. O., Banfield, J. K., Hales, C. A., Hopkins, A., Norris, R. P., Seymour, N.,
4256 Chow, K. E., Herzog, A., Huynh, M. T., Lenc, E., Mao, M. Y., & Middelberg, E.
4257 (2015). ATLAS - I. third release of 1.4 GHz mosaics and component catalogues.
4258 *MNRAS*, 453, 4020–4036. <https://doi.org/10.1093/mnras/stv1866>
- 4259 Galvin, T. J., Huynh, M., Norris, R. P., Wang, X. R., Hopkins, E., Wong, O. I., Sha-
4260 bala, S., Rudnick, L., Alger, M. J., & Polsterer, K. L. (2019). Radio Galaxy Zoo:
4261 Knowledge transfer using rotationally invariant self-organizing maps. *PASP*,
4262 131(1004). <https://doi.org/10.1088/1538-3873/ab150b>
- 4263 Galvin, T. J., Huynh, M. T., Norris, R. P., Wang, X. R., Hopkins, E., Polsterer, K., Ralph,
4264 N. O., O'Brien, A. N., & Heald, G. H. (2020). Cataloguing the radio-sky with
4265 unsupervised machine learning: A new approach for the SKA era. *MNRAS*,
4266 497, 2730–2758. <https://doi.org/10.1093/mnras/staa1890>
- 4267 Garofalo, D., & Singh, C. B. (2019). FR0 radio galaxies and their place in the radio
4268 morphology classification. *ApJ*, 871(2). <https://doi.org/10.3847/1538-4357/aaf056>
- 4269
- 4270 Garon, A. F., Rudnick, L., Wong, O. I., Jones, T. W., Kim, J.-A., Andernach, H., Sha-
4271 bala, S. S., Kapińska, A. D., Norris, R. P., de Gasperin, F., Tate, J., & Tang, H.
4272 (2019). Radio Galaxy Zoo: The distortion of radio galaxies by galaxy clusters.
4273 *AJ*, 157(3). <https://doi.org/10.3847/1538-3881/aaff62>
- 4274 Gebhard, T. D., Kilbertus, N., Harry, I., & Schölkopf, B. (2019). Convolutional neural
4275 networks: A magic bullet for gravitational-wave detection? *Phys. Rev. D*, 100.
4276 <https://doi.org/10.1103/PhysRevD.100.063015>
- 4277 Gendre, M. A., & Wall, J. V. (2008). The Combined NVSS-FIRST Galaxies (CoNFIG)
4278 sample — I. sample definition, classification and evolution. *MNRAS*, 390, 819–
4279 828. <https://doi.org/10.1111/j.1365-2966.2008.13792.x>
- 4280 Gilyazev, R. A., & Turdakov, D. Y. (2018). Active learning and crowdsourcing: A survey
4281 of optimization methods for data labeling. *Programming and Computer Software*,
4282 44(6), 476–491. <https://doi.org/10.1134/S0361768818060142>
- 4283 Godfrey, L. E. H., & Shabala, S. S. (2016). Mutual distance dependence drives the ob-
4284 served jet-power-radio-luminosity scaling relations in radio galaxies. *MNRAS*,
4285 456, 1172–1184. <https://doi.org/10.1093/mnras/stv2712>
- 4286 Goldstein, J., S. J., & Reed, J. A. (1984). Double faraday rotation toward 3C 27. *ApJ*, 283,
4287 540–545. <https://doi.org/10.1086/162337>
- 4288 Gopal-Krishna, & Wiita, P. J. (2000). Extragalactic radio sources with hybrid morphol-
4289 ogy: Implications for the Fanaroff-Riley dichotomy. *A&A*, 363, 507–516.
- 4290 Grant, J. K., Taylor, A. R., Stil, J. M., Landecker, T. L., Kothes, R., Ransom, R. R., &
4291 Scott, D. (2010). The DRAO planck deep fields: The polarization properties of
4292 radio galaxies at 1.4 GHz. *ApJ*, 714, 1689–1701. [https://doi.org/10.1088/0004-637X/714/2/1689](https://doi.org/10.1088/0004-
4293 637X/714/2/1689)
- 4294 Grant, J. K. (2011). *Polarised radio sources: A study of luminosity, redshift and flux density*
4295 (Doctoral dissertation). University of Calgary.

- 4296 Groves, B., & Kewley, L. (2007). Distinguishing active galactic nuclei and star forma-
 4297 tion. *Pathways Through an Eclectic Universe*.
- 4298 Guo, P., Duan, F., Wang, P., Yao, Y., Yin, Q., Xin, X., Li, D., Qian, L., Wang, S., Pan, Z.,
 4299 & Zhang, L. (2019). Pulsar candidate classification using generative adversary
 4300 networks. MNRAS, 490(4), 5424–5439. <https://doi.org/10.1093/mnras/stz2975>
- 4301 Hardcastle, M. J., & Croston, J. H. (2020). Radio galaxies and feedback from agn jets.
 4302 *New Astronomy Reviews*, 88. <https://doi.org/10.1016/j.newar.2020.101539>
- 4303 Hardcastle, M. J., & Krause, M. G. H. (2013). Numerical modelling of the lobes of radio
 4304 galaxies in cluster environments. MNRAS, 430, 174–196. <https://doi.org/10.1093/mnras/sts564>
- 4305 Hardcastle, M. J., & Krause, M. G. H. (2014). Numerical modelling of the lobes of radio
 4306 galaxies in cluster environments — II. magnetic field configuration and observ-
 4307 ability. MNRAS, 443, 1482–1499. <https://doi.org/10.1093/mnras/stu1229>
- 4308 Hardcastle, M. J., Williams, W. L., Best, P. N., Croston, J. H., Duncan, K. J., Röttgering,
 4309 H. J. A., Sabater, J., Shimwell, T. W., Tasse, C., Callingham, J. R., & et al. (2019).
 4310 Radio-loud AGN in the first LoTSS data release: The lifetimes and environmen-
 4311 tal impact of jet-driven sources. A&A, 622. <https://doi.org/10.1051/0004-6361/201833893>
- 4312 He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recogni-
 4313 tion, In *Proceedings of the ieee conference on computer vision and pattern recognition*.
- 4314 Heald, G. (2008). The Faraday rotation measure synthesis technique. *Proceedings of the
 4315 International Astronomical Union*, 4(S259), 591–602. <https://doi.org/10.1017/S1743921309031421>
- 4316 Heckman, T. M., & Best, P. N. (2014). The coevolution of galaxies and supermassive
 4317 black holes: Insights from surveys of the contemporary universe. ARA&A, 52(1),
 4318 589–660. <https://doi.org/10.1146/annurev-astro-081913-035722>
- 4319 Helfand, D. J., White, R. L., & Becker, R. H. (2015). The last of FIRST: The final catalog
 4320 and source identifications. ApJ, 801(1). <https://doi.org/10.1088/0004-637x/801/1/26>
- 4321 Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolta,
 4322 M. R., Halpern, M., Hill, R. S., Odegard, N., Page, L., Smith, K. M., Weiland, J. L.,
 4323 Gold, B., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., ... Wright,
 4324 E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) ob-
 4325 servations: Cosmological parameter results. ApJS, 208. <https://doi.org/10.1088/0067-0049/208/2/19>
- 4326 Hložek, R., Ponder, K. A., Malz, A. I., Dai, M., Narayan, G., Ishida, E. E. O., Allam, J., T.,
 4327 Bahmanyar, A., Biswas, R., Galbany, L., & et al. (2020). Results of the photometric
 4328 LSST astronomical time-series classification challenge (PLAsTiCC). *arXiv e-prints*, arXiv:2012.12392v1 [astro-ph.IM].
- 4329 Holzinger, A. (2016). Interactive machine learning for health informatics: When do we
 4330 need the human-in-the-loop? *Brain Informatics*, 3(2), 119–131.
- 4331 Hurley-Walker, N., Callingham, J. R., Hancock, P. J., Franzen, T. M. O., Hindson, L.,
 4332 Kapińska, A. D., Morgan, J., Offringa, A. R., Wayth, R. B., Wu, C., Zheng, Q.,
 4333 Murphy, T., Bell, M. E., Dwarakanath, K. S., For, B., Gaensler, B. M., Johnston-
 4334

- 4340 Hollitt, M., Lenc, E., Procopio, P., ... Williams, C. L. (2017). GaLactic and Ex-
 4341 tragalactic All-sky Murchison Widefield Array (GLEAM) survey - I. a low-
 4342 frequency extragalactic catalogue. *MNRAS*, *464*, 1146–1167. <https://doi.org/10.1093/mnras/stw2337>
- 4344 Husemann, B., & Harrison, C. M. (2018). Reality and myths of AGN feedback. *Nature Astronomy*, *2*(3), 196–197. <https://doi.org/10.1038/s41550-018-0407-2>
- 4345 Jarrett, T. H., Cluver, M. E., Magoulas, C., Bilicki, M., Alpaslan, M., Bland-Hawthorn,
 4346 J., Brough, S., Brown, M. J. I., Croom, S., Driver, S., Holwerda, B. W., Hopkins,
 4347 A. M., Loveday, J., Norberg, P., Peacock, J. A., Popescu, C. C., Sadler, E. M., Tay-
 4348 lor, E. N., Tuffs, R. J., & Wang, L. (2017). Galaxy and mass assembly (GAMA):
 4349 Exploring the WISE web in G12. *ApJ*, *836*(2).
- 4350 Jarrett, T. H., Cohen, M., Masci, F., Wright, E., Stern, D., Benford, D., Blain, A., Carey, S.,
 4351 Cutri, R. M., Eisenhardt, P., Lonsdale, C., Mainzer, A., Marsh, K., Padgett, D.,
 4352 Petty, S., Ressler, M., Skrutskie, M., Stanford, S., Surace, J., ... Yan, D. L. (2011).
 4353 The Spitzer-WISE survey of the ecliptic poles. *ApJ*, *735*(2).
- 4354 Johnston, S., Bailes, M., Bartel, N., Baugh, C., Bietenholz, M., Blake, C., Braun, R., Brown,
 4355 J., Chatterjee, S., Darling, J., Deller, A., Dodson, R., Edwards, P. G., Ekers, R.,
 4356 Ellingsen, S., Feain, I., Gaensler, B. M., Havercorn, M., Hobbs, G., ... Wolleben,
 4357 M. (2007). Science with the Australian Square Kilometre Array Pathfinder. *PASA*,
 4358 *24*, 174–188. <https://doi.org/10.1071/AS07033>
- 4359 Kapinska, A. D. (2020). EMU: The Evolutionary Map of the Universe radio sky survey.
 4360 *American Astronomical Society Meeting Abstracts*, *236*.
- 4361 Kapińska, A. D., Terentev, I., Wong, O. I., Shabala, S. S., Andernach, H., Rudnick, L.,
 4362 Storer, L., Banfield, J. K., Willett, K. W., Gasperin, F. d., Lintott, C. J., López-
 4363 Sánchez, Á. R., Middelberg, E., Norris, R. P., Schawinski, K., Seymour, N., &
 4364 Simmons, B. (2017). Radio Galaxy Zoo: A search for hybrid morphology radio
 4365 galaxies. *AJ*, *154*(6), 253. <https://doi.org/10.3847/1538-3881/aa90b7>
- 4366 Kimball, A. E., & Ivezić, Ž. (2008). A unified catalog of radio objects detected by NVSS,
 4367 FIRST, WENSS, GB6, and SDSS. *AJ*, *136*, 684–712. [https://doi.org/10.1088/0004-6256/136/2/684](https://doi.org/10.1088/0004-

 4368 6256/136/2/684)
- 4369 Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *3rd Interna-
 4370 tional Conference for Learning Representations*, arXiv:1412.6980v9 [cs.LG].
- 4371 Kochanek, C. S., Pahre, M. A., Falco, E. E., Huchra, J. P., Mader, J., Jarrett, T. H., Chester,
 4372 T., Cutri, R., & Schneider, S. E. (2001). The K-band galaxy luminosity function.
 4373 *ApJ*, *560*, 566–579. <https://doi.org/10.1086/322488>
- 4374 Koribalski, B. S., Staveley-Smith, L., Westmeier, T., Serra, P., Spekkens, K., Wong, O. I.,
 4375 Lee-Waddell, K., Lagos, C. D. P., Obreschkow, D., Ryan-Weber, E. V., Zwaan,
 4376 M., Kilborn, V., Bekiaris, G., Bekki, K., Bigiel, F., Boselli, A., Bosma, A., Catinella,
 4377 B., Chauhan, G., ... Wolf, C. (2020). WALLABY - an SKA pathfinder HI survey.
 4378 *Ap&SS*, *365*. <https://doi.org/10.1007/s10509-020-03831-4>
- 4379 Kormendy, J., & Ho, L. C. (2013). Coevolution (or not) of supermassive black holes
 4380 and host galaxies. *ARA&A*, *51*(1), 511–653.

- 4382 Kormendy, J., & Richstone, D. (1995). Inward bound—the search for supermassive
 4383 black holes in galactic nuclei. *ARA&A*, 33. <https://doi.org/10.1146/annurev.aa.33.090195.003053>
- 4385 Kozieł-Wierzbowska, D., Vale Asari, N., Stasińska, G., Herpich, F. R., Sikora, M., Ży-
 4386 wucka, N., & Goyal, A. (2020). Identifying radio active galactic nuclei among
 4387 radio-emitting galaxies. *arXiv e-prints*, arXiv:2012.00367v1 [astro-ph.GA].
- 4388 Lacy, M., Baum, S. A., Chandler, C. J., Chatterjee, S., Clarke, T. E., Deustua, S., English,
 4389 J., Farnes, J., Gaensler, B. M., Gugliucci, N., Hallinan, G., Kent, B. R., Kimball, A.,
 4390 Law, C. J., Lazio, T. J. W., Marvil, J., Mao, S. A., Medlin, D., Mooley, K., ... Yoon,
 4391 I. (2020). The Karl G. Jansky Very Large Array Sky Survey (VLASS). science
 4392 case and survey design. *PASP*, 132(1009). <https://doi.org/10.1088/1538-3873/ab63eb>
- 4394 Laing, R. A., & Bridle, A. H. (1987). Rotation measure variation across M84. *MNRAS*,
 4395 228, 557–571. <https://doi.org/10.1093/mnras/228.3.557>
- 4396 Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predic-
 4397 tive uncertainty estimation using deep ensembles, In *Advances in neural infor-
 4398 mation processing systems*.
- 4399 Law, C. J., Gaensler, B. M., Bower, G. C., Backer, D. C., Bauermeister, A., Croft, S.,
 4400 Forster, R., Gutierrez-Kraybill, C., Harvey-Smith, L., Heiles, C., Hull, C., Keat-
 4401 ing, G., MacMahon, D., Whysong, D., Williams, P. K. G., & Wright, M. (2011).
 4402 Spectropolarimetry with the Allen Telescope Array: Faraday rotation toward
 4403 bright polarized radio galaxies. *ApJ*, 728(1). <https://doi.org/10.1088/0004-637X/728/1/57>
- 4405 Leahy, J. P., Bridle, A. H., & Strom, R. G. (n.d.). An Atlas of DRAGNs. Retrieved June
 4406 13, 2020, from <http://www.jb.man.ac.uk/atlas/index.html>
- 4407 Leahy, J. P., & Perley, R. A. (1991). VLA images of 23 extragalactic radio sources. *AJ*,
 4408 102, 537–561. <https://doi.org/10.1086/115892>
- 4409 Leahy, J. P., Pooley, G. G., & Riley, J. M. (1986). The polarization of classical double-
 4410 radio sources. *MNRAS*, 222, 753–785. <https://doi.org/10.1093/mnras/222.4.753>
- 4411 LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel,
 4412 L. D. (1989). Backpropagation applied to handwritten zip code recognition.
Neural Computation, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- 4414 LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied
 4415 to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- 4416 Ledlow, M. J., & Owen, F. N. (1996). 20 cm VLA survey of Abell clusters of galaxies.
 4417 VI. radio/optical luminosity functions. *AJ*, 112. <https://doi.org/10.1086/117985>
- 4418 Lin, H., Li, X., & Zeng, Q. (2020). Pulsar candidate sifting using multi-input convolu-
 4419 tion neural networks. *ApJ*, 899. <https://doi.org/10.3847/1538-4357/aba838>
- 4420 Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick,
 4421 M. J., Nichol, R. C., Szalay, A., Andreeescu, D., Murray, P., & Vandenberg, J.
 4422 (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies
 4423 from the Sloan Digital Sky Survey. *MNRAS*, 389, 1179–1189. <https://doi.org/10.1111/j.1365-2966.2008.13689.x>

- 4425 Livingston, J. D., McClure-Griffiths, N. M., Gaensler, B. M., Seta, A., & Alger, M. J.
4426 (2021). Heightened Faraday complexity in the inner 1 kpc of the Galactic cen-
4427 tre. MNRAS. <https://doi.org/10.1093/mnras/stab253>
- 4428 Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., Surace, J., Shupe, D., Xu, C., Oliver,
4429 S., Padgett, D., Fang, F., Conrow, T., Franceschini, A., Gautier, N., Griffin, M.,
4430 Hacking, P., Masci, F., Morrison, G., O'Linger, J., Owen, F., Pérez-Fournon, I., ...
4431 Serjeant, S. (2003). SWIRE: The SIRTF Wide-Area Infrared Extragalactic Survey.
4432 PASP, 115, 897–927. <https://doi.org/10.1086/376850>
- 4433 Lovelace, R. V. E. (1976). Dynamo model of double radio sources. *Nature*, 262(5570),
4434 649–652. <https://doi.org/10.1038/262649a0>
- 4435 Lukic, V., Brüggen, M., Banfield, J. K., Wong, O. I., Rudnick, L., Norris, R. P., & Simmons,
4436 B. (2018). Radio Galaxy Zoo: Compact and extended radio source classification
4437 with deep learning. MNRAS, 476, 246–260. <https://doi.org/10.1093/mnras/sty163>
- 4438 Lukic, V., Brüggen, M., Mingo, B., Croston, J. H., Kasieczka, G., & Best, P. N. (2019).
4439 Morphological classification of radio galaxies: Capsule networks versus convo-
4440 lutional neural networks. MNRAS, 487, 1729–1744. <https://doi.org/10.1093/mnras/stz1289>
- 4441 Ma, Y. K., Mao, S. A., Stil, J., Basu, A., West, J., Heiles, C., Hill, A. S., & Betti, S. K.
4442 (2019a). A broad-band spectro-polarimetric view of the NVSS rotation mea-
4443 sure catalogue - I. breaking the $n\pi$ -ambiguity. MNRAS, 487(3), 3432–3453. <https://doi.org/10.1093/mnras/stz1325>
- 4444 Ma, Z., Xu, H., Zhu, J., Hu, D., Li, W., Shan, C., Zhu, Z., Gu, L., Li, J., Liu, C., & Wu,
4445 X. (2019b). A machine learning based morphological classification of 14,245
4446 radio AGNs selected from the Best-Heckman sample. ApJS, 240(2), 34. <https://doi.org/10.3847/1538-4365/aaf9a2>
- 4447 Ma, Z., Zhu, J., Li, W., & Xu, H. (2018). Radio galaxy morphology generation using
4448 DNN autoencoder and Gaussian mixture models. *arXiv e-prints*, 1806, arXiv:
4449 1806.00398v1 [cs.CV].
- 4450 Machado Poletti Valle, L. F., Avestruz, C., Barnes, D. J., Farahi, A., Lau, E. T., & Nagai,
4451 D. (2020). SHAPing the gas: Understanding gas shapes in dark matter haloes
4452 with interpretable machine learning. *arXiv e-prints*, 2011, arXiv:2011.12987v1
4453 [astro-ph.CO].
- 4454 Machalski, J., Jamrozy, M., Stawarz, L., & Koziel-Wierzbowska, D. (2011). Under-
4455 standing giant radio galaxy J1420-0545: Large-scale morphology, environment, and
4456 energetics. ApJ, 740(2). <https://doi.org/10.1088/0004-637X/740/2/58>
- 4457 Macquart, J.-P., Bailes, M., Bhat, N. D. R., Bower, G. C., Bunton, J. D., Chatterjee, S.,
4458 Colegate, T., Cordes, J. M., D'Addario, L., Deller, A., & et al. (2010). The Com-
4459 mensal Real-Time ASKAP Fast-Transients (CRAFT) survey. PASA, 27(3), 272–
4460 282. <https://doi.org/10.1071/AS09082>
- 4461 Marconi, A., & Hunt, L. K. (2003). The relation between black hole mass, bulge mass,
4462 and near-infrared luminosity. ApJ, 589(1), L21–L24. <https://doi.org/10.1086/375804>

- 4468 Marshall, P. J., Lintott, C. J., & Fletcher, L. N. (2015). Ideas for citizen science in astronomy. *ARA&A*, 53, 247–278. <https://doi.org/10.1146/annurev-astro-081913-035959>
- 4471 Mauch, T., & Sadler, E. M. (2007). Radio sources in the 6dFGS: Local luminosity functions at 1.4GHz for star-forming galaxies and radio-loud AGN. *MNRAS*, 375, 931–950. <https://doi.org/10.1111/j.1365-2966.2006.11353.x>
- 4474 McConnell, D., Hale, C. L., Lenc, E., Banfield, J. K., Heald, G., Hotan, A. W., Leung, J. K., Moss, V. A., Murphy, T., O'Brien, A., Pritchard, J., Raja, W., Sadler, E. M., Stewart, A., Thomson, A. J. M., Whiting, M., Allison, J. R., Amy, S. W., Anderson, C., ... Westmeier, T. (2020). The Rapid ASKAP Continuum Survey I: Design and first results. *PASA*, 37. <https://doi.org/10.1017/pasa.2020.41>
- 4479 Menon, A. K., Van Rooyen, B., Ong, C. S., & Williamson, R. C. (2015). Learning from corrupted binary labels via class-probability estimation, In *Proceedings of the 32nd International Conference on International Conference on Machine Learning — Volume 37*.
- 4483 Middelberg, E., Norris, R. P., Cornwell, T. J., Voronkov, M. A., Siana, B. D., Boyle, B. J., Ciliegi, P., Jackson, C. A., Huynh, M. T., Berta, S., Rubele, S., Lonsdale, C. J., Ivison, R. J., & Smail, I. (2008). Deep Australia Telescope large area survey radio observations of the European large area ISO survey S1/Spitzer wide-area infrared extragalactic field. *AJ*, 135, 1276–1290. <https://doi.org/10.1088/0004-6256/135/4/1276>
- 4489 Miller, D. (2017). Stargazing Live viewers find four-planet solar system. Retrieved January 16, 2021, from <https://www.abc.net.au/news/2017-04-06/stargazing-live-four-planets-discovered-in-new-solar-system/8423142>
- 4492 Mingo, B., Croston, J. H., Hardcastle, M. J., Best, P. N., Duncan, K. J., Morganti, R., Röttgering, H. J. A., Sabater, J., Shimwell, T. W., Williams, W. L., Brienza, M., Gurkan, G., Mahatma, V. H., Morabito, L. K., Prandoni, I., Bondi, M., Ineson, J., & Mooney, S. (2019). Revisiting the Fanaroff–Riley dichotomy and radio-galaxy morphology with the LOFAR Two-Metre Sky Survey (LoTSS). *MNRAS*, 488(2), 2701–2721. <https://doi.org/10.1093/mnras/stz1901>
- 4498 Miyashita, Y., Ideguchi, S., Nakagawa, S., Akahori, T., & Takahashi, K. (2019). Performance test of QU-fitting in cosmic magnetism study. *MNRAS*, 482(2), 2739–2749. <https://doi.org/10.1093/mnras/sty2862>
- 4501 Mohr, P. J., Newell, D. B., Barry N. Taylor, & Tiesinga, E. (2019). The NIST reference on constants, units, and uncertainty. NIST Physical Measurement Laboratory. <https://physics.nist.gov/cuu/Reference/about.html>
- 4504 Morganti, R. (2017). The many routes to AGN feedback. *Frontiers in Astronomy and Space Sciences*, 4. <https://doi.org/10.3389/fspas.2017.00042>
- 4506 Mostert, R. I. J., Duncan, K. J., Röttgering, H. J. A., Polsterer, K. L., Best, P. N., Brienza, M., Brüggen, M., Hardcastle, M. J., Jurlin, N., Mingo, B., Morganti, R., Shimwell, T., Smith, D., & Williams, W. L. (2021). Unveiling the rarest morphologies of the LOFAR Two-metre Sky Survey radio source population with self-organised maps. *A&A*, 645. <https://doi.org/10.1051/0004-6361/202038500>

- 4511 Mukherjee, D., Bicknell, G. V., Sutherland, R., & Wagner, A. (2016). Relativistic jet
4512 feedback in high-redshift galaxies - I. dynamics. MNRAS, 461, 967–983. <https://doi.org/10.1093/mnras/stw1368>
- 4514 Mukherjee, D., Bicknell, G. V., Wagner, A. Y., Sutherland, R. S., & Silk, J. (2018). Rela-
4515 tivistic jet feedback - III. feedback on gas discs. MNRAS, 479, 5544–5566. <https://doi.org/10.1093/mnras/sty1776>
- 4517 Murphy, T., Chatterjee, S., Kaplan, D. L., Banyer, J., Bell, M. E., Bignall, H. E., Bower,
4518 G. C., Cameron, R. A., Coward, D. M., Cordes, J. M., Croft, S., Curran, J. R.,
4519 Djorgovski, S. G., Farrell, S. A., Frail, D. A., Gaensler, B. M., Galloway, D. K.,
4520 Gendre, B., Green, A. J., ... Williams, P. K. G. (2013). VAST: An ASKAP survey
4521 for variables and slow transients. PASA, 30. <https://doi.org/10.1017/pasa.2012.006>
- 4523 Norris, R. P. (2017a). Discovering the unexpected in astronomical survey data. PASA, 34. <https://doi.org/10.1017/pasa.2016.63>
- 4525 Norris, R. P. (2017b). Extragalactic radio continuum surveys and the transformation of
4526 radio astronomy. Nature Astronomy, 1, 671–678. <https://doi.org/10.1038/s41550-017-0233-y>
- 4528 Norris, R. P., Afonso, J., Appleton, P. N., Boyle, B. J., Ciliegi, P., Croom, S. M., Huynh,
4529 M. T., Jackson, C. A., Koekemoer, A. M., Lonsdale, C. J., Middelberg, E., Mobasher,
4530 B., Oliver, S. J., Polletta, M., Siana, B. D., Smail, I., & Voronkov, M. A. (2006).
4531 Deep ATLAS radio observations of the Chandra Deep Field-South/Spitzer wide-
4532 area infrared extragalactic field. AJ, 132(6), 2409–2423. <https://doi.org/10.1086/508275>
- 4534 Norris, R. P., Hopkins, A. M., Afonso, J., Brown, S., Condon, J. J., Dunne, L., Feain, I.,
4535 Hollow, R., Jarvis, M., Johnston-Hollitt, M., Lenc, E., Middelberg, E., Padovani,
4536 P., Prandoni, I., Rudnick, L., Seymour, N., Umana, G., Andernach, H., Alexan-
4537 der, D. M., ... Zhao, G.-B. (2011). EMU: Evolutionary Map of the Universe.
4538 PASA, 28, 215–248. <https://doi.org/10.1071/AS11021>
- 4539 O'Sullivan, S. P., Brown, S., Robishaw, T., Schnitzeler, D. H. F. M., McClure-Griffiths,
4540 N. M., Feain, I. J., Taylor, A. R., Gaensler, B. M., Land ecker, T. L., Harvey-Smith,
4541 L., & Carretti, E. (2012). Complex Faraday depth structure of active galactic nu-
4542 clei as revealed by broad-band radio polarimetry. MNRAS, 421(4), 3300–3315.
4543 <https://doi.org/10.1111/j.1365-2966.2012.20554.x>
- 4544 O'Sullivan, S. P., Gaensler, B. M., Lara-López, M. A., Velzen, S. v., Banfield, J. K., &
4545 Farnes, J. S. (2015). The magnetic field and polarization properties of radio
4546 galaxies in different accretion states. ApJ, 806(1). <https://doi.org/10.1088/0004-637X/806/1/83>
- 4548 O'Sullivan, S. P., Purcell, C. R., Anderson, C. S., Farnes, J. S., Sun, X. H., & Gaensler,
4549 B. M. (2017). Broad-band, radio spectro-polarimetric study of 100 radiative-
4550 mode and jet-mode AGN. MNRAS, 469. <https://doi.org/10.1093/mnras/stx1133>
- 4551 Owens, E. A., Griffiths, R. E., & Ratnatunga, K. U. (1996). Using oblique decision trees
4552 for the morphological classification of galaxies. MNRAS, 281. <https://doi.org/10.1093/mnras/281.1.153>

- 4554 Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge
4555 and Data Engineering*, 22(10), 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- 4557 Panwar, M., Prabhakar, Sandhu, P. K., Wadadekar, Y., & Jain, P. (2020). Alignment of
4558 radio galaxy axes using FIRST catalogue. *MNRAS*. <https://doi.org/10.1093/mnras/staa2975>
- 4560 Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison,
4561 A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch, In *NIPS
4562 2017 Workshop Autodiff*.
- 4563 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel,
4564 M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau,
4565 D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine
4566 learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- 4567 Polsterer, K. L., Gieseke, F., & Igel, C. (2015). Automatic galaxy classification via
4568 machine learning techniques: Parallelized rotation/flipping invariant kohonen maps
4569 (PINK), In *Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV)*.
- 4570 Pracy, M. B., Ching, J. H. Y., Sadler, E. M., Croom, S. M., Baldry, I. K., Bland-Hawthorn,
4571 J., Brough, S., Brown, M. J. I., Couch, W. J., Davis, T. M., Drinkwater, M. J.,
4572 Hopkins, A. M., Jarvis, M. J., Jelliffe, B., Jurek, R. J., Loveday, J., Pimbblet, K. A.,
4573 Prescott, M., Wisnioski, E., & Woods, D. (2016). GAMA/WiggleZ: The 1.4 GHz
4574 radio luminosity functions of high- and low-excitation radio galaxies and their
4575 redshift evolution to $z = 0.75$. *MNRAS*, 460, 2–17. <https://doi.org/10.1093/mnras/stw910>
- 4577 Proctor, D. D. (2006). Comparing pattern recognition feature sets for sorting triples in
4578 the FIRST database. *ApJS*, 165, 95–107. <https://doi.org/10.1086/504801>
- 4579 Proctor, D. D. (2011). Morphological annotations for groups in the FIRST database.
4580 *ApJS*, 194. <https://doi.org/10.1088/0067-0049/194/2/31>
- 4581 Proctor, D. D. (2016). A selection of giant radio sources from NVSS. *ApJS*, 224. <https://doi.org/10.3847/0067-0049/224/2/18>
- 4583 Ralph, N. O., Norris, R. P., Fang, G., Park, L. A. F., Galvin, T. J., Alger, M. J., Andernach,
4584 H., Lintott, C., Rudnick, L., Shabala, S., & Wong, O. I. (2019). Radio Galaxy Zoo:
4585 Unsupervised clustering of convolutionally auto-encoded radio-astronomical
4586 images. *PASP*, 131(1004). <https://doi.org/10.1088/1538-3873/ab213d>
- 4587 Randall, K., Hopkins, A., Norris, R., Zinn, P.-C., Middelberg, E., Mao, M., & Sharp, R.
4588 (2012). Spectral index properties of milliJansky radio sources. *MNRAS*, 421(2),
4589 1644–1660. <https://doi.org/10.1111/j.1365-2966.2012.20422.x>
- 4590 Raouf, M., Shabala, S. S., Croton, D. J., Khosroshahi, H. G., & Bernyk, M. (2017). The
4591 many lives of active galactic nuclei — II: The formation and evolution of radio
4592 jets and their impact on galaxy evolution. *MNRAS*, 471, 658–670. <https://doi.org/10.1093/mnras/stx1598>
- 4594 Raykar, V. C., Yu, S., Zhao, L. H., Valadez, G. H., Florin, C., Bogoni, L., & Moy, L. (2010).
4595 Learning from crowds. *Journal of Machine Learning Research*, 11(43), 1297–1322.

- 4596 Rayner, D. P., Norris, R. P., & Sault, R. J. (2000). Radio circular polarization of active
4597 galaxies. *MNRAS*, *319*, 484–496. <https://doi.org/10.1111/j.1365-8711.2000.03854.x>
- 4599 Reiprich, T. H., Veronica, A., Pacaud, F., Ramos-Ceja, M. E., Ota, N., Sanders, J., Kara,
4600 M., & Erben, T. (2020). The Abell 3391/95 galaxy cluster system. A 15 Mpc
4601 intergalactic medium emission filament, a warm gas bridge, infalling matter
4602 clumps, and (re-) accelerated plasma discovered by combining SRG/eROSITA
4603 data with ASKAP/EMU and DECam data. *A&A*. <https://doi.org/10.1051/0004-6361/202039590>
- 4605 Richter, G. A. (1975). Search for optical identifications in the 5C3-radio survey. II —
4606 statistical treatment and results. *Astronomische Nachrichten*, *296*, 65–81. <https://doi.org/10.1002/asna.19752960203>
- 4608 Rodman, P. E., Turner, R. J., Shabala, S. S., Banfield, J. K., Wong, O. I., Andernach,
4609 H., Garon, A. F., Kapińska, A. D., Norris, R. P., & Rudnick, L. (2019). Radio
4610 Galaxy Zoo: Observational evidence for environment as the cause of radio
4611 source asymmetry. *MNRAS*, *482*, 5625–5641. <https://doi.org/10.1093/mnras/sty3070>
- 4613 Rowley, H. A., Baluja, S., & Kanade, T. (1996). Human face detection in visual scenes,
4614 In *Advances in neural information processing systems*.
- 4615 Rybicki, G. B., & Lightman, A. P. (2008). *Radiative processes in astrophysics*. John Wiley
4616 & Sons.
- 4617 Sadler, E. M., Ekers, R. D., Mahony, E. K., Mauch, T., & Murphy, T. (2014). The local
4618 radio-galaxy population at 20 GHz. *MNRAS*, *438*, 796–824. <https://doi.org/10.1093/mnras/stt2239>
- 4620 Sadler, E. M., Jackson, C. A., Cannon, R. D., McIntyre, V. J., Murphy, T., Bland-Hawthorn,
4621 J., Bridges, T., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., de Propris,
4622 R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Lahav,
4623 O., ... Taylor, K. (2002). Radio sources in the 2dF galaxy redshift survey — II.
4624 local radio luminosity functions for AGN and star-forming galaxies at 1.4 GHz.
4625 *MNRAS*, *329*(1), 227–245. <https://doi.org/10.1046/j.1365-8711.2002.04998.x>
- 4626 Saikia, D. J., & Salter, C. J. (1988). Polarization properties of extragalactic radio sources.
4627 *ARA&A*, *26*, 93–144. <https://doi.org/10.1146/annurev.aa.26.090188.000521>
- 4628 Sajina, A., Lacy, M., & Scott, D. (2005). Simulating the Spitzer mid-infrared color-color
4629 diagrams. *ApJ*, *621*, 256–268. <https://doi.org/10.1086/426536>
- 4630 Sakelliou, I., Hardcastle, M. J., & Jetha, N. N. (2008). 3C40 in Abell194: Can tail radio
4631 galaxies exist in a quiescent cluster? *MNRAS*, *384*, 87–93. <https://doi.org/10.1111/j.1365-2966.2007.12465.x>
- 4633 Samudre, A., George, L., Bansal, M., & Wadadekar, Y. (2020). Data-efficient classification
4634 of radio galaxies. *arXiv e-prints*, *2021*, arXiv:2011.13311v1 [astro-ph.IM].
- 4635 Scalzo, R. A., Yuan, F., Childress, M. J., Möller, A., Schmidt, B. P., Tucker, B. E., Zhang,
4636 B. R., Onken, C. A., Wolf, C., Astier, P., Betoule, M., & Regnault, N. (2017). The
4637 SkyMapper Transient Survey. *PASA*, *34*. <https://doi.org/10.1017/pasa.2017.24>
- 4638 Schmidt, M. (1968). Space distribution and luminosity functions of quasi-stellar radio
4639 sources. *ApJ*, *151*. <https://doi.org/10.1086/149446>

- 4640 Segal, G., Parkinson, D., Norris, R. P., & Swan, J. (2019). Identifying complex sources in
 4641 large astronomical data sets using a coarse-grained complexity measure. *PASP*,
 4642 131(1004). <https://doi.org/10.1088/1538-3873/ab0068>
- 4643 Shabala, S. S. (2018). The role of environment in the observed fundamental plane of
 4644 radio active galactic nuclei. *MNRAS*, 478, 5074–5080. <https://doi.org/10.1093/mnras/sty1328>
- 4645 Shabala, S. S., & Godfrey, L. E. H. (2013). Size dependence of the radio-luminosity-
 4646 mechanical-power correlation in radio galaxies. *ApJ*, 769(2). <https://doi.org/10.1088/0004-637X/769/2/129>
- 4647 Shimwell, T. W., Tasse, C., Hardcastle, M. J., Mechev, A. P., Williams, W. L., Best, P. N.,
 4648 Röttgering, H. J. A., Callingham, J. R., Dijkema, T. J., de Gasperin, F., Hoang,
 4649 D. N., Hugo, B., Mirmont, M., Oonk, J. B. R., Prandoni, I., Rafferty, D., Sabater,
 4650 J., Smirnov, O., van Weeren, R. J., ... Wilber, A. (2019). The LOFAR Two-metre
 4651 Sky Survey. II. First data release. *A&A*, 622. <https://doi.org/10.1051/0004-6361/201833559>
- 4652 Sieland, M. (2015). Snapshot Serengeti: An Unexpected Discovery. Retrieved January
 4653 16, 2021, from <https://www.mico-project.eu/snapshot-serengeti-an-unexpected-discovery/>
- 4654 Sikora, M., & Begelman, M. C. (2013). Magnetic flux paradigm for radio loudness of
 4655 active galactic nuclei. *ApJ*, 764. <https://doi.org/10.1088/2041-8205/764/2/L24>
- 4656 Stokes, G. G. (1851). On the composition and resolution of streams of polarized light
 4657 from different sources. *Transactions of the Cambridge Philosophical Society*, 9, 399.
- 4658 Storrie-Lombardi, M. C., Lahav, O., Sodre, L., & Storrie-Lombardi, L. J. (1992). Morphological
 4659 classification of galaxies by artificial neural networks. *MNRAS*, 259.
 4660 <https://doi.org/10.1093/mnras/259.1.8P>
- 4661 Strauss, M. A., Weinberg, D. H., Lupton, R. H., Narayanan, V. K., Annis, J., Bernardi,
 4662 M., Blanton, M., Burles, S., Connolly, A. J., Dalcanton, J., Doi, M., Eisenstein,
 4663 D., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Kent, S., Kim, R. S. J.,
 4664 Knapp, G. R., ... Zehavi, I. (2002). Spectroscopic target selection in the Sloan
 4665 Digital Sky Survey: The main galaxy sample. *AJ*, 124, 1810–1824. <https://doi.org/10.1086/342343>
- 4666 Sun, X. H., Rudnick, L., Akahori, T., Anderson, C. S., Bell, M. R., Bray, J. D., Farnes,
 4667 J. S., Ideguchi, S., Kumazaki, K., O'Brien, T., & et al. (2015). Comparison of algorithms
 4668 for determination of rotation measure and Faraday structure. I. 1100–
 4669 1400 MHz. *AJ*, 149(2), 60. <https://doi.org/10.1088/0004-6256/149/2/60>
- 4670 Surace, J., Shupe, D., Fang, F., Lonsdale, C., Gonzalez-Solares, E., Hatziminaoglou, E.,
 4671 Siana, B., Babbedge, T., Polletta, M., Rodighiero, G., et al. (2005). The SWIRE
 4672 data release 2: Image atlases and source catalogs for ELAIS-N1, ELAIS-N2,
 4673 XMM-LSS, and the Lockman Hole. *Spitzer Science Centre, California Institute of
 4674 Technology, Pasadena, CA*.
- 4675 Tang, H., Scaife, A. M. M., & Leahy, J. P. (2019). Transfer learning for radio galaxy
 4676 classification. *MNRAS*, 488, 3358–3375. <https://doi.org/10.1093/mnras/stz1883>

- 4682 Taylor, A. R., & Jagannathan, P. (2016). Alignments of radio galaxies in deep radio
4683 imaging of ELAIS N1. *MNRAS*, 459(1), L36–L40. <https://doi.org/10.1093/mnrasl/slw038>
- 4685 Taylor, A. R., Stil, J. M., Grant, J. K., Landecker, T. L., Kothes, R., Reid, R. I., Gray, A. D.,
4686 Scott, D., Martin, P. G., Boothroyd, A. I., Joncas, G., Lockman, F. J., English,
4687 J., Sajina, A., & Bond, J. R. (2007). Radio polarimetry of the ELAIS N1 field:
4688 Polarized compact sources. *ApJ*, 666, 201–211. <https://doi.org/10.1086/519786>
- 4689 Taylor, A. R., Stil, J. M., & Sunstrum, C. (2009). A rotation measure image of the sky.
4690 *ApJ*, 702(2), 1230–1236. <https://doi.org/10.1088/0004-637X/702/2/1230>
- 4691 The Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., Lim,
4692 P. L., Crawford, S. M., Conseil, S., Shupe, D. L., Craig, M. W., Dencheva, N.,
4693 Ginsburg, A., VanderPlas, J. T., Bradley, L. D., Pérez-Suárez, D., de Val-Borro,
4694 M., Aldcroft, T. L., Cruz, K. L., Robitaille, T. P., Tollerud, E. J., ... Astropy Con-
4695 tributors. (2018). The Astropy Project: Building an open-science project and
4696 status of the v2.0 core package. *AJ*, 156(3). <https://doi.org/10.3847/1538-3881/aabc4f>
- 4698 Treichel, K., Rudnick, L., Hardcastle, M. J., & Leahy, J. P. (2001). Spectral structure in FR
4699 II radio galaxies and jets. *ApJ*, 561(2), 691–702. <https://doi.org/10.1086/323254>
- 4700 Tucker, B. (2017). Stargazing citizens join search for Planet Nine. Retrieved January 16,
4701 2021, from <https://www.abc.net.au/news/2017-04-04/stargazing-live-citizens-join-search-for-planet-nine/8414832>
- 4703 Tully, R. B., & Fisher, J. R. (1977). A new method of determining distances to galaxies.
4704 *A&A*, 54, 661–673.
- 4705 Turner, R. J., & Shabala, S. S. (2015). Energetics and lifetimes of local radio active galac-
4706 tic nuclei. *ApJ*, 806. <https://doi.org/10.1088/0004-637X/806/1/59>
- 4707 Urry, C. M., & Padovani, P. (1995). Unified schemes for radio-loud active galactic nuclei.
4708 *PASP*, 107(715), 803.
- 4709 Van Eck, C. L., Havercorn, M., Alves, M. I. R., Beck, R., de Bruyn, A. G., Enßlin, T.,
4710 Farnes, J. S., Ferrière, K., Heald, G., Horellou, C., Horneffer, A., Iacobelli, M.,
4711 Jelić, V., Martí-Vidal, I., Mulcahy, D. D., Reich, W., Röttgering, H. J. A., Scaife,
4712 A. M. M., Schnitzeler, D. H. F. M., ... Sridhar, S. S. (2017). Faraday tomography
4713 of the local interstellar medium with LOFAR: Galactic foregrounds towards IC
4714 342. *A&A*, 597. <https://doi.org/10.1051/0004-6361/201629707>
- 4715 van Velzen, S., Falcke, H., Schellart, P., Nierstenhöfer, N., & Kampert, K.-H. (2012).
4716 Radio galaxies of the local universe: All-sky catalog, luminosity functions, and
4717 clustering. *A&A*, 544. <https://doi.org/10.1051/0004-6361/201219389>
- 4718 Verheijen, M. A. W., Oosterloo, T. A., van Cappellen, W. A., Bakker, L., Ivashina, M. V.,
4719 & van der Hulst, J. M. (2008). Apertif, a focal plane array for the WSRT, In
4720 *AIP Conference Series 1035, The Evolution of Galaxies Through the Neutral Hydrogen*
4721 *Window*.
- 4722 Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D.,
4723 Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett,
4724 M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern,
4725 R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algo-

- 4769 - a deep learning classifier for radio morphologies. *MNRAS*, 482, 1211–1230.
4770 <https://doi.org/10.1093/mnras/sty2646>
- 4771 Zevin, M., Coughlin, S., Bahaadini, S., Besler, E., Rohani, N., Allen, S., Cabero, M.,
4772 Crowston, K., Katsaggelos, A. K., Larson, S. L., & et al. (2017). Gravity Spy:
4773 Integrating advanced LIGO detector characterization, machine learning, and
4774 citizen science. *Classical and Quantum Gravity*, 34. <https://doi.org/10.1088/1361-6382/aa5cea>
- 4775 Zhang, C., Wang, C., Hobbs, G., Russell, C. J., Li, D., Zhang, S.-B., Dai, S., Wu, J.-W.,
4776 Pan, Z.-C., Zhu, W.-W., Toomey, L., & Ren, Z.-Y. (2020). Applying saliency-map
4777 analysis in searches for pulsars and fast radio bursts. *A&A*, 642. <https://doi.org/10.1051/0004-6361/201937234>
- 4778 Zhang, J., Wu, X., & Sheng, V. S. (2016). Learning from crowdsourced labeled data: A
4779 survey. *Artificial Intelligence Review*, 46(4), 543–576. <https://doi.org/10.1007/s10462-016-9491-9>
- 4780 Zhang, L. (2020). Transfer adaptation learning: A decade survey. *arXiv e-prints*, arXiv:
4781 1903.04687v2 [cs.CV].
- 4782 Zhu, X.-P., Dai, J.-M., Bian, C.-J., Chen, Y., Chen, S., & Hu, C. (2019). Galaxy morphology
4783 classification with deep convolutional neural networks. *Astrophysics and Space Science*, 364, 55. <https://doi.org/10.1007/s10509-019-3540-1>
- 4784 Zhu, Y., Ong, C. S., & Huttley, G. A. (2020). Machine learning techniques for classifying
4785 the mutagenic origins of point mutations. *Genetics*, 215(1), 25–40. <https://doi.org/10.1534/genetics.120.303093>
- 4786 Zhuang, M.-Y., & Ho, L. C. (2020). The interplay between star formation and black
4787 hole accretion in nearby active galaxies. *ApJ*, 896. <https://doi.org/10.3847/1538-4357/ab8f2e>
- 4788 Zubovas, K., & King, A. (2012). Clearing out a galaxy. *ApJ*, 745. <https://doi.org/10.1088/2041-8205/745/2/L34>
- 4789 Zubovas, K., Sabulis, K., & Naujalis, R. (2014). Collapse and fragmentation of molecular
4790 clouds under pressure. *MNRAS*, 442(3), 2837–2854. <https://doi.org/10.1093/mnras/stu1048>
- 4791 Żywucka, N., Koziel-Wierzbowska, D., & Goyal, A. (2020). Catalogue with visual
4792 morphological classification of 32,616 radio galaxies with optical hosts, In *Proceedings of the international astronomical union symposium 356: Nuclear activity in galaxies across cosmic time*.
- 4793
- 4794
- 4795
- 4796
- 4797
- 4798
- 4799
- 4800
- 4801
- 4802