



Universiteit
Leiden
The Netherlands

Faint quasars at very low frequencies

Retana Montenegro, E.F.

Citation

Retana Montenegro, E. F. (2019, October 16). *Faint quasars at very low frequencies*. Retrieved from <https://hdl.handle.net/1887/79263>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/79263>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The following handle holds various files of this Leiden University dissertation:
<http://hdl.handle.net/1887/79263>

Author: Retana Montenegro, E.F.

Title: Faint quasars at very low frequencies

Issue Date: 2019-10-16

Faint Quasars at Very Low Frequencies



Edwin Retana-Montenegro
Leiden Observatory
Leiden University

A thesis submitted for the degree of
Doctor of Philosophy

2019

Faint Quasars at Very Low Frequencies

Proefschrift

ter verkrijging

van de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,

volgens besluit van het College voor Promoties

te verdedigen op dinsdag 16 October 2019

klokke 10.00 uur

door

Edwin Fernando Retana Montenegro

geboren te San José, Costa Rica

in 1986

Promotiecommissie

Promotor: Prof. dr. Huub Röttgering (Leiden University)

Co-promotor: Dr. Reinout van Weeren (Leiden University)

Overige leden: Prof. dr. Marijn Franx (Leiden University)

Prof. dr. Paul van der Werf (Leiden University)

Prof. dr. Peter Barthel (University of Groningen)

Dr. Lingyu Wang (SRON)

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Quasi-Stellar objects: Short historical perspective | 1 |
| 1.2 | Quasar Properties | 2 |
| 1.3 | Supermassive Black Holes | 4 |
| 1.4 | Very Low-Frequency Radio Astronomy | 5 |
| 1.5 | Outline of this thesis | 7 |
| 1.6 | Future prospects | 9 |
| 2 | Probing the Radio Loud/Quiet AGN dichotomy with quasar clustering | 11 |
| 2.1 | Introduction | 12 |
| 2.2 | Data | 17 |
| 2.2.1 | Sloan Digital Sky Survey | 17 |
| 2.2.2 | FIRST survey | 18 |
| 2.2.3 | Cross-matching of the SDSS and FIRST catalogs | 19 |
| 2.2.4 | Final quasar sample | 22 |
| 2.3 | Clustering of quasars | 24 |
| 2.3.1 | Two-point correlation functions | 24 |
| 2.3.2 | Error estimation | 27 |
| 2.3.3 | Bias, dark matter halo and black hole mass estimations | 28 |
| 2.4 | Results | 29 |
| 2.4.1 | Projected correlation function $w_p(r_p)$ | 29 |
| 2.4.2 | Quasar bias factors | 32 |
| 2.4.3 | Bias and host halo mass redshift evolution | 35 |
| 2.4.4 | Clustering as a function of radio-loudness | 35 |
| 2.4.5 | Clustering as function of BH masses | 37 |
| 2.4.6 | Clustering as a function of redshift | 38 |
| 2.4.7 | Clustering and AGN unification theories | 40 |
| 2.4.8 | The role of mergers in quasar radio-activity | 42 |
| 2.4.9 | Black hole properties involved in quasar triggering | 44 |
| 2.5 | Summary | 45 |
| 2.6 | Acknowledgements | 48 |
| 3 | Deep LOFAR 150 MHz imaging of the Boötes field: Unveiling the faint low-frequency sky | 49 |
| 3.1 | Introduction | 50 |
| 3.2 | Observations | 53 |
| 3.3 | Data reduction | 54 |
| 3.3.1 | Direction independent calibration | 54 |
| 3.3.2 | Direction dependent calibration | 57 |
| 3.3.3 | Combined facet imaging | 61 |
| 3.4 | Images and sources catalog | 62 |
| 3.4.1 | Final mosaic | 62 |
| 3.4.2 | Noise analysis and source extraction | 63 |
| 3.4.3 | Astrometry | 64 |
| 3.4.4 | Bandwidth and time smearing | 67 |
| 3.4.5 | Flux density scale accuracy | 68 |

| | | |
|----------|--|------------|
| 3.4.6 | Resolved sources | 69 |
| 3.4.7 | Completeness and reliability | 70 |
| 3.4.8 | Source catalog | 72 |
| 3.5 | Source counts | 75 |
| 3.5.1 | Size distribution and resolution bias | 75 |
| 3.5.2 | Visibility area | 79 |
| 3.5.3 | Completeness and reliability | 79 |
| 3.5.4 | Multiple-component sources | 81 |
| 3.5.5 | Differential source counts | 81 |
| 3.5.6 | Cosmic variance | 82 |
| 3.6 | Conclusions | 86 |
| 3.7 | Acknowledgements | 86 |
| 4 | On the Selection of High-z Quasars Using LOFAR Observations | 87 |
| 4.1 | Introduction | 88 |
| 4.1.1 | Method Overview | 90 |
| 4.1.2 | Optical selection | 90 |
| 4.1.2.1 | Selection of Ly α break objects | 90 |
| 4.1.2.2 | Separating quasars and stars | 90 |
| 4.1.3 | Mid-infrared selection | 91 |
| 4.1.4 | LOFAR detection | 91 |
| 4.1.5 | Visual inspection | 91 |
| 4.1.6 | Fitting the UV/optical to MIR spectral energy distributions of the candidate quasar sample | 92 |
| 4.2 | Results | 92 |
| 4.2.1 | Selecting candidate quasars in the NDWFS-Botes field | 92 |
| 4.2.1.1 | Data | 92 |
| 4.2.1.2 | Candidate quasars selection | 93 |
| 4.2.1.3 | Performance of the selection method | 99 |
| 4.2.1.4 | Effect of the radio spectral index distribution on the candidate quasar selection | 99 |
| 4.3 | Limitations | 101 |
| 4.4 | Summary | 102 |
| 4.5 | Conflict of Interest Statement | 103 |
| 4.6 | Author Contributions | 103 |
| 4.7 | Funding | 103 |
| 5 | The luminosity function of LOFAR radio-selected quasars at $1.4 \leq z \leq 5.0$ in the NDWFS-Boötes field | 104 |
| 5.1 | Introduction | 105 |
| 5.2 | Data | 109 |
| 5.2.1 | NOAO Deep Wide-field survey | 109 |
| 5.2.2 | SDSS, Pan-STARRS1, WISE, and Spitzer surveys | 110 |
| 5.2.3 | Spectroscopic quasars with optical and mid-infrared photometry | 113 |
| 5.3 | Classification | 113 |
| 5.3.1 | Training sample | 113 |
| 5.3.2 | Target sample | 116 |
| 5.3.3 | Classification algorithms | 117 |
| 5.3.3.1 | Random forest | 117 |
| 5.3.3.2 | Support vector machines | 118 |
| 5.3.3.3 | Bootstrap aggregation on K-nearest neighbors | 118 |
| 5.3.3.4 | Performance | 118 |
| 5.3.3.5 | Classification results | 119 |
| 5.3.3.6 | Radio data | 120 |
| 5.4 | Photometric redshifts | 121 |
| 5.4.1 | Nadaraya-Watson kernel regression | 121 |
| 5.4.2 | Quasar training sample | 123 |
| 5.4.3 | Redshift estimation | 123 |
| 5.4.4 | Final quasar sample | 126 |
| 5.4.5 | LOFAR and wedge-based mid-infrared selection of quasars | 131 |
| 5.5 | Luminosity Function of radio-selected quasars | 139 |

| | | |
|-------|--|------------|
| 5.5.1 | Selection completeness and accuracy of photometric redshifts | 139 |
| 5.5.2 | Simulated Quasar Spectra | 140 |
| 5.5.3 | K-correction | 141 |
| 5.5.4 | Quasar Luminosity function | 142 |
| 5.6 | Results | 145 |
| 5.6.1 | Model-fitting | 145 |
| 5.6.2 | Comparison to previous works | 148 |
| 5.6.3 | Density evolution of RSQs | 152 |
| 5.6.4 | Contribution of RSQs to IGM Photoionization | 156 |
| 5.7 | Discussion | 160 |
| 5.7.1 | The origins of radio-emission in RSQs | 160 |
| 5.7.2 | The environment of RSQs | 161 |
| 5.7.3 | RSQs and their location in spectroscopic parameter spaces | 162 |
| 5.8 | Conclusions | 163 |
| 5.A | Appendix: A sample of false color RGB ($R=B_W$, $G=R$, $B=I$) images | 165 |
| | References | 168 |
| | Dutch Summary | 193 |
| | English summary | 200 |
| | Publications | 205 |
| | Curriculum vitae | 207 |
| | Acknowledgements | 208 |

Chapter 1

Introduction

1.1 Quasi-Stellar objects: Short historical perspective

The development of radio-telescopes in the nineteen-sixties made this decade a golden age for astronomy. The first two important discoveries were the cosmic background radiation and pulsars. The third was that of quasars (acronym for quasi stellar radio source) by Dutch-American astronomer Maarten Schmidt. In 1963, Schmidt obtained the spectrum of the star-like object coinciding with the position of the radio source number 273 in the (3C) 3rd Cambridge Catalog of Radio Sources (Edge et al. 1959). The spectrum looked unlike the spectrum of any star because it contained broad-emission lines, unlike stellar spectra which contain narrow-absorption lines. **Even** more puzzling was the fact that the emission lines did not match any of the known chemical elements. Schmidt quickly realized that these lines corresponded to hydrogen emission lines, which were shifted towards the red end of the spectrum by a factor of 0.158, which corresponds to a receding velocity of one-sixth of the velocity of light (Schmidt 1963). The implicit distance indicates that the light from the source 3C273 had traveled by almost two billion years before reaching Earth. To be visible at such distances implies that 3C273 is very powerful and luminous, suggesting the operation of powerful energetic mechanisms

unknown at the time of its discovery. Schmidt's pioneering work on the quasar 3C273 marked the beginning of a journey of discovery that continues until this day, with the number of quasars discovered continuously increasing, and the confirmation of that some already exist less than seven hundred million years after the big bang (Bañados et al. 2018).

The discovery of the first quasar represents a historic landmark in astronomy for several reasons. First, it is a superb example of the synergy between radio and optical astronomy leading to the discovery of a new class of astronomical objects. Secondly, until then, most cosmologists and astronomers believed in Fred Hoyle's steady state theory (Hoyle 1948), which proposes that the density of the universe remains unchanged during its expansion because new matter (stars and galaxies) is continuously being created. According to this theory, the expanded space refills with new stars and galaxies, so that the universe in the present is not different from how it was in the past and how it will be in the future. The discovery of distant quasars implies that the universe in the past looked different compared to the current universe. This implies that the universe is evolving, which contradicts the steady state theory. Thirdly, it was the basis of the recognition of ubiquitousness of black holes (BHs) in the universe, which are now an essential part of the theories and models of formation and evolution of galaxies and stars.

1.2 Quasar Properties

By the end of the sixties, Schmidt had studied a large sample of quasars and defined what he considered to be their properties: star-like morphology associated with a radio source, variable light, large UV flux, broad emission lines, and large redshift (Schmidt 1969). We now know that not every quasar has all of these properties. For instance, only a fraction ($\sim 10\%$ in traditional radio surveys) of the quasars can be associated with a bright radio counterpart. One property that is common though is their high UV rest-frame luminosity. Therefore it is possible to detect them at very large distances (and hence at earlier epochs). An important difference between stellar sources and quasars is

their spectral energy distributions (SEDs). The former can be described approximately by single-temperature black-body SEDs, while the latter can be characterized roughly using a power-law SED.

The temporal variability of quasars was one of their first properties to be studied in detail (Matthews & Sandage 1963; Smith & Hoffleit 1963), and confirmed to be an important property of active galactic nuclei (AGN). The origin of temporal variability in quasars is still under investigation, however instabilities in the accretion disk or jets have been suggested as explanations. AGN variability can be exploited using a technique called reverberation mapping (RM) to probe the size and structure of the broad-line region (BLR), and to obtain estimates of the BH masses (Peterson 1988; Peterson et al. 2004). Using this technique, observing campaigns monitor the continuum and emission-line brightness of quasars; the time-delays between brightness measurements can be used to derive the size of the region where the emission-line comes from. Assuming that classical Newtonian mechanics describes the motion of gas in the vicinity of the central BH, its mass can be estimated using Kepler's laws. RM has provided good estimates for the BH masses of low- z quasars (e.g. Kaspi et al. 2000), but it is not suitable for high-redshift, high-mass BH sources due to the longer variability scales (Kaspi et al. 2004; Lira et al. 2018). Finally, assuming that the motion of the gas in the BLR is virialized, RM provides the basis to obtain estimates of BH masses of high- z quasars using single-epoch spectra (Kaspi et al. 2000; McLure & Dunlop 2004).

The presence of strong, broad emission lines is a defining characteristic of quasars. These emission lines include the hydrogen Ly $_{\alpha}$ λ 1215, the hydrogen Balmer-series lines (H $_{\alpha}$ λ 6563, H $_{\beta}$ λ 4861, H $_{\gamma}$ λ 4340), and prominent lines of abundant ions such as MgII λ 2798, CIII λ 1909, and CIV λ 1549. These spectral features (especially the Ly $_{\alpha}$ emission line) make the colors of quasars very different from those of galaxies and most stars. In practice, this implies that the majority of quasars can be identified using 3 broadband optical filters: one containing the Ly $_{\alpha}$ emission, one blueward (the dropout band), and one redward. In fact, a large fraction of the ~ 592000 quasars currently known (Flesch 2015) have been discovered using color selection.

The quasar population is also divided according to their radio properties. The

majority of quasars (radio-quiet quasars, RQQs) have weak or absent radio emission, while a small fraction of 10 – 15% have strong radio emission (radio-loud quasars, RLQs). RLQs are often associated with bright radio sources characterized by powerful collimated jets (Bridle et al. 1994; Mullin et al. 2008), while RQQs usually remain radio-undetected in wide-field shallow radio surveys (White et al. 2007; Shen et al. 2009). This division still remains a point of discussion. Some authors have found that RLQs and RQQs have very similar properties (e.g. McLure & Dunlop 2001; Dunlop et al. 2003; Barvainis et al. 2005; Rochais et al. 2014), while others have demonstrated that there are important differences between them (e.g. Sulentic et al. 2003, 2007; Sikora et al. 2007; Kratzer & Richards 2015).

1.3 Supermassive Black Holes

Supermassive black holes (SMBHs) are compact astrophysical objects with masses of $10^6 M_{\odot} \lesssim M \lesssim 10^9 M_{\odot}$. According to the theory of general relativity, SMBHs are capable of deforming spacetime in a that both matter and electromagnetic radiation cannot escape from inside them. BHs are found as a solution to the Einstein field equations (EFEs, Einstein 1916) that describe the gravitational field outside a spherical mass that has neither electric charge nor angular momentum (Schwarzschild 1916). A rotating uncharged BH can be described using the solutions to the EFEs obtained by Roy Kerr in 1963 (Kerr 1963). A few years after the publication of the discovery of the first quasar by Schmidt, Donald Lynden-Bell proposed that SMBHs are located at the center of galaxies, and are responsible for powering quasars (Lynden-Bell 1969). The presence of central SMBHs in galaxies provides a satisfactory explanation for many of the observational properties of quasars such as: their extreme bolometric luminosity due to radiation emitted from accreting matter into the BH, the rapid orbital motion of the filaments and gas clouds in the local vicinity of the BH that can explain the broadening of the emission lines, and the compact emission region and its related short time-scale variability.

1.4 Very Low-Frequency Radio Astronomy

Radio astronomy had its origins at very low-frequencies (10-300 MHz) with the serendipitous discovery of radio-emission coming from the Galactic center in 1933 by Karl G. Jansky, using an antenna designed to receive radio waves at a frequency of 20.5 MHz (Jansky 1933). In 1937, radio-engineer Grote Reber designed and built a steerable paraboloid reflector that enabled him to confirm Jansky's discovery (Reber 1940a,b), and to carry out the the first systematic radio-survey at 160 MHz (Reber 1944). In his radio contour maps, radio-emission is aligned with the shape of the Milky way and clearly its center is clear visible, along with concentrations towards the direction of the constellations Cygnus, Cassiopeia, Canis Major, and Sagitarius. Additionally, as mentioned earlier, low-frequency radio observations played a crucial role in the discovery of the first quasar by Schmidt (1963), who used the 3rd Cambridge Catalog of Radio Sources (Edge et al. 1959) deduced from observations by the Cambridge Interferometer operating at 159 MHz.

The pursuit of higher sensitivities and higher angular resolutions drove the development of radio-astronomy in the second half of the twentieth century towards frequencies above 1 GHz. At the beginning, this was an obvious choice to obtain higher resolutions, as the angular resolution of an interferometer is inversely proportional to its observing frequency. However, the main reason behind this bifurcated focus is that the Earth's ionosphere poses an additional challenge at low-frequencies. Electrons in the ionosphere disturb the incoming wavefronts causing a phase shift that increases linearly with wavelength (Kassim et al. 1993). In the conventional self-calibration (Pearson & Readhead 1984), it is assumed that the phase shifts across the field-of-view (FOV) are constant (not changing with direction) therefore it is possible to determine a phase correction for each antenna for each polarization that is valid for the entire FOV. The wide fields of low-frequency radio-telescopes imply that sources significantly separated within the FOV will have lines-of-sight through the ionosphere with different electron densities, and therefore their phase shifts differ. The assumption that the differential phase shifts across the FOV can be ignored breaks down, and simple self-calibration is not applicable.

The beginning of the twenty-first century witnessed a renewed interest in very low-

frequency radio-astronomy. This was driven by advances in modern computing and radio-interferometry technology, development of new calibration algorithms, the scientific motivation of probing the relatively unexplored very low-frequency parameter space, and construction of the Square Kilometer Array (SKA, Schilizzi 2005). The SKA is the largest radio-telescope ever proposed, and will be built in Australia and several African countries including South Africa; and various SKA “pathfinders” projects operating at low-frequencies have been built to pave the way for the SKA. These projects include the Long Wavelength Array (LWA; Taylor 2007), the Murchison Wide-field Array (MWA; Lonsdale et al. 2009; Tingay et al. 2013), and the Low Frequency Array (LOFAR; van Haarlem et al. 2013). These radio-telescopes will serve as testbeds in which to evaluate the technologies, observing strategies, calibration algorithms, and computational challenges that will be eventually used in the construction and operation of the SKA.

Radio-telescopes such as the Jansky Very Large Array (JVLA) and Giant Metrewave Radio Telescope (GRMT) are based on a steerable antenna design, while LOFAR is based on phased-array technology. A phased-array radio-telescope is composed of stations that contain a certain number of dipoles at fixed orientation. Currently, (as of January 2018) there are thirty eight stations distributed across the Netherlands, with an additional thirteen stations located in Germany, France, United Kingdom, Ireland, Sweden, and Poland. There are two different types of dipole antennas: Low Band Antenna (LBA) and High Band Antenna (HBA), optimized to operate at 10-80 MHz and 120-240 MHz, respectively. The signals from each dipole are digitized and combined to create a digital beam. The fact that the beams are digital makes it possible to create different combinations of pointing directions and observing frequencies, limited only by the total bandwidth of the radio-telescope. Effectively, the large instantaneous FOV and multi-beam capabilities make LOFAR a powerful sky-survey machine.

Low-frequency radio-telescopes operate in a regime where ionospheric distortions have a significant impact on the observations. The lack of calibration algorithms that efficiently correct for ionospheric distortions is one of the difficulties that hindered the progress of low-frequency radio-astronomy in the past. In the last fifteen years, several direction-dependent calibration algorithms have been developed, including field-based

calibration (Cotton et al. 2004), Source Peeling and atmospheric modeling (SPAM Intema et al. 2009; Intema 2014), SAGECal (Kazemi et al. 2011; Yatawatta et al. 2013), facet-calibration (van Weeren et al. 2016; Williams et al. 2016), and kMS/DDFACET (Tasse 2014; Smirnov & Tasse 2015; Tasse et al. 2018)

1.5 Outline of this thesis

Quasars represent the active phase of SMBHs, and are among the most luminous, powerful, and energetic objects known in the universe. The goal of this thesis is to use low-frequency and high-frequency radio observations to address the following questions:

- Is the radio loud/quiet quasar dichotomy real?
- Can deep low-frequency radio observations be used to effectively select high- z quasars?
- How does the faint radio-selected quasar population evolve with redshift?
- Is the environment of quasars related to the origin of their radio-emission?

In this thesis, the main tools used are low-frequency and high-frequency radio imaging, spectroscopic quasar catalogs, and ancillary optical and infrared data. Below there is a detailed description of the chapter contents.

In **Chapter 2**, we investigate the clustering properties of 45441 RQQs and 3493 RLQs drawn jointly from the Sloan Digital Sky Survey (SDSS, York et al. 2000; Schneider et al. 2010) and Faint Images of the Radio Sky at 20 cm (FIRST, Becker et al. 1995) in the range $0.3 < z < 2.3$. From the clustering properties, we deduce that RLQs in our sample inhabit massive dark matter haloes with masses of $M_{\text{DMH}} \gtrsim 10^{13.5} h^{-1} M_{\odot}$ at all redshifts, which corresponds to the mass scale of galaxy groups and galaxy clusters. RQQs reside in less massive haloes of a few times $\sim 10^{12} h^{-1} M_{\odot}$. Additionally, the clustering signal shows a dependence on BH mass, with the quasars powered by the most massive BHs clustering more strongly than quasars having less massive BHs. These results suggest that there is a real dichotomy between RLQs and RQQs, at least considering the masses of their host dark matter haloes.

Chapter 3 presents a deep radio-survey (with a central rms of $55 \mu\text{Jy}$) of the NOAO Deep Wide-field Survey (NDWFS) Boötes field (Jannuzi & Dey 1999) conducted with LOFAR at 120-168 MHz. This field has a large wealth of multi-wavelength data available. A total of 55 hours of LOFAR data have been calibrated using the directional-dependent calibration method presented by van Weeren et al. 2016. The final mosaic has an angular resolution of $3.98'' \times 6.45''$ and the resulting catalog contains 10091 radio sources (5σ limit) over an area of 20 deg^2 . Our differential source counts present a flattening below sub-mJy flux densities, which agrees with previous results from higher frequency surveys. This flattening has been argued to be due to an increasing contribution of star-forming galaxies and faint AGN. Moreover, the contribution of cosmic variance to the scatter in source counts measurements is evaluated. We find that the scatter due to cosmic variance is larger than the Poissonian errors of the source counts, and it may explain the discrepancies from previously reported source counts at flux densities $S < 1 \text{ mJy}$.

Chapter 4 describes a method to identify candidate quasars that combines optical/infrared color selection with 5σ LOFAR detections at 150 MHz. This method is applied in a region of $\sim 9 \text{ deg}^2$ located in the NDWFS-Boötes field using the LOFAR mosaic presented in Chapter 3, along with multi-wavelength data available for this region. The effect of the radio spectral index distribution on the selection of candidate quasars is investigated by combining the LOFAR observations with Westerbork Synthesis Radio Telescope (WSRT) imaging at 1400 MHz (de Vries et al. 2002). The candidate quasars detected by LOFAR and WSRT have a steep distribution of spectral indices with a median value of $\alpha_{150-1400 \text{ MHz}} = -0.73 \pm 0.07$. For the candidates undetected by WSRT, we find an upper limit for the distribution of spectral indexes of $\alpha_{\text{upp}} < -0.75$. As the upcoming LOFAR wide area surveys (Röttgering et al. 2011) are much deeper than the traditional 1.4 GHz surveys like NVSS (Condon et al. 1998) and FIRST (Becker et al. 1995), the combination of LOFAR observations with optical/infrared imaging will be an excellent fishing ground for obtaining large samples of quasars.

Chapters 2 and 4 demonstrate that the combination of optical, mid-infrared, and radio data can provide valuable insights into the SMBHs powering quasar activity. In

this respect, **Chapter 4** capitalizes on the wealth of radio, optical, and mid-infrared data available and the ever-growing number of quasars to identify RSQs in the NDWFS-Boötes field. This provides a robust statistical sample to draw conclusions regarding the evolution of RSQs across cosmic time, and possible origins of their radio emission. The identification of faint RSQs is an essential step in understanding the radio-loudness distribution dichotomy in quasars.

In **Chapter 5**, we use machine learning (ML) algorithms to compile a sample of quasars to investigate the luminosity function of quasars detected by LOFAR (radio-selected quasars, RSQs). The sample comprises 134 objects, including both photometrically-selected candidate quasars (51) and spectroscopically confirmed quasars (83). The depth of our LOFAR observations allows us to detect the radio-emission of quasars that otherwise would be classified as radio-quiet. In our final sample, a fraction of 66% of the quasars are fainter than $M_{1450} < -24.0$, a regime where the luminosity function of RSQs has not been studied before. Our results agree with a pure luminosity evolution model at $z < 2.4$ and luminosity evolution and density evolution model at redshift $z > 2.4$. By comparing the spatial density of RSQs with that of faint quasars at similar redshifts, we find that RSQs may compose to up $31 \pm 22\%$ of the *total* (radio-detected and radio-undetected) faint quasar population. This fraction, within uncertainties, seems to remain roughly constant with redshift.

1.6 Future prospects

The study of quasars is a very active field of research and will continue to be so in the years to come. Radio (LoTSS, Shimwell et al. 2018; FIRST, Becker et al. 1995), optical (SDSS, York et al. 2000; PanSTARRS, Chambers et al. 2016), mid-infrared (Spitzer; WISE, Cutri 2013), and spectroscopic (SDSS Schneider et al. 2010; BOSS Pâris et al. 2018; LAMOST, Yao et al. 2019; SHELLQs Matsuoka et al. 2016) surveys have already observed hundreds of square degrees in the Northern hemisphere, and will provide large samples containing hundreds of thousands of quasars, such as those used in this thesis. Future telescopes will have higher sensitivity, and will make it possible to

further increase the number of known quasars. Planned surveys with facilities such as the Large Synoptic Survey Telescope (LSST, Ivezić et al. 2019), 4MOST (de Jong et al. 2012), Dark Energy Spectroscopic Instrument (DESI, DESI Collaboration et al. 2016), Square Kilometre Array (SKA, Schilizzi 2005), Wide Field Infrared Survey Telescope (WFIRST, Spergel et al. 2015) and Euclid (**Laureijs et al. 2011**) will expand the number of known quasars at all redshifts to the order of millions. This will open a new unexplored parameter space for the study of quasar activity. It is expected that observations obtained with these surveys will have an important role in addressing many of the open questions regarding quasars. Some of these include: when the first quasars were formed?, what is the nature of faint quasars?, and what is the role of quasars in cosmic reionization?.

Chapter 2

Probing the Radio

Loud/Quiet AGN dichotomy with quasar clustering

Abstract:

We investigate the clustering properties of 45441 radio-quiet quasars (RQQs) and 3493 radio-loud quasars (RLQs) drawn from a joint use of the Sloan Digital Sky Survey (SDSS) and Faint Images of the Radio Sky at 20 cm (FIRST) surveys in the range $0.3 < z < 2.3$. This large spectroscopic quasar sample allow us to investigate the clustering signal dependence on radio-loudness and black hole (BH) virial mass. We find that RLQs are clustered more strongly than RQQs in all the redshift bins considered. We find a real-space correlation length of $r_0 = 6.59_{-0.24}^{+0.33} h^{-1} \text{Mpc}$ and $r_0 = 10.95_{-1.58}^{+1.22} h^{-1} \text{Mpc}$ for RQQs and RLQs, respectively, for the full redshift range. This implies that RLQs are found in more massive host haloes than RQQs in our samples, with mean host halo masses of $\sim 4.9 \times 10^{13} h^{-1} M_{\odot}$ and $\sim 1.9 \times 10^{12} h^{-1} M_{\odot}$, respectively. Comparison

Retana-Montenegro and Röttgering, A&A, 600, A97 (2017)

with clustering studies of different radio source samples indicates that this mass scale of $\gtrsim 1 \times 10^{13} h^{-1} M_{\odot}$ is characteristic for the bright radio-population, which corresponds to the typical mass of galaxy groups and galaxy clusters. The similarity we find in correlation lengths and host halo masses for RLQs, radio galaxies and flat-spectrum radio quasars agrees with orientation-driven unification models. Additionally, the clustering signal shows a dependence on black hole (BH) mass, with the quasars powered by the most massive BHs clustering more strongly than quasars having less massive BHs. We suggest that the current virial BH mass estimates may be a valid BH proxies for studying quasar clustering. We compare our results to a previous theoretical model that assumes that quasar activity is driven by cold accretion via mergers of gas-rich galaxies. While the model can explain the bias and halo masses for RQQs, it cannot reproduce the higher bias and host halo masses for RLQs. We argue that other BH properties such as BH spin, environment, magnetic field configuration, and accretion physics must be considered to fully understand the origin of radio-emission in quasars and its relation to the higher clustering.

2.1 Introduction

Quasars are luminous active galactic nuclei (AGN) powered by supermassive black holes (SMBHs) (Salpeter 1964; Lynden-Bell 1969). The role of AGN activity in galaxy formation and evolution processes is still not well understood. Evidence for a co-evolution scenario is provided by the empirical relationship between the host galaxy velocity dispersion and the mass of their central black holes (BHs) (Ferrarese & Merritt 2000; Gebhardt et al. 2000). At low- z , the analysis of stars and gas dynamics in the nucleus of nearby galaxies (Davies et al. 2006; de Francesco et al. 2006; Pastorini et al. 2007; de Francesco et al. 2008; Siopis et al. 2009; Walsh et al. 2013) and the reverberation mapping technique (Peterson 1988; Peterson et al. 2004; Doroshenko et al. 2012; Grier et al. 2012) have found that the most massive galaxies harbour the most massive BHs. At high- z , virial BH mass (M_{BH}) estimations based on single-epoch spectra employing empirical scaling relations (e.g. Kaspi et al. 2000; McLure & Dunlop 2004; Shen et al.

2008) suggest that SMBHs with masses $> 10^9 M_\odot$ were already in place at $z \gtrsim 5$ (Willott et al. 2003; Jiang et al. 2007b; Mortlock et al. 2011; Yi et al. 2014).

Because of their high-luminosity, quasars are excellent tracers of the large-scale structure up to $z \sim 6$. Recent large optical surveys using wide field integral spectrographs, such as the Sloan Digital Sky Survey (SDSS, York et al. 2000) and the 2dF QSO Redshift Survey (2QZ, Croom et al. 2004) have revealed thousands of previously unknown quasars. These newly detected quasars can be used to construct large statistical samples to study quasar clustering in detail across cosmic time. Several authors have found that quasars have correlation lengths of $r_0 = 5 h^{-1} - 8.5 h^{-1}$ Mpc at $0.8 < z < 2.0$, indicating that they reside in massive dark matter haloes (DMH) with masses of $\sim 10^{12} - 10^{13} M_\odot$ (e.g. Porciani et al. 2004; Myers et al. 2006; da Ângela et al. 2008; Ross et al. 2009; Shen et al. 2009).

Such clustering measurements provide a means to probe the outcome of any cosmological galaxy formation model (Springel et al. 2005; Hopkins et al. 2008), to understand how SMBH growth takes place (di Matteo et al. 2005; Bonoli et al. 2009; Shankar et al. 2010b), to define the quasar host galaxies characteristic masses (Shankar et al. 2010a; Fanidakis et al. 2013b), and to comprehend the interplay between its environment and the accretion modes (Fanidakis et al. 2013a).

Recently, galaxy clustering studies at intermediate and high redshift (Brown et al. 2000; Daddi et al. 2003; Coil et al. 2006; Meneux et al. 2009; Barone-Nugent et al. 2014; Skibba et al. 2014) have confirmed a strong correlation between galaxy luminosity and clustering amplitude, previously found at lower redshifts (Guzzo et al. 1997; Zehavi et al. 2005; Loh et al. 2010; Zehavi et al. 2011). This suggests that most the luminous galaxies reside in more overdense regions than less luminous ones. For quasar clustering, the picture is less clear. Several authors have found a weak clustering dependency on optical luminosity (e.g., Adelberger et al. 2005; Croom et al. 2005; Porciani & Norberg 2006; Myers et al. 2006; da Ângela et al. 2008; Shanks et al. 2011; Shen et al. 2013; Eftekharzadeh et al. 2015). These clustering results are in disagreement with the biased halo clustering idea, in which more luminous quasars reside in the most massive haloes, and therefore should have larger correlation lengths. A weak dependency on

the luminosity could imply that host halo mass and quasar luminosity are not tightly correlated, and both luminous and faint quasars reside in a broad range of host DMH masses. However, these conclusions can be affected because the quasar samples are flux-limited, and therefore often have small dynamical range in luminosity. In addition, the intrinsic scatter for the different observables, such as the luminosity, emission line width, and stars velocity dispersion, leads to uncertainties in derivables such as halo, galaxy, and BH masses, which in turn could mask any potential correlation between the observables and derivables. For instance, Croom (2011) assigned aleatory quasar velocity widths to different objects and re-determined their BH masses. They found that the differences between the randomized and original BH masses are marginal. This implies that the low dispersion in broad-line velocity widths provides little additional information to virial BH mass estimations.

Shen et al. (2009) divided their SDSS sample into bins corresponding to different quasar properties: optical luminosity, virial BH mass, quasar color, and radio-loudness. They found that the clustering strength depends weakly on the optical luminosity and virial BH masses, with the 10% most luminous and massive quasars being more clustered than the rest of the sample. Additionally, their radio-loud sample shows a larger clustering amplitude than their radio-quiet sources. Previous observations at low and intermediate redshift of the environments of radio galaxies and radio-loud AGNs suggest that these reside in denser regions compared with control fields (e.g., Miley et al. 2006; Wylezalek et al. 2013). At $z \gtrsim 1.5$, Mpc-sized dense regions have not yet virialized within a single cluster-sized DMH and are considered to be the progenitors of present day galaxy clusters (Kurk et al. 2004; Miley & De Breuck 2008). These results suggest that there is a relationship between radio-loud AGNs and the environment in which these sources reside (see Miley & De Breuck 2008 for a review).

Although the first known quasars were discovered as radio sources, only a fraction of $\sim 10\%$ are radio-loud (Sandage 1965). Radio-loud quasars (RLQs) and Radio-quiet quasars (RQQs) share similar properties over a wide wavelength range of the electromagnetic spectrum, from $100 \mu m$ to the X-ray bands. The main difference between both categories is the presence of powerful jets in RLQs (e.g. Bridle et al.

1994; Mullin et al. 2008). However, there is evidence that RQQs have weak radio jets (Ulvestad et al. 2005; Leipski et al. 2006). How these jets form is still a matter of debate and their physics is not yet completely understood. Several factors such as accretion rate (Lin et al. 2010; Fernandes et al. 2011), BH spin (Blandford & Znajek 1977; Sikora et al. 2007; Fernandes et al. 2011; van Velzen & Falcke 2013), BH mass (Laor 2000; Dunlop et al. 2003; Chiaberge & Marconi 2011), and quasar environment (Fan et al. 2001; Ramos Almeida et al. 2013), but most probably a combination of them, may be responsible for the conversion of accreted material into well-collimated jets. This division into RLQs and RQQs still remains a point of discussion. Some authors advocate the idea that radio-loudness (R , radio-to-optical flux ratio) distribution for optical-selected quasars is bimodal (Kellermann et al. 1989; Miller et al. 1990; Ivezić et al. 2002; Jiang et al. 2007a), while others have confirmed a very broad range for the radio-loudness parameter, questioning its bimodality nature (Cirasuolo et al. 2003; Singal et al. 2011, 2013).

An important question in the study of the bimodality for the quasar population is which physics sets the characteristic mass scale of quasar host halos and the BHs that power them. Specifically, studying the threshold for BH mass associated with the onset of significant radio activity is crucial for addressing basic questions about the physical process involved. According to the spectral analysis of homogeneous quasar samples, RLQs are associated to massive BHs with $M_{\text{BH}} \gtrsim 10^9$, while RQQs are linked to BHs with $M_{\text{BH}} \lesssim 10^8$ (Laor 2000; Jarvis & McLure 2002; Metcalf & Magliocchetti 2006). Other studies found that there is no such upper cutoff in the masses for RQQs and they stretch across the full range of BH masses (Oshlack et al. 2002; Woo & Urry 2002; McLure & Jarvis 2004).

An alternative way to indirectly infer BH masses for radio-selected samples is to use spatial clustering measurements. Most previous clustering analyses for radio selected sources have found they are strongly clustered with correlation lengths $r_0 \gtrsim 11 h^{-1}$ Mpc (Peacock & Nicholson 1991; Magliocchetti et al. 1998; Overzier et al. 2003). Magliocchetti et al. (2004) studied the clustering properties for a sample of radio galaxies drawn from the Faint Images of the Radio Sky at 20 cm (FIRST, Becker et al. 1995) and 2dF

Galaxy Redshift surveys (2dFGRS, Colless et al. 2001) and found that they reside in typical DMH mass of $M_{\text{DMH}} \sim 10^{13.4} M_{\odot}$, with a BH mass of $\sim 10^9 M_{\odot}$, a value consistent with BH mass estimations using composite spectra. A comparable limit for the BH mass was found by Best et al. (2005) analyzing a SDSS radio-AGN sample at low- z . Clustering measurements of the two-point correlation function for RLQs (e.g. Croom et al. 2005; Shen et al. 2009) obtained r_0 values consistent with those of radio galaxies. On the other hand, Donoso et al. (2010) found that RLQs are less clustered than radio galaxies, however, their sample was relative smaller.

Clustering statistics offer an efficient way to explore the connections between AGN types, including radio, X-ray, and infrared selected AGNs (Hickox et al. 2009); obscured and unobscured quasars (Hickox et al. 2011; Allevato et al. 2014b; DiPompeo et al. 2015); radio galaxies (Magliocchetti et al. 2002; Wake et al. 2008; Fine et al. 2011); blazars (Allevato et al. 2014a); and AGNs and galaxy populations: Seyferts and normal galaxies; and optical quasars and submillimeter galaxies (Hickox et al. 2012). These findings open up the possibility to explain the validity and simplicity of unification schemes (e.g. Antonucci 1993; Urry & Padovani 1995) for radio AGNs with clustering.

The purpose of the present study is to measure the quasar clustering signal, study its dependency on radio-loudness and BH virial mass, and derive the typical masses for the host haloes and the SMBHs that power these quasars. We use a sample of approximately 48000 uniformly selected spectroscopic quasars drawn from the SDSS DR7 (Shen et al. 2011) at $0.3 \leq z \leq 2.2$. In Section 3.2, we present our sample obtained from the joint use of the SDSS DR7 and FIRST surveys. The methods used for the clustering measurement are introduced in Section 2.3. We discuss our results for the measurement of the two-point correlation function for both RLQs and RQQs in Section 2.4. In addition, we compare our findings with previous results from the literature. Finally, in Section 3.6, we summarize our conclusions. Throughout this paper, we adopt a lambda cold dark matter cosmological model with the matter density $\Omega_m = 0.30$, the cosmological constant $\Omega_{\Lambda} = 0.70$, the Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and the rms mass fluctuation amplitude in spheres of size $8 h^{-1} \text{ Mpc}$ $\sigma_8 = 0.84$.

2.2 Data

2.2.1 Sloan Digital Sky Survey

The SDSS I/II was a photometric and spectroscopic survey of approximately one-fourth of the sky using a dedicated wide-field 2.5m telescope (Gunn et al. 1998). The resulting imaging provides photometric observations in five bands: u , g , r , i , and z (Fukugita et al. 1996). The selection for spectroscopic follow-up for the quasars at low redshift ($z \leq 3$) is done in the $ugri$ color space with a limiting magnitude of $i \leq 19.1$ (Richards et al. 2002). At high-redshift ($z \geq 3$), the selection is performed in $griz$ color space with $i < 20.2$. The quasar candidates are assigned to 3° diameter spectroscopic plates by a tiling algorithm (Blanton et al. 2003) and observed with double spectrographs with a resolution of $\lambda/\Delta\lambda \sim 2000$. Each plate hosts 640 fibers and two fibers cannot be closer than $55''$, which corresponds to a projected distance of $0.6 - 1.5 h^{-1}$ Mpc for $0.3 < z < 2.3$. This restriction is called fiber collisions, and causes a deficit of quasar pairs with projected separations ≤ 2 Mpc. We did not attempt to compensate for pair losses due to fiber collisions, therefore we only model our results for projected distances ≥ 2 Mpc.

We exploit the Shen et al. (2011) value-added catalog that is based on the main SDSS DR7 quasar parent sample Schneider et al. (2010). We select a flux limited $i = 19.1$ sample of 48338 quasars with $0.3 \leq z \leq 2.3$ from the Shen et al. (2011) catalog with the flag `UNIFORM_TARGET=1`. This sample includes both RLQs and RQQs selected uniformly by the quasar target selection algorithm presented in Richards et al. (2002). For quasar clustering studies, it is critical to use statistical samples that have been constructed using only one target selection algorithm. Therefore, this sample excludes SDSS objects with non-fatal photometric errors and are selected for spectroscopic follow-up based only on their radio detection in the FIRST survey (see Richards et al. 2002 for more details). The combination of quasars selected employing different target selections could lead to the appearance of potential systematics in the resulting sample. This includes higher clustering strength at large scales (Ross et al. 2009). Previous studies using uniform samples have shown that these are very stable and insensitive to

systematic effects such as dust reddening, and bad photometry (Ross et al. 2009; Shen et al. 2009, 2013).

2.2.2 FIRST survey

The FIRST survey (Becker et al. 1995) is a radio survey at 1.4 GHz that aims to map 10000 square degrees of the North and South Galactic Caps using the NRAO Very Large Array. The FIRST radio observations are done using the B-array configuration providing an angular resolution of $\sim 5''$ with positional accuracy better than $1''$ at a limiting radio flux density of 1 mJy (5σ) for point sources. FIRST was designed to have an overlap with the SDSS survey, and yields a 40% identification rate for optical counterparts at the $m_V \sim 23$ (SDSS limiting magnitude).

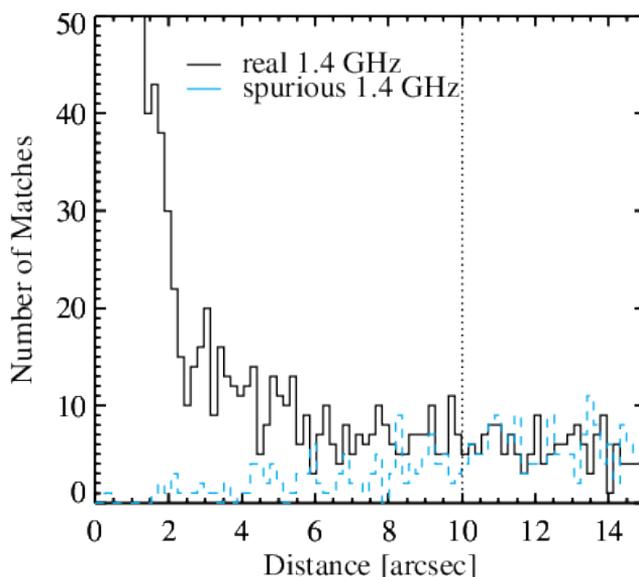


Figure 2.1: The solid histogram shows the distance distribution for SDSS quasar counterparts to $S_{1.4\text{GHz}} \geq 1.0$ mJy FIRST radio sources. Cyan dashed histogram indicates the distribution for spurious associations, which are obtained by vertically shifting the quasar positions by $1'$.

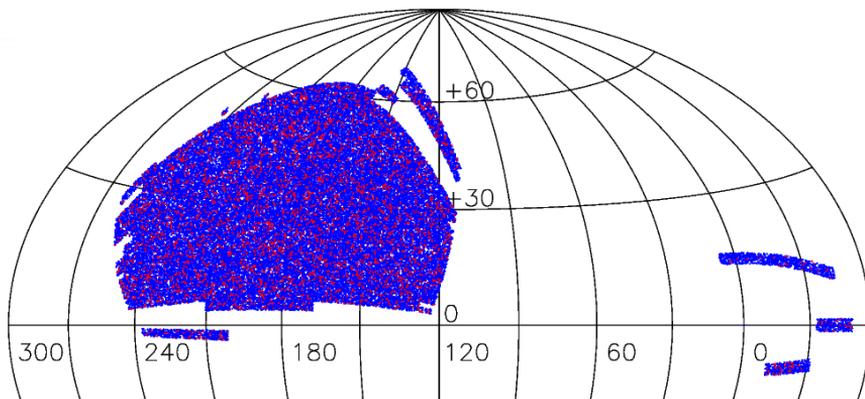


Figure 2.2: Aitoff projection for the sky coverage of the SDSS DR7 uniform quasar sample from Shen et al. (2011). RQQs are denoted by blue points, while the RLQs are represented by red points. See Section 3.2 for a description of the methodology employed in the selection for the RLQs.

2.2.3 Cross-matching of the SDSS and FIRST catalogs

The quasar catalog provided by Schneider et al. (2010) is matched to the FIRST catalog taking sources with position differences less than $2''$. However, this short distance prevents the identification of quasars with diffuse or complex radio emission. Therefore, to account for RLQs possibly missed by the original matching, we cross-matched the SDSS and FIRST catalogs with larger angular distances. To choose the upper limit for a new matching radius, we vertically shifted the quasar positions by $1'$ and proceeded to match again with the FIRST catalog. Shown by a solid line in Fig. 2.1 we reproduce the distribution of angular distances between SDSS objects and their nearest FIRST counterpart, and by a dashed line the we show distribution of spurious matches. The distribution of real matches presents a peak and a declining tail that flattens with increasing distance. Both distributions are at the same level at $\sim 10''$. This radius will be used as the maximum angular separation for matching the SDSS and FIRST surveys. This value is a good compromise between the maximum number of real identifications and keeping the spurious associations to a minimum. The total number of newly identified radio quasars with angular offsets between $2''$ and $10''$ is 409.

Some statistical matching methods, such as the likelihood ratio (LR), have been pro-

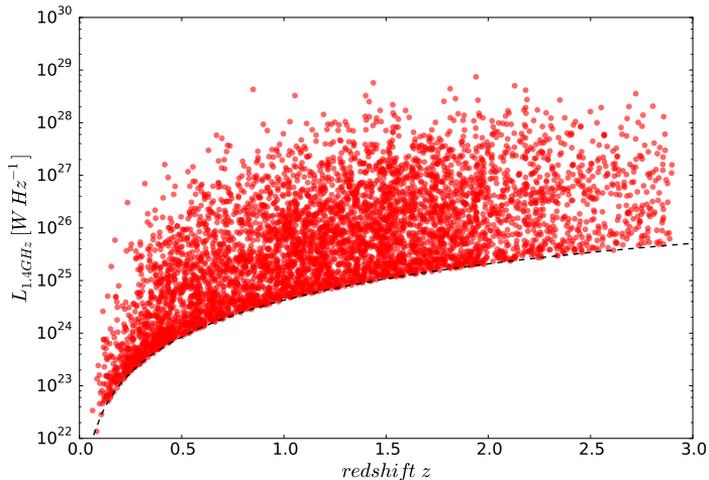


Figure 2.3: The 1.4 GHz restframe radio luminosity for the RLQs (red) detected in the FIRST radio survey. We assume a radio spectral index of 0.70, and a flux limit of 1.0 mJy. The dashed lines show the luminosity limit for the FIRST survey flux limit.

posed to robustly cross-match radio and optical surveys (e.g., Sutherland & Saunders 1992). Sullivan et al. (2004) showed that when the positional uncertainties for both radio and optical catalogues are small, the LR technique and positional coincidence yield very similar results. This is the case for both catalogs used in this work, which have accurate astrometry ($\sim 0.1''$ for SDSS, $\sim 1''$ for FIRST). The contamination rate by random coincidences (El Boucheffry & Cress 2007; Lindsay et al. 2014b) is:

$$P_C = \pi r_s^2 \rho, \quad (2.1)$$

where r_s is the matching radius, and $\rho \simeq 5.6 \text{ deg}^{-2}$ is the quasar surface density. For $r_s = 2''$, the expected number of contaminants in the RLQs sample is 2, while for $r_s = 10''$ this rate increases to 61. This small contamination fraction ($< 2\%$ from the total radio sample) is unlikely to affect our clustering measurements.

The sensibility for the FIRST survey is not uniform across the sky, with fluctuations due to different reasons, such as hardware updates, observing strategies, target declination, and increasing noise in the neighborhood of bright sources (Becker et al. 1995). Despite

all these potential limitations, the detection limit for most of the targeted sky is a peak flux density of 1 mJy (5σ), with only an equatorial strip having a slightly deeper detection threshold due to the combination of two observing epochs. We refer the interested reader to Helfand et al. (2015), where the impact of all the above mentioned aspects is discussed extensively. The flux limit of 1 mJy is considered only for peak flux density instead of integrated flux density. Hence a source with peak fluxes individually smaller than the detection threshold but with total flux greater than this value could not appear in our radio sample. In particular, lobe-dominated quasars (see Fanaroff & Riley 1974; hereafter FR2) with peak fluxes less than the flux limit suffer from a systematic incompleteness in comparison to core-dominated quasars (FRI). We investigate how not taking into account FIRST resolution effects could possibly affect our RLQ clustering measurements. We estimate the weights for RLQs with fluxes less than 5 mJy using the completeness curve from Jiang et al. (2007b), which takes into account the source morphology and rms values in the FIRST survey for SDSS quasars. We find that including a weighting scheme does not affect the clustering signal for RLQs.

We define a quasar to be radio-loud if it has a detection in the FIRST with a flux above 1 mJy, and radio-quiet if it is undetected in the radio survey. To minimize incompleteness due to the FIRST flux limit while retaining the maximum numbers of quasars for clustering measurements, we consider two radio-luminosity cuts: $L_{1.4\text{ GHz}} > 4 \times 10^{24} \text{ W Hz}^{-1}$ for $0.3 < z < 1.0$; and $L_{1.4\text{ GHz}} > 1 \times 10^{25} \text{ W Hz}^{-1}$ for $1.0 < z < 2.3$. Our parent sample then comprises a total of 45441 RQQs and 3493 RLQs with $0.3 < z < 2.3$, which corresponds to a radio-loud/-quiet source fraction of $\sim 7.2\%$. This ratio is in agreement with previous studied quasar samples (e.g., Jiang et al. 2007a; Hodge et al. 2011). This choice for the redshift range avoids the poor completeness at high- z due to color confusion with stars in the *ugri* color cube. The sky coverage of our final quasar sample of 6248 deg^2 is shown in Fig. 2.2. We calculate the radio-luminosity adopting a mean radio spectral index of $\alpha_{\text{rad}} = 0.7$ (where $S_\nu \propto \nu^{-\alpha}$) and applying the usual k-correction for the luminosity estimation. Fig. 2.3 shows the radio-luminosity for our quasar sample. The quasar distribution in the optical-luminosity redshift plane is displayed in Fig. 2.4. The normalized redshift and optical-luminosity distributions for

both samples show a good degree of similarity, this allows a direct comparison of their clustering measurements. We confirm this by applying two Kolmogorov-Smirnov (K-S) tests, which indicate a probability for the redshift and luminosity redshift distributions of 95% and 97%, respectively, that both samples (RLQs and RQQs) are drawn from the same parent distribution.

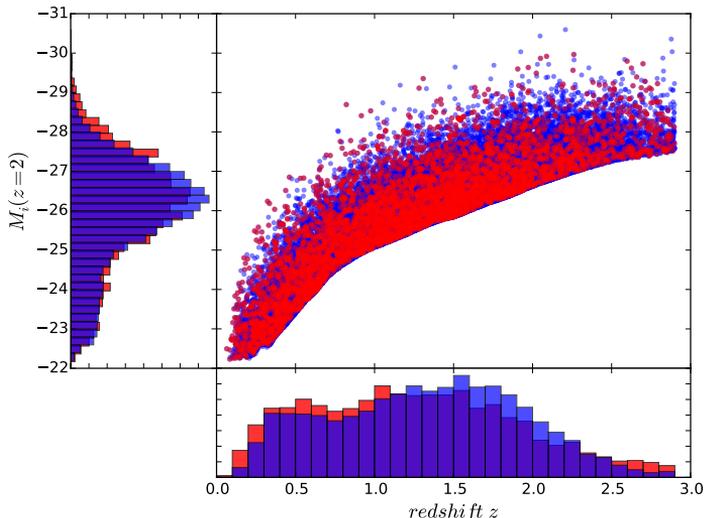


Figure 2.4: The distribution of RLQs (red) and RQQs (blue) in the optical-luminosity space. The absolute magnitude in the i -band at $z = 2$ $M_i(z = 2)$ is calculated using the K-correction from Richards et al. (2006). The left and bottom panels show the $M_i(z = 2)$ and redshift histograms. The normalized redshift and optical-luminosity distributions are displayed in the left and bottom panels. The normalized distributions for both samples show a good degree of similarity, allowing a direct comparison of their clustering measurements.

2.2.4 Final quasar sample

The final spectroscopic quasar sample restricted to $0.3 < z < 2.3$ provides an excellent dataset for probing the clustering dependence based on physical properties such as radio-loudness or BH virial mass. It is possible to explore how clustering depends on these properties to some degree across different redshift intervals. Previous quasar clustering studies (e.g., Croom et al. 2005; Ross et al. 2009; Shen et al. 2009) were limited by their sample size ($\lesssim 30000$ quasars) and studied the correlation function for RLQs in only one

Table 2.1: Main properties of our quasar samples. The bar denotes the median values.

| Sample | \bar{M}_{BH} [$\log(M_{\odot})$] | \bar{L}_{Bol} [$10^{46} \text{ erg s}^{-1}$] | $\bar{L}_{1.4 \text{ GHz}}$ [$10^{26} \text{ W Hz}^{-1}$] |
|---|--|--|--|
| $0.3 \leq z \leq 2.3$ | | | |
| All | 9.21 | 4.72 | - |
| RQQs | 9.19 | 3.57 | - |
| RLQs | 9.36 | 5.69 | 8.32 |
| $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ | 9.23 | 1.48 | - |
| $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ | 8.82 | 2.14 | - |
| $0.3 \leq z \leq 1.0$ | | | |
| RQQs | 8.80 | 0.90 | - |
| RLQs | 9.35 | 6.43 | 2.54 |
| $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ | 9.20 | 0.79 | - |
| $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ | 8.77 | 0.85 | - |
| $1.0 \leq z \leq 2.3$ | | | |
| RQQs | 9.15 | 4.70 | - |
| RLQs | 9.07 | 5.39 | 10.6 |
| $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ | 9.23 | 2.57 | - |
| $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ | 8.84 | 2.69 | - |

redshift bin corresponding to the entire redshift range of the sample. We take advantage of the higher quasar numbers of our sample and divide each redshift bin into smaller bins using radio-loudness and the virial BH masses as indicators, and still obtain a good S/N for the correlation function of the samples in our analysis. The $M_{\text{BH}} - z$ space is not uniformly populated. We limit our analysis to two mass samples that are separated according to their BH mass: $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ and $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$. The redshift distributions for these two mass bins are very different, with more massive BHs peaking at $z \sim 2$, while less massive at $z \sim 0.5$ (see Fig. 2.5). This hampers a direct comparison between their clustering measurements. Thus, we create control samples by randomly selecting quasars from the initial BH mass samples that are matched by their optical luminosity distribution. We verify that the resulting samples can be compared by applying a K-S test to the new redshift distributions. This indicates a probability of 97% that the mass samples are drawn from the same parent distribution. The properties for all the quasar samples are presented in Table 2.2.

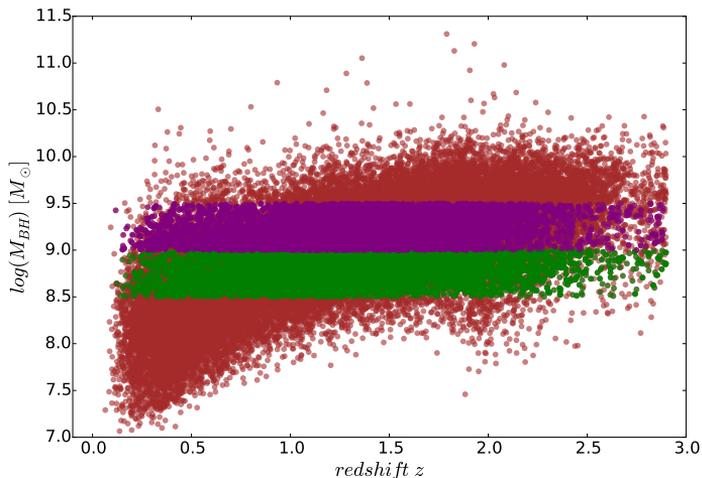


Figure 2.5: The quasar distribution in the virial BH mass plane. The quasars selected to match in optical luminosity with masses $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ are indicated with green color, and the objects with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ are represented by purple points. The properties of the mass samples are summarized in Table 2.2.

2.3 Clustering of quasars

2.3.1 Two-point correlation functions

The two-point correlation function (TPCF) $\xi(r)$ describes the excess probability of finding a quasar at a redshift distance r from a quasar selected randomly over a random distribution. To constraint this function, we create random catalogs with the same angular geometry and the same redshift distribution as the data with at least 70 times the number of quasars in the data sets to minimize the impact of Poisson noise. The redshift distributions corresponding to the different quasar samples are shown by the solid lines in Fig. 2.6.

The TPCF is estimated using the minimum variance estimator suggested by Landy & Szalay (1993)

$$\xi_{\text{LS}} = \frac{DD - 2DR + RR}{RR}, \quad (2.2)$$

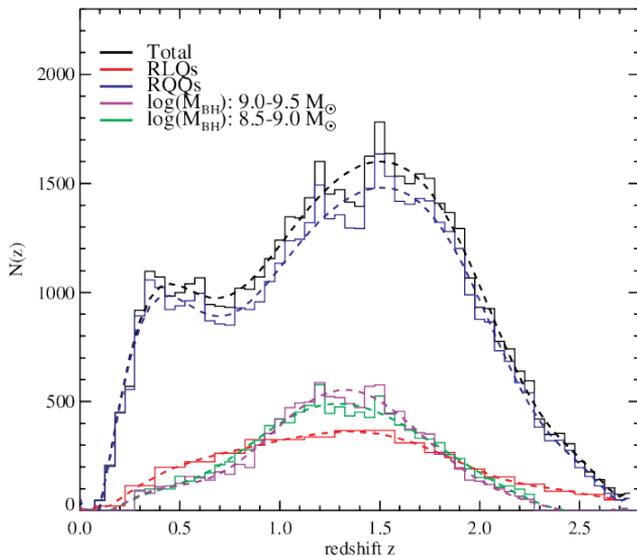


Figure 2.6: Redshift distributions for the total quasar sample (black), RQOs (blue), RLQs (red), quasars with $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ (green) and $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ (purple). The mass samples are matched in optical luminosity at each redshift interval (see Section 2.4.1 for more details). The solid lines are fitted polynomials used to generate the random quasar catalogs used in the correlation function estimations.

where DD is the number of distinct data pairs, RR is the number of different random pairs, and DR is the number of cross-pairs between the real and random catalogs within the same bin. All pair counts are normalized by n_{QSO} and n_R , respectively, the mean number densities in the quasar and random catalogs. We verify our estimates using the Hamilton estimator (Hamilton 1993), and find a good agreement of the results for both estimators within the error bars, although the LS estimator is preferred because it is less sensitive to edge effects.

In reality, observed TPCFs are distorted both at large and small scales. On smaller scales, quasars have peculiar non-linear velocities that cause an elongation along the line of sight, which is referred as the Finger of God effect (Jackson 1972). At larger scales, the coherent motion of quasars that are infalling onto still-collapsing structures produces a flattening of the clustering pattern to the observer. This distortion is called the Kaiser effect (Kaiser 1987).

Because of the existing bias mentioned earlier in redshift-space, a different approach is

used to minimize the distortion effects in the clustering signal (Davis & Peebles 1983). Following Fisher et al. (1994), we use the separation vector, $\mathbf{s} = \mathbf{s}_1 - \mathbf{s}_2$, and the line of sight vector, $\mathbf{l} = \mathbf{s}_1 + \mathbf{s}_2$; where \mathbf{s}_1 and \mathbf{s}_2 are the redshift-space position vectors. From these, it is possible to define the parallel and perpendicular distances for the pairs as:

$$\pi = \frac{|\mathbf{s} \cdot \mathbf{l}|}{|\mathbf{l}|}, \quad r_p = \sqrt{\mathbf{s} \cdot \mathbf{s} - \pi^2}. \quad (2.3)$$

Now, we can compute the correlation function $\xi(r_p, \pi)$ in a two-dimensional grid using the LS estimator, as in eq. (2.2). Because the redshift distortions only affect the distances in the π – direction, we integrate along this component and project it on the r_p – axis to obtain the projected correlation function

$$\frac{w_p(r_p)}{r_p} = \frac{2}{r_p} \int_0^\infty \xi(r_p, \pi) d\pi, \quad (2.4)$$

which is independent of redshift-space distortions, as it measures the clustering signal as a function of the quasar separation in the perpendicular direction to the line of sight. In practice, it is not feasible to integrate eq. (2.4) to infinity, thus an upper limit π_{max} to the integral shall be chosen to be a good compromise between the impact of noise and a reliable calculation of the measured signal. We try several π upper limits by fitting w_p to a power-law of the form (Davis & Peebles 1983)

$$w_p(r_p) = r_p \left(\frac{r_0}{r_p} \right)^\gamma \left[\frac{\Gamma(\frac{1}{2}) \Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})} \right], \quad (2.5)$$

where r_0 is the real-space correlation length, and γ the power-law slope. We use the range $2.0 \leq r_p \leq 130 h^{-1} \text{ Mpc}$ to determine the scale at which the clustering signal is stable (Fig. 2.7). We find that above $\pi = 63.1 h^{-1} \text{ Mpc}^{-1}$, the fluctuations in the

correlation length are within uncertainties and have poorer S/N. Thus, we take this value as our upper integration limit π_{max} , which is within the range $40 - 70 h^{-1} \text{Mpc}^{-1}$ of previous quasar clustering studies (e.g. Porciani et al. 2004; Ross et al. 2009).

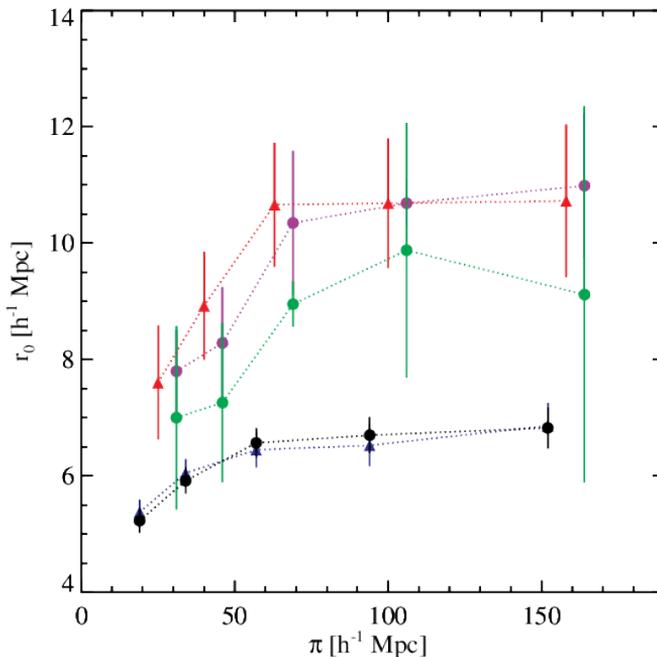


Figure 2.7: Real-space correlation length r_0 vs the parallel direction to the line of sight π for the full quasar sample (black circles), $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ sample (purple circles), $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ sample (green circles), RQQs (blue triangles), and RLQs (red triangles). For clarity, the mass samples have been shifted by $\pi = 6 h^{-1} \text{Mpc}$, and the full and RQQs samples by $\pi = 6 h^{-1} \text{Mpc}$.

2.3.2 Error estimation

We calculate the errors from the data itself by using the delete-one jackknife method (Norberg et al. 2009). We divide the survey into N_{sub} different sub-samples, and *delete* one sample at a time to compute the correlation function for $N_{sub} - 1$ sub-samples. This process is repeated N_{sub} times to obtain the correlation function for bin i in the jackknife sub-sample k , denoted by ξ_i^k . We can write the jackknife covariance matrix (e.g. Scranton et al. 2002; Norberg et al. 2009) as

$$C_{ij} = \frac{N_{sub} - 1}{N_{sub}} \sum_{k=1}^{N_{sub}} (\xi_i^k - \xi_i) (\xi_j^k - \xi_j), \quad (2.6)$$

with ξ_i the correlation function for all data at each bin i . We employ a total of $N_{sub} = 24$ sub-samples for our error estimations. Each sub-sample is chosen to be an independent cosmological volume with approximately the same number of quasars. The off-diagonal elements in the covariance matrix are small at large scales and could potentially insert some noise into the inverse matrix (Ross et al. 2009; Shen et al. 2009). Therefore, we employ only diagonal elements for the χ^2 fitting.

2.3.3 Bias, dark matter halo and black hole mass estimations

According to the linear theory of structure formation, the bias parameter b relates the clustering amplitude of large-scale structure tracers and the underlying dark matter distribution. The quasar bias parameter can be defined as

$$b^2 = w_{\text{QSO}}(r_p, z) / w_{\text{DM}}(r_p, z), \quad (2.7)$$

where w_{QSO} and w_{DM} are the quasar and dark matter correlation functions (Peebles 1980), respectively. We estimate the bias factor using the halo model approach, in which w_{DM} has two contributions: the 1-halo and 2-halo terms. The first term is related to quasar pairs from within the same halo, and the second one is the contribution from quasars pairs in different haloes. As the latter term dominates at large separations, we can neglect the 1-halo term and write w_{DM} as (Hamana et al. 2002)

$$w_{\text{DM}}(r_p, z) = w_{\text{DM}}^{2-h}(r_p, z) = r_p \int_{r_p}^{\infty} \frac{r \xi_{\text{DM}}^{2-h}(r)}{\sqrt{r^2 - r_p^2}} dr, \quad (2.8)$$

with

$$\xi_{\text{DM}}^{2-h}(r) = \frac{1}{2\pi^2} \int P^{2-h}(k) k^2 j_0(kr) dk, \quad (2.9)$$

where k is the wavelength number, h refers to the halo term, $P^{2-h}(k)$ is the Fourier transform of the linear power spectrum (Efstathiou et al. 1992) and $j_0(x)$ is the spherical Bessel function of the first kind.

With the bias factor, it is possible to derive the typical mass for the halo in which the quasars reside. We follow the procedure described in previous AGN clustering studies (e.g., Myers et al. 2007; Krumpe et al. 2010; Allevalo et al. 2014b) using the ellipsoidal gravitational collapse model of Sheth et al. (2001) and the analytical approximations of van den Bosch (2002).

2.4 Results

2.4.1 Projected correlation function $w_p(r_p)$

First, we check the consistency of our results by calculating the real-space TPCF for the entire quasar sample in the interval $0.3 \leq z \leq 2.3$ and compare it with previous clustering studies. We select a fitting range of $2 \leq r_p \leq 130 h^{-1} \text{ Mpc}$ to have a distance coverage similar to previous quasar clustering studies (e.g., Shen et al. 2009). To determine the appropriate values for our TPCFs, we fit eq. 2.5 with r_0 and γ as free parameters using a χ^2 minimization technique. We find a real-space correlation length of $r_0 = 6.81_{-0.30}^{+0.29} h^{-1} \text{ Mpc}$ and a slope of $\gamma = 2.10_{-0.05}^{+0.05}$, which is in good agreement with the results of Ross et al. (2009) for the SDSS DR5 quasar catalog, and Ivashchenko et al. (2010) for their SDSS DR7 uniform quasar catalog. Subsequently, we derive the best-fit r_0 and γ values for all the quasars samples. The best-fitting values and their respective errors are presented in Table 2.2.

We then split each redshift range according to their radio-loudness and virial BH mass to study the clustering dependence on these properties. The results of our clustering analysis for the different quasar sub-samples as a function of radio-loudness are presented

Table 2.2: Best-fitting correlation function model parameters for the quasar samples. The range for the fits is $2.0 \leq r \leq 130 h^{-1}$ Mpc.

| Sample | z | N_{QSO} | r_0 [h^{-1} Mpc] | γ | χ^2 | DOF | b | M_{DMH} [$h^{-1} M_{\odot}$] |
|---|------|------------------|--------------------------|------------------------|----------|-----|-----------------|--|
| $0.3 \leq z \leq 2.3$ | | | | | | | | |
| All | 1.30 | 48338 | $6.81^{+0.29}_{-0.30}$ | $2.10^{+0.05}_{-0.05}$ | 20.17 | 7 | 2.00 ± 0.08 | $2.33^{+0.41}_{-0.38} \times 10^{12}$ |
| RQs | 1.30 | 45441 | $6.59^{+0.33}_{-0.24}$ | $2.09^{+0.10}_{-0.09}$ | 19.60 | 7 | 2.01 ± 0.08 | $2.38^{+0.42}_{-0.38} \times 10^{12}$ |
| RLQs | 1.32 | 3493 | $10.95^{+1.22}_{-1.58}$ | $2.29^{+0.53}_{-0.34}$ | 1.06 | 7 | 3.14 ± 0.34 | $1.23^{+0.47}_{-0.39} \times 10^{13}$ |
| $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ | 1.31 | 11356 | $8.53^{+1.57}_{-2.25}$ | $1.84^{+0.21}_{-0.20}$ | 0.69 | 6 | 2.64 ± 0.42 | $6.57^{+0.43}_{-0.31} \times 10^{12}$ |
| $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ | 1.31 | 11356 | $10.45^{+0.79}_{-0.98}$ | $2.36^{+0.18}_{-0.17}$ | 1.99 | 6 | 2.99 ± 0.43 | $1.02^{+0.55}_{-0.42} \times 10^{13}$ |
| $0.3 \leq z \leq 1.0$ | | | | | | | | |
| RQs | 0.65 | 13219 | $6.85^{+0.45}_{-0.40}$ | $2.04^{+0.08}_{-0.07}$ | 2.74 | 7 | 1.52 ± 0.09 | $3.53^{+1.07}_{-0.91} \times 10^{12}$ |
| RLQs | 0.71 | 1019 | $18.39^{+1.75}_{-2.01}$ | $2.40^{+0.19}_{-0.16}$ | 1.95 | 4 | 4.63 ± 0.58 | $1.16^{+0.37}_{-0.33} \times 10^{14}$ |
| $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ | 0.74 | 2604 | $10.90^{+1.97}_{-2.48}$ | $1.54^{+0.15}_{-0.14}$ | 0.54 | 6 | 2.83 ± 0.45 | $2.89^{+1.56}_{-1.23} \times 10^{13}$ |
| $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ | 0.74 | 2604 | $15.26^{+2.15}_{-2.09}$ | $2.29^{+0.56}_{-0.36}$ | 1.11 | 6 | 3.56 ± 1.02 | $5.59^{+5.10}_{-3.53} \times 10^{13}$ |
| $1.0 \leq z \leq 2.3$ | | | | | | | | |
| RQs | 1.58 | 31102 | $6.61^{+0.80}_{-0.70}$ | $2.13^{+0.10}_{-0.09}$ | 14.70 | 6 | 2.21 ± 0.10 | $1.89^{+0.38}_{-0.34} \times 10^{12}$ |
| RLQs | 1.56 | 2474 | $13.76^{+1.64}_{-1.86}$ | $2.21^{+0.37}_{-0.22}$ | 2.14 | 4 | 4.33 ± 0.57 | $2.01^{+0.84}_{-0.69} \times 10^{13}$ |
| $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ | 1.47 | 9446 | $8.00^{+0.96}_{-1.28}$ | $1.88^{+0.16}_{-0.15}$ | 0.23 | 7 | 2.51 ± 0.34 | $3.98^{+2.33}_{-1.70} \times 10^{12}$ |
| $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ | 1.47 | 9446 | $11.39^{+0.67}_{-0.95}$ | $2.60^{+0.22}_{-0.2}$ | 0.63 | 6 | 3.94 ± 0.32 | $1.79^{+0.46}_{-0.40} \times 10^{13}$ |

in the left panels of Fig. 2.9.

The best-fitting parameters in the interval $0.3 \leq z \leq 2.3$ are $r_0 = 10.95_{-1.58}^{+1.22}$ Mpc, $\gamma = 2.29_{-0.34}^{+0.53}$ for the RLQs and $r_0 = 6.59_{-0.24}^{+0.33} h^{-1}$ Mpc, $\gamma = 2.09_{-0.09}^{+0.10}$ for the RQQs (see Table 2.2). The latter fit is poor with $\chi^2 = 19.60$ and 7 dof, while the former, with the same number of data points, is more acceptable, with $\chi^2 = 1.06$. It is clear from our clustering measurements that RLQs are more strongly clustered than RQQs. The two additional redshift bins show similar trends, with RLQs in the low- z bin clustering more strongly.

In order to check our results, we estimate the correlation function for 100 randomly selected quasar sub-samples chosen from the RQQs with the same number of quasars as RLQs in the corresponding redshift interval. The randomly selected quasar samples present similar clustering lengths to those of RQQs.

We also fit the correlation function over a more restricted range to examine the impact of different distance scales on the clustering measurements. Using $2 \leq r_p \leq 35 h^{-1}$ Mpc, we obtain a model with a somewhat smaller correlation scale-length $r_0 = 6.04_{-0.60}^{+0.51} h^{-1}$ Mpc and a flatter slope $\gamma = 1.72_{-0.10}^{+0.10}$ for RQQs in the full sample. The model matches the data better, resulting in $\chi^2 = 1.06$ and 4 dof. This may signal a change in the TPCF with scale; the transition between the one-halo and two-halo terms may be responsible for the $w_p(r_p)$ distortion on smaller scales (e.g., Porciani et al. 2004). Our remaining non-radio samples show a similar trend of improving the fits at smaller distances. For RLQs, we obtain $(r_0, \gamma) = (9.75_{-1.60}^{+1.90}, 2.70_{-0.60}^{+0.50})$ with $\chi^2 = 2.77$ and 4 dof. The changes in the parameters are within the error bars.

We use the virial BH mass estimations based on single-epoch spectra to investigate whether or not quasar clustering depends on BH mass. The emission line which is employed to determine the fiducial virial mass depends on the redshift interval (see Shen et al. 2008 for a description).

First, we divide the quasar samples using the median virial BH mass in redshift intervals of $\Delta z = 0.05$ following Shen et al. (2009). Although this approach yields samples with comparable redshift distributions, it mixes quasars regardless of their luminosity and could wash out any true dependence on M_{BH} . Indeed, the mass samples following

this scheme hardly show any significant differences in their clustering with correlation lengths similar to those of RQQs. Thus, we proceed to create mass samples with two M_{BH} intervals: $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ and $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$, as described in Sec. 2.2.4. The right-hand panels in Figure 2.9 show $w_p(r_p)$ for these BH mass-selected samples. It can be seen that quasars with higher BH masses have stronger clustering. For $0.3 \leq z \leq 2.3$, we obtain $r_0 = 8.535^{+1.57}_{-2.25} h^{-1} \text{ Mpc}$, $\gamma = 1.84^{+0.21}_{-0.20}$ for quasars with $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$; and $r_0 = 10.45^{+0.79}_{-0.98} h^{-1} \text{ Mpc}$, $\gamma = 2.36^{+0.18}_{-0.17}$ for BH masses in the range $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$. In the other z -bins, the resulting trend is similar, with the low- z bin showing the larger clustering amplitudes. These trends hold when the distance is restricted to $2 \leq r_p \leq 35 h^{-1} \text{ Mpc}$, with no significant variations in r_0 and γ due to the larger uncertainties at these scales.

2.4.2 Quasar bias factors

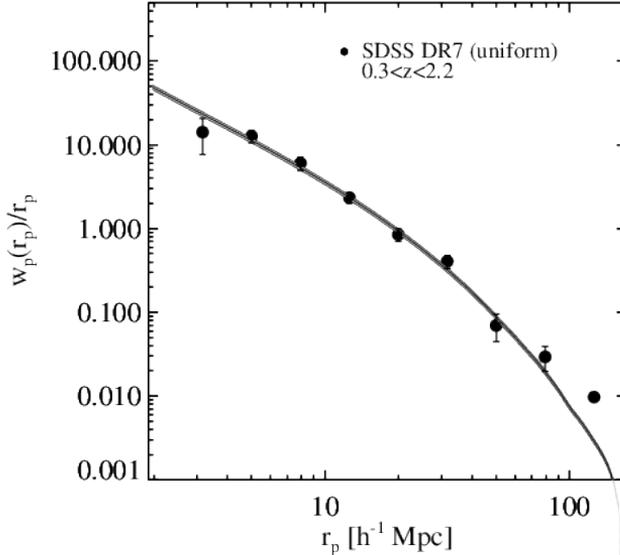


Figure 2.8: Real-space correlation function for the SDSS DR7 quasar uniform sample with $0.3 \leq z \leq 2.3$. **The solid line denotes the model $w_{\text{QSO}}(r_p)$ defined in eq. 2.8 and the shaded areas are the $1 - \sigma$ uncertainties.** Errors bars are the square root of the diagonal elements from the covariance matrix computed using the jackknife method.

We compute the quasar bias factors over the scales $2.0 \leq r_p \leq 130 h^{-1} \text{ Mpc}$ using the $w_{\text{DM}}(r_p)$ model in eq. (2.8). Again, this distance scale has been chosen to have a good overlap with previous SDSS quasar clustering studies (e.g., Shen et al. 2009; Ross et al. 2009). The best-fit bias values and the corresponding typical DMH masses for quasar samples are shown in Table 2.2. We find that the SDSS DR7 quasars at $\bar{z} = 1.30$ (Figure 2.8) have a bias of $b = 2.00 \pm 0.08$. Previous bias estimates from 2QZ (Croom et al. 2005) and 2SLAQ (da Ângela et al. 2008) surveys are consistent with our results within the 1σ error bars.

The left panel on Figure 2.9 compares the projected real-space TPCF w_p/r_p for the RLQs (red) and RQQs (blue). Optically selected quasars are significantly less clustered than radio quasars in the three redshift bins analyzed, which implies that they are less biased objects. Indeed, the RLQs and RQQs, with mean redshifts of $\bar{z} = 1.20$ and $\bar{z} = 1.28$, have bias equivalent to $b = 3.14 \pm 0.34$ and $b = 2.01 \pm 0.08$, respectively. These bias factors correspond to typical DMH masses of $1.23_{-0.39}^{+0.47} \times 10^{13} h^{-1} M_{\odot}$ and $2.38_{-0.38}^{+0.42} \times 10^{12} h^{-1} M_{\odot}$, respectively. We obtain similar results for RQQs in the other two redshift bins with $\bar{z} = 0.65$ and $\bar{z} = 1.58$, respectively, (see Table 2.2). There are considerable differences between the low- z and high- z bins results for RLQs, with low- z RLQs residing in more massive haloes with masses of $1.16_{-0.33}^{+0.37} \times 10^{14} h^{-1} M_{\odot}$.

The projected correlation functions for the mass samples are shown in Fig. 2.9 (right panels), and the corresponding best-fit bias parameters are reported in Table 2.2. We find $b = 2.64 \pm 0.42$ for quasars with $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$, and $b = 2.99 \pm 0.43$ for the objects with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ in the full redshift interval. There is a clear trend: the quasars powered by the most massive BHs are more clustered than quasars with less massive BHs. These quasars are more biased than RQQs, but less than radio quasars. In the other z -bins, the b values are comparable to those of the full sample. This implies larger halo masses for the low- z quasars.

We also estimate the bias over $2.0 \leq r_p \leq 35 h^{-1} \text{ Mpc}$. RQQs in the three bins show hardly almost no difference within the uncertainties. The resulting bias for RLQs is $b = 3.11 \pm 0.42$ at $0.3 \leq z \leq 2.2$, which is approximately 1% smaller in comparison

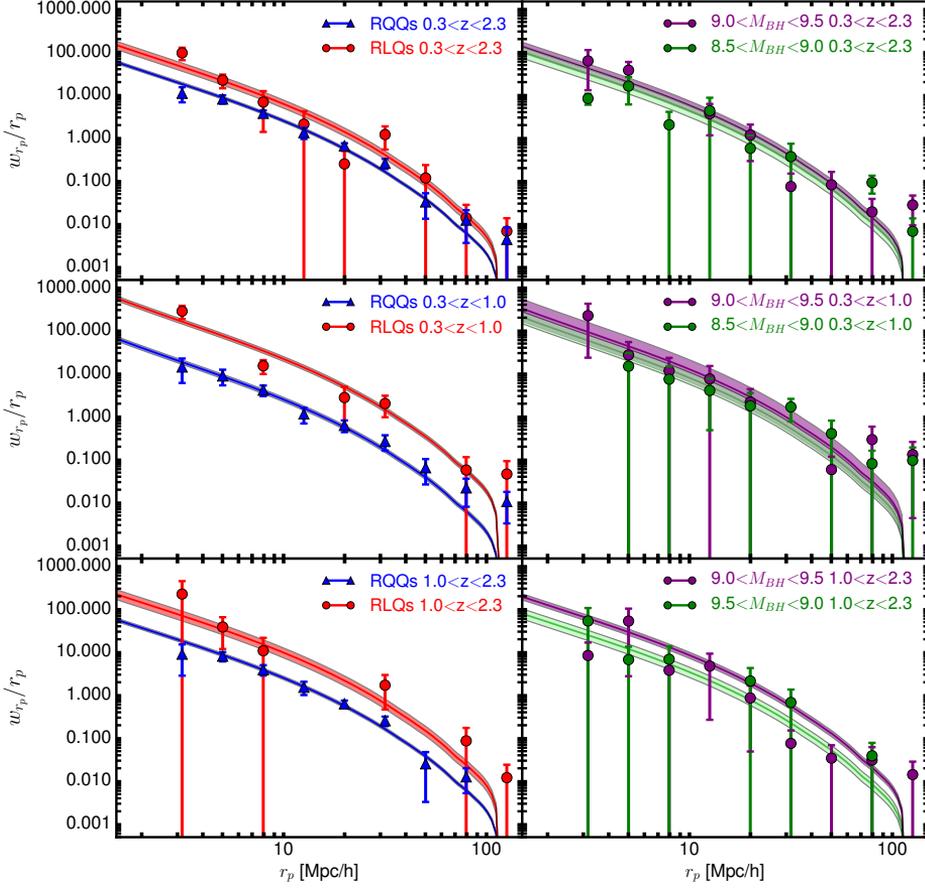


Figure 2.9: Projected correlation functions for the radio-loudness (left) and BH mass (right) samples corresponding to the redshift intervals defined in Table 2.2. The thin lines in each panel represent the term $b^2 w_{\text{DM}}(r_p)/r_p$ for each sample, where the shaded areas correspond to the $1 - \sigma$ errors in the bias factor.

to the bias at $2.0 \leq r_p \leq 130 h^{-1} \text{ Mpc}$. Therefore, restricting the bias does not affect our conclusions for the radio samples. For the mass samples, they remain virtually the same when the range is restricted.

2.4.3 Bias and host halo mass redshift evolution

In Figure 2.10 (left panel), we show our bias estimates for RQQs and RLQs (red and gray triangles, respectively). It can be seen that the bias is a strong function of redshift. In the same plot, we show the previous bias estimates from the optical spectroscopic quasar samples (gray symbols) as well as radio-loud AGNs (green and orange symbols). Our estimates for both RQQs and RLQs are consistent with previous works. The expected redshift evolution tracks of DMH masses based on the models from Sheth et al. (2001) are shown by dashed lines in Fig. 2.10. RQQs follow a track of constant mass a few times $10^{12} h^{-1} M_{\odot}$, while the majority of RLQs and radio sources approximately follow a track of $\sim 10^{14.0-13.5} h^{-1} M_{\odot}$ within the error bars.

2.4.4 Clustering as a function of radio-loudness

Even though the number of radio sources is only $\sim 7.6\%$ of the total number of quasars, it is clear from the left-hand panels of Fig. 2.9 that RLQs are considerably more clustered than RQQs in all the redshift bins. The stronger clustering presented by RLQs suggests that these inhabit more massive haloes than their radio-quiet counterparts. The RLQs typical halo mass of $> 1 \times 10^{13} h^{-1} M_{\odot}$ is characteristic of galaxy groups and small clusters, while the typical mass of a few times $10^{12} h^{-1} M_{\odot}$ for RQQs is typical of galactic haloes. The higher DMH mass presented by RLQs in the low- z bin is similar to the halo mass of galaxy clusters, which is usually $> 1 \times 10^{14} h^{-1} M_{\odot}$.

The right-hand panel in Fig. 2.10 presents the DMH masses against redshift for the same samples as in the left-hand panel. Our new mass estimates for RLQs and RQQs are generally consistent with those derived in previous works (e.g., Croom et al. 2005; Porciani & Norberg 2006; Ross et al. 2009; Shen et al. 2009). We denote the typical halo masses for the two quasar populations using dashed lines. This suggests that the

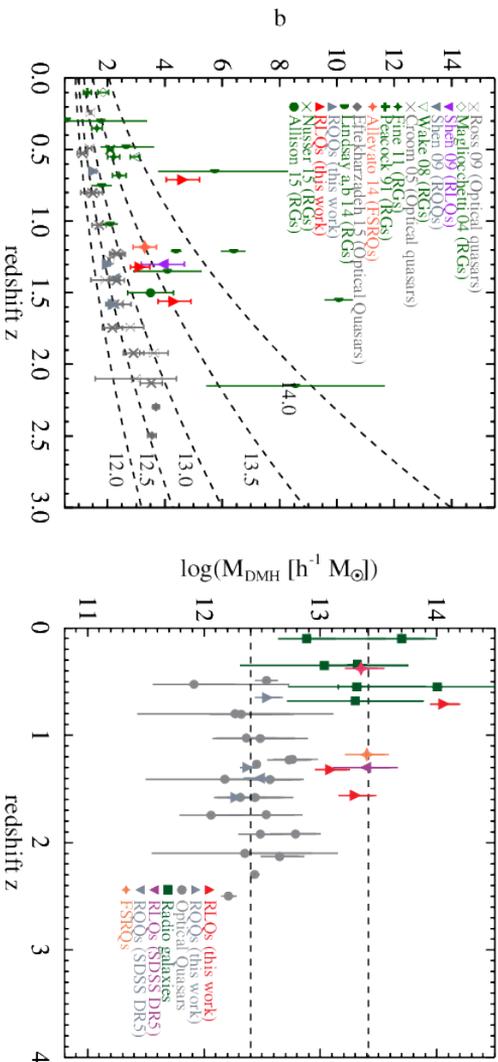


Figure 2.10: Left: The **derived linear bias** parameter b as a function of redshift for radio and optical AGN samples represented by the corresponding legend. Red or gray triangles represent the RLQs or RQs sub-samples of this work, respectively. The dashed lines denote the expected redshift evolution of DMH masses based on the models from Sheth et al. (2001) with $\log(M_{\text{DMH}}/h^{-1}M_{\odot}) = [12.0, 13.0, 13.5, 14.0]$. Right: Typical DMH masses M_{DMH} against redshift for RLQs and RQs from our sample (red and gray triangles, respectively), RLQs and RQs from SDSS DR5 (purple and gray downward triangles, Shen et al. 2009, respectively), optical quasars (gray circle, Croom et al. 2005; Ross et al. 2009; Eftekharzadeh et al. 2015), radio galaxies (RGs, dark green squares, Peacock & Nicholson 1991; Magliocchetti et al. 2004; Wake et al. 2008; Fine et al. 2011; Lindsay et al. 2014b,a; Allison et al. 2015; Nusser & Tiwari 2015), and flat-spectrum radio quasars (FSRQs) (orange star, Allevato et al. 2014a). For comparison, we show with dashed lines the mass values corresponding to $\log(M_{\text{DMH}}/h^{-1}M_{\odot}) = [12.4, 13.4]$. When bias and mass estimations are not provided by the authors we use the reported power-law best-fitting values to estimate b and M_{DMH} (Peebles 1980; Krumpke et al. 2010).

difference between the typical host halo masses for RLQs and RQQs is constant with redshift, with the haloes hosting RLQs being approximately one order of magnitude more massive.

2.4.5 Clustering as function of BH masses

Our clustering measurements for the $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ and $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ show a clear dependence on virial BH masses. This trend is apparent in Fig. 2.9 (right panels) for all the redshift bins considered. Moreover, this is reflected in our M_{BH} predictions for the mass samples in Figure 2.11. The quasars powered by SMBHs with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ present larger clustering amplitudes than those with less massive BH masses in the range $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$. Table 2.2 indicates that both RLQs and the quasars with BH masses of $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ have larger correlation lengths than RQQs and quasars with $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$. However, RLQ clustering is at least slightly stronger in all the redshift bins analyzed. It is important to remark that the use of virial estimators to calculate the BH masses is subject to large uncertainties (e.g., Shen et al. 2008; Shen & Liu 2012; Assef et al. 2012) leading to significant biases and scatter around the true BH mass values, which could potentially weaken any clustering dependence on BH mass. Nevertheless, our results give some validity to their use in clustering analyses.

Fig. 2.11 shows the redshift evolution of the ratio between the DMH and the average virial BH masses for our quasar samples. The different lines mark the ratio for each quasar sample denoted by the plot legend. The ratios reproduce the trend for the clustering amplitudes in all the samples: RLQs and quasars with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ cluster more strongly than RQQs and quasars with $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$, respectively. Quasars with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ present clustering comparable to RLQs. Also, it is evident that the ratios are larger at low- z due to the host haloes being more massive and the virial BH masses showing no significant changes with redshift (see Table 2.1).

An important point to consider is the cause of stronger clustering: is the stronger clustering for the high-mass quasars due to the fact that they are radio loud, or are the RLQs more clustered due to the fact that they have higher BH masses. We can

address this by examining the distribution of RLQs on the virial BH mass plane. This distribution is not restricted to high BH masses only. Instead, RLQs present BH masses in all the ranges sampled, indicating that their radio-emission rather than high BH mass is responsible for the stronger clustering in RLQs. However, for the high-mass sample only a fraction of $\sim 6\%$ is radio-loud, which translates to approximately 700 RLQs, which is not large enough to obtain a reliable clustering signal. For the high-mass sample minus the radio-quasars, we do obtain a clustering amplitude similar to those including radio objects. Therefore, we conclude that the stronger clustering for both samples is mainly due to the intrinsic properties of each sample. This point needs to be addressed using forthcoming quasar samples with higher quasar numbers.

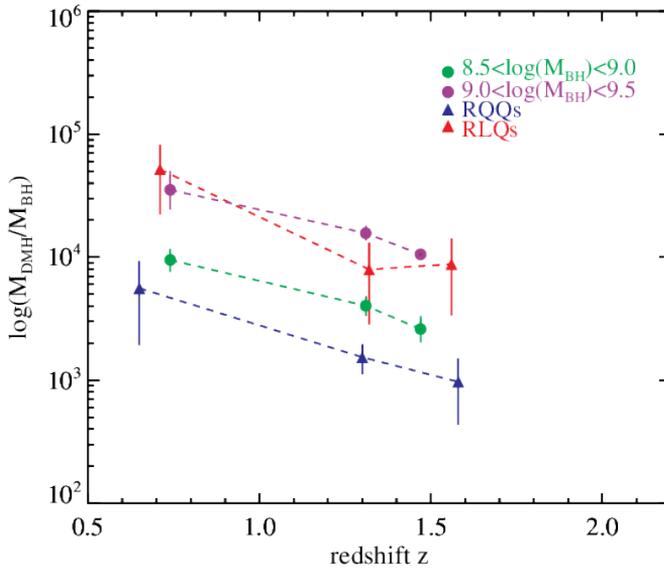


Figure 2.11: **Ratio between the DMH and the average virial BH masses for our quasar samples as a function of redshift.**

2.4.6 Clustering as a function of redshift

In Fig. 2.12, we show our r_0 measurements along with results from previous works for radio galaxies (Peacock & Nicholson 1991; Magliocchetti et al. 2004; Wake et al.

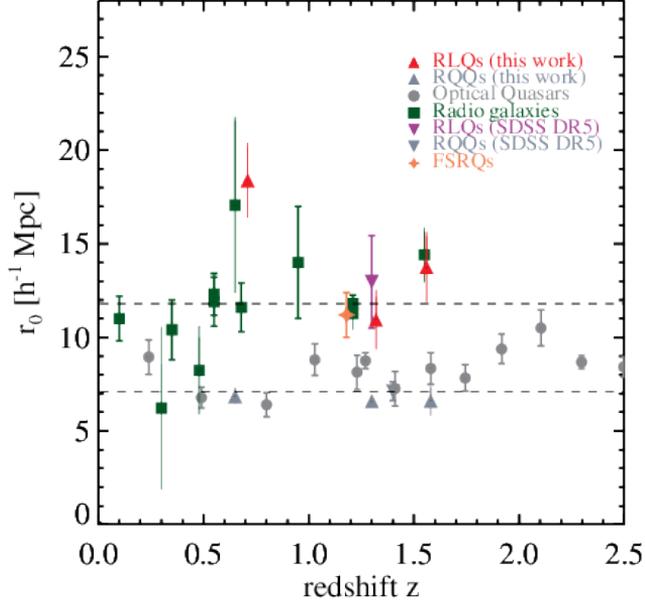


Figure 2.12: Different values for the real-space correlation length r_0 against redshift for RLQs and RQs from SDSS DR5 (purple and gray downward triangles, Shen et al. 2009), optical quasars (gray circles, Croom et al. 2005; Ross et al. 2009; Eftekharzadeh et al. 2015), radio galaxies (dark green squares, Peacock & Nicholson 1991; Magliocchetti et al. 2004; Wake et al. 2008; Fine et al. 2011; Lindsay et al. 2014b), and FSRQs (orange star, Allevato et al. 2014a). The r_0 values for RLQ and RRQ in our sample are represented by red and gray upward triangles, respectively. For comparison, we show the r_0 values corresponding to $r_0 = [11.8, 7.1] h^{-1} \text{ Mpc}$ (dashed lines). The results from Lindsay et al. (2014b) are derived assuming linear clustering.

2008; Fine et al. 2011; Lindsay et al. 2014b; Allison et al. 2015; Nusser & Tiwari 2015), optically-selected quasars (Ross et al. 2009; Croom et al. 2005; Eftekharzadeh et al. 2015), and γ -selected blazars (Allevato et al. 2014a). In these samples, the typical 1.4 GHz radio-luminosities for AGNs is 10^{23} - $10^{26} \text{ W Hz}^{-1}$ which is representative of FRI sources, whilst for our sample the average radio-luminosity is $\sim 8 \times 10^{26} \text{ W Hz}^{-1}$, which is near the boundary between FRI and FRII sources.

A systematic trend with redshift is observed in Fig. 2.12, which indicates that the majority of radio sources considered have clustering lengths over the entire redshift range considered ($0 < z < 2.3$). This is consistent with the trend from Fig. 2.10, where the majority of radio sources seem to inhabit haloes of $M_{\text{DMH}} > 1 \times 10^{13}$ at

all redshifts. The simplest interpretation of this result is that a considerable part of the bright radio population resides in massive haloes with large correlation lengths. Our new RLQ clustering measurements for the full sample and high- z bin agree, within the errors bars, with the previous single estimation from Shen et al. (2009) using the SDSS DR5 quasar sample, while the low- z bin correlation amplitude is consistent with Lindsay et al. (2014b).

Overzier et al. (2003) measured the angular TPCF for the NVSS survey (Condon et al. 1998) and concluded that lower luminosity radio sources ($\leq 10^{26} \text{ W Hz}^{-1}$) present typical correlation lengths of $r_0 \lesssim 6 h^{-1} \text{ Mpc}$, whilst the brighter radio sources ($> 10^{26} \text{ W Hz}^{-1}$), mainly FR II type, have significantly larger scale lengths of $r_0 \gtrsim 14 h^{-1} \text{ Mpc}$. Our findings are consistent with Overzier et al. (2003) predictions for the bright radio population. It is possible that the weaker correlation length presented by lower radio-luminosity samples in Fig. 2.12 indicates a mild clustering dependence on radio-luminosity. However, our RLQs sample is still too small to draw firm conclusions on the radio luminosity dependence as the increasing errors for these luminosity-limited samples mean we cannot satisfactorily distinguish between them

The DMH masses for RLQs and quasars with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ at $0.3 < z < 1.0$, are approximately $> 1 \times 10^{14} h^{-1} M_{\odot}$, which is the typical value for cluster-size haloes. Moreover, these halo masses are larger than the corresponding haloes for quasar samples at $z > 1.0$. This suggests that the environments in which these objects reside is different from those of their high- z counterparts. Additionally, the radio source clustering amplitudes are similar to the clustering scale of massive galaxy clusters (e.g., Bahcall et al. 2003). This almost certainly reveals a connection between quasar radio-emission and galaxy cluster formation that must be explored in detail with data from forthcoming radio surveys.

2.4.7 Clustering and AGN unification theories

Our clustering results hint at an interesting point regarding the relationship between RLQs and radio galaxies in AGN classifications, which consider these AGNs as the same source type seen from different angles (e.g., Urry & Padovani 1995). Thus, we would

expect that different AGN types such as radio galaxies and RLQs, should have similar clustering properties. The real-space correlation lengths for RLQs (red triangles) and other radio sources including, radio galaxies (green squares), are shown in Fig. 2.12. We see that there is a reasonable consistency for most r_0 values up to $z \lesssim 2.3$. We identify the same trend in Fig. 2.10 (right panel), where bright radio sources seem to inhabit haloes of approximately constant mass of $\gtrsim 10^{13.5} h^{-1} M_\odot$. Our clustering study seems to support the validity of unification models at least for RLQs and radio galaxies with relatively median radio-luminosities ($\gtrsim 1 \times 10^{23} \text{ W Hz}^{-1}$).

Allevato et al. (2014a) studied the clustering properties of a γ -selected sample of blazars divided into BL Lacs and flat-spectrum radio quasars (FSRQs). In the context of unification models, FSRQs are associated with intrinsically powerful FRII radio galaxies, while BL Lacs are related to weak FRI radio galaxies. From a clustering point of view, as explained before, luminous blazars should have similar clustering properties to radio galaxies. In Figs. 2.10 and 2.12, we denote by a orange star, the DMH mass and correlation length for FSRQs, respectively, found by Allevato et al. (2014b). FSRQs show a similar M_{DMH} value to those of radio galaxies and RLQs, supporting a scenario in which radio AGNs such as quasars, radio galaxies and powerful blazars are similar from a clustering perspective and reside in massive hosting haloes providing the ideal place to fuel the most massive and powerful BHs.

Based on an analysis of the cross-correlation function for radio galaxies, RLQs and a reference sample of luminous red galaxies Donoso et al. (2010) concluded that the clustering for RLQs is weaker in comparison with radio galaxies. This is apparently at odds with previous clustering measurements and our results. However, there are several differences between Donoso's and our sample that must be considered. First, Donoso's sample is significantly smaller with only 307 RLQs at $0.35 < z < 0.78$. Secondly, in the common range between the two samples where the TPCF is computed, their clustering signal has large uncertainties. Thirdly, they compute the clustering for objects with radio-luminosities restricted to $> 10^{25} \text{ W Hz}^{-1}$. We employ the same luminosity cut only for the high- z bin, while for the low- z bin only sources brighter than $> 4 \times 10^{24} \text{ W Hz}^{-1}$ are considered. The mean luminosity for both redshift bins is

$> 2 \times 10^{26} \text{ W Hz}^{-1}$ (see Table 2.1). Therefore, comparable radio-luminosity cuts were used for both samples. For these reasons, it is difficult to draw any conclusions from comparison with the Donoso results.

2.4.8 The role of mergers in quasar radio-activity

We compare our clustering measurements with the theoretical framework for the growth and evolution of SMBHs introduced by Shen (2009). This model links the quasar properties and host halo mass with quasar activity being triggered by major galaxy mergers. The bias factor is a function of the instantaneous luminosity and redshift, with most luminous quasars having larger host-halo masses. The rate of quasar activity is controlled by the fraction parameter f_{QSO} , which involves exponential cutoffs at both high and low mass ends assigned according to phenomenological rules. At low masses, the cutoffs prevent quasar activity on the smallest postmerger haloes, while those at the highest masses cause that gas accretion to become inefficient and subsequent BH growth stops. Figure 2.13 presents the predicted linear bias as a function of bolometric luminosity at $z = 0.65$ (left) and $z = 1.40$ (right). In the low- z bin ($0.3 \leq z \leq 1.0$), the model can reproduce the bias for the RQQs. However, the quasar merger model disagrees with the higher bias value for RLQs. At high- z ($1.0 \leq z \leq 2.3$), the consistency between the model predictions and the measured bias for RQQs for the high- z bin and the complete quasar sample worsens. The bias luminosity-dependent trend predicted by the model seems to be followed slightly better by the RLQs than in the low- z bin.

The discrepancy between the merger-driven model predictions and our bias values might indicate differences in the fueling channels for both quasar types. First, our bias estimates for RQQs in the context of the Shen et al. (2009) framework favor accretion of cold gas via galaxy mergers (referred to as cold-gas accretion). These M_{DH} masses are in agreement with the halo mass-scale of a few times $\gtrsim 10^{12} h^{-1} M_{\odot}$ predicted by merger-driven models for optical quasars (e.g., Croom et al. 2005; Ross et al. 2009). In contrast, the bias results for RLQs, which correspond to halo masses of $\gtrsim 10^{13} h^{-1} M_{\odot}$, cannot be reproduced by models that assume that quasar activity is solely triggered by

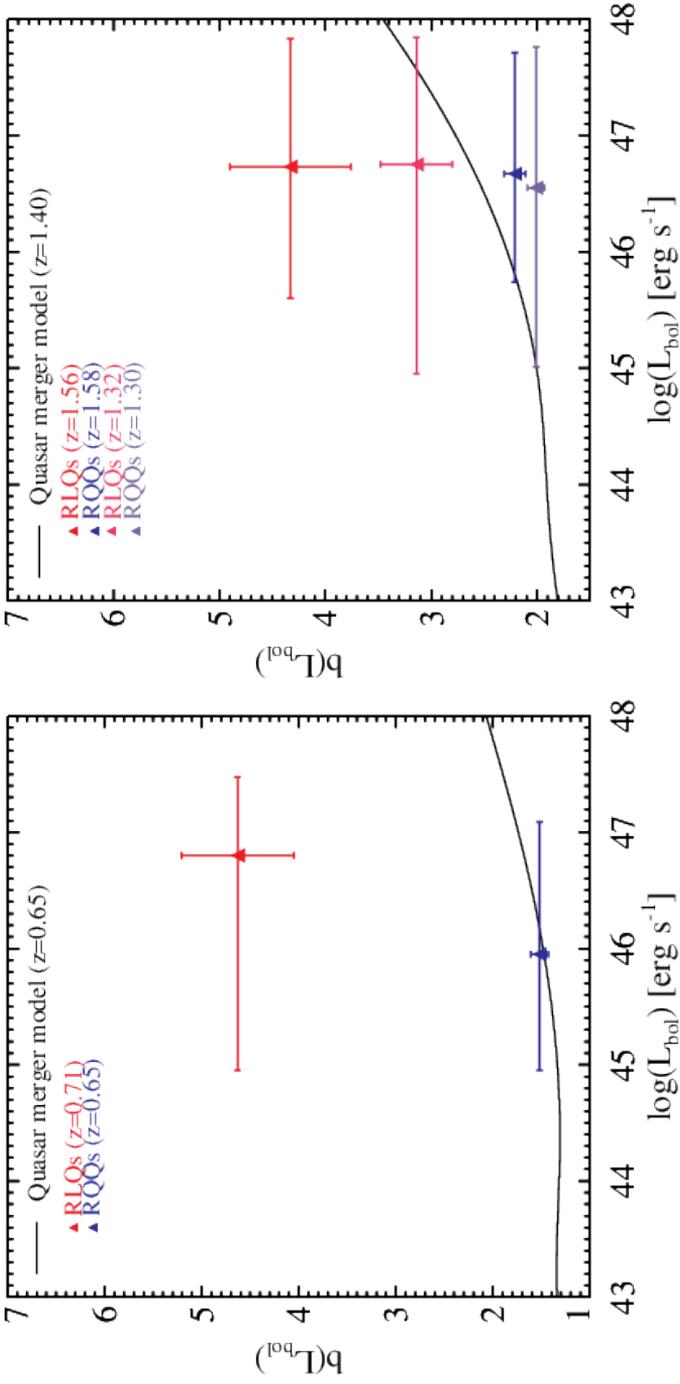


Figure 2.13: Bias parameter b as a function of bolometric luminosity for our RLQs and RQQs in the ranges $0.3 \leq z \leq 1.0$ (left) and $1.0 \leq z \leq 2.0$ (right). Errors in the L_{bol} axis are the dispersion values for each different quasar sample. The solid lines in both panels denote the predicted bias luminosity evolution according to the Shen (2009) model, which predicts that quasar activity is triggered by galaxy mergers.

typical galaxy mergers.

A similar difference in DMH masses has been reported in clustering studies for X-ray selected AGNs with moderate luminosity ($L_{bol} \sim 10^{43-46} \text{ erg s}^{-1}$) (Gilli et al. 2005, 2009; Starikova et al. 2011; Allevato et al. 2011; Mountrichas et al. 2013; Mountrichas & Georgakakis 2012). The DMH masses of X-Ray AGNs are approximately $10^{13} h^{-1} M_{\odot}$, which is significantly higher in comparison with relatively bright optical quasars ($L_{bol} \gtrsim 10^{46} \text{ erg s}^{-1}$) with $\gtrsim 10^{12} h^{-1} M_{\odot}$ (Croom et al. 2005; Ross et al. 2009). Several authors have observationally (Allevato et al. 2011; Mountrichas & Georgakakis 2012; Allevato et al. 2014b) and theoretically (Fanidakis et al. 2012, 2013a) interpreted these two mass scales as evidence favoring different accretion channels for each AGN population. Fanidakis et al. (2013a), using semi-analytical galaxy formation models, found that cold gas fuelling cannot reproduce the DMH masses from X-Ray AGN clustering studies. Instead, they found that when gas cooled from quasi-hydrostatic hot-gas haloes (i.e., known as hot-mode; Croton et al. 2006) is included, a much better agreement with the DMH masses derived from X-Ray AGN clustering studies is obtained.

The differences in DMH masses for X-Ray AGNs and optical quasars is reminiscent of our results for RQQs and RLQs. This may suggest that the contribution of hot-gas accretion increases for more massive haloes, such as those hosting X-Ray AGNs and RLQs. However, this scenario for RLQs still needs to be confronted with more detailed simulations and models to further constrain the physics of BH accretion.

2.4.9 Black hole properties involved in quasar triggering

As considering only cold accretion via mergers cannot explain the mass scales associated with RQQs and RLQs, it is important to take into account different mechanisms related to quasar activity. For instance, the massive haloes where these RLQs are embedded must have an important role in determining the BH properties and the onset of radio activity. Indeed, the BH spin could be altered by environmental conditions: either by means of coherent gas accretion, or by BH-BH mergers. In the spin paradigm proposed by Wilson & Colbert (1995), the rapidly spinning BHs are associated with radio-loud

AGNs, whilst the slower spinning ones are considered to be radio-quiet. Objects above a certain spin threshold could have the necessary energy to produce powerful relativistic jets (Blandford & Znajek 1977). The intrinsic scatter on the BH spin values required to power the jets may reproduce the different morphologies and the shape of the luminosity function at radio wavelengths (Fanidakis et al. 2011). Another plausible scenario is a two-way interaction between RLQs jets and the surrounding intergalactic medium, as suggested by the morphological associations of radio continuum with extended optical emission (van Breugel et al. 1985), and bent radio structures in nearby radio active galaxies (O’Dea & Owen 1986). As radio jets propagate into a dense interstellar medium they suffer from both depolarization and decollimation that yield an enhancement in their radio brightness (Begelman et al. 1984). The luminosity boosting for these objects may help to make them just bright enough to be detectable above the FIRST survey flux limit. Finally, the magnetic field configurations derived from polarimetry studies (e.g., Bridle & Perley 1984) indicate that the magnetic field in FR-II radio-galaxies is predominantly aligned along the jet for most of its length, whereas FR-I objects are characterized by perpendicular and parallel components. This may suggest a correlation between the DMH mass and the efficiency in producing the magnetic field alignment required to produce brighter radio emission.

In conclusion, the interplay between all these BH properties in triggering radio activity is still poorly understood. Additional observational and theoretical efforts are required to obtain a better comprehension of the origins of radio-emission in quasars.

2.5 Summary

In this study, we have investigated the quasar clustering dependence on radio-loudness and BH virial mass, by using a sample of approximately 48000 spectroscopically confirmed quasars at $0.3 \leq z \leq 2.3$ drawn from SDSS DR7 quasar catalog (Shen et al. 2011; Schneider et al. 2010). Our radio sample consists of FIRST-detected quasars. The main conclusions of this paper are the following:

1. We studied the spatial clustering of quasars at $0.3 \leq z \leq 2.3$ over the scales

$2.0 \leq r_0 \leq 130 h^{-1}$ Mpc. For RQQs, we find a real-space correlation length equal to $r_0 = 6.59_{-0.24}^{+0.33} h^{-1}$ Mpc with a slope of $\gamma = 2.09_{-0.09}^{+0.10}$. RLQs are more strongly clustered than RQQs with $r_0 = 10.95_{-1.58}^{+1.22} h^{-1}$ Mpc, $\gamma = 2.29_{-0.34}^{+0.53}$.

2. We estimated the linear bias for RQQs and RLQs by splitting the quasar sample according to radio-loudness, and find $b = 2.01 \pm 0.08$ and $b = 3.14 \pm 0.34$, respectively, for the full redshift interval.
3. We investigated the clustering dependency on BH virial mass using quasar samples with $8.5 \leq \log(M_{\text{BH}}) \leq 9.0$ and $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ constructed to have comparable optical luminosity distributions. We find a dependence on BH mass, with the quasars powered by the most massive BHs having larger correlation lengths. These results suggest that BH virial mass estimations based on broad emission lines may be valid BH mass proxies for clustering studies.
4. Using our best-fit bias values, we find that RLQs in our sample inhabit massive haloes with masses of $M_{\text{DMH}} \gtrsim 10^{13.5} h^{-1} M_{\odot}$ at all redshifts, which corresponds to the mass scale of galaxy groups and galaxy clusters. RQQs reside in less massive haloes of a few times $\sim 10^{12} h^{-1} M_{\odot}$.
5. RQQs have smaller DMH masses in comparison with RLQs. The BH mass selected samples have larger DMH masses than RQQs, but smaller DMH masses than those of radio quasars. However, RLQs have the most massive DMHs in all the redshift bins considered. We considered the ratio between the DMH and average virial BH masses for all the samples. The ratios present the same above-mentioned trends.
6. Within our quasar sample, we do detect significant correlations between quasar clustering and redshift for RLQs up to $z \lesssim 2.3$. At low- z , RLQs and quasars with $9.0 \leq \log(M_{\text{BH}}) \leq 9.5$ have clustering amplitudes of $r_0 \gtrsim 18 h^{-1}$ Mpc, comparable to those of today's massive galaxy clusters. Our real-space clustering length r_0 estimate for the full samples agrees very well with the majority of previous complementary and independent clustering estimates for radio galaxies and RLQs.

7. We used radio-loudness to separate the quasar sample into RLQs and RQQs. Our clustering measurements suggest that there are differences between RLQs and RQQs in terms of halo and BH mass scales. Our result is consistent with the hierarchical clustering scenario, in which most massive galaxies harboring the most massive BHs form in the highest density peaks, thus cluster more strongly than less massive galaxies in typical peaks. This is confirmed by clustering analysis of the mass samples and their dependence on M_{BH} .
8. Comparing our linear bias and DMH mass estimates with the theoretical predictions of the merger-driven model from Shen (2009), we find that this model cannot explain the larger bias and DHM masses for RLQs, suggesting that cold accretion driven by galaxy mergers is unlikely to be the main fueling channel for RLQs with $M_{\text{DMH}} \geq 10^{13} h^{-1} M_{\odot}$. Conversely, merger model predictions agree well with our bias and host mass estimates for RQQs, with $M_{\text{DMH}} \gtrsim 10^{12} h^{-1} M_{\odot}$.
9. The disagreement between the bias luminosity-dependent trend predicted by the Shen (2009) merger model and our bias estimates for RLQs suggests a scenario where the radio emission is a complex phenomenon that may depend on several BH properties such as: BH spin, environment, magnetic field configuration, and accretion physics.
10. The similarity in clustering amplitude and host halo masses for radio-galaxies, radio-selected AGNs, RLQs, and FSRQs is in line with the idea that the different spectral features for these radio sources depend only on the orientation angle and not on the environment in which they are embedded, supporting orientation-driven unification models (Urry & Padovani 1995 and references therein). Donoso et al. (2010) found that the clustering properties for RLQs and radio galaxies differ, with the latter displaying a stronger clustering. In principle, these results are in tension with our results and previous clustering studies of radio sources (e.g. Magliocchetti et al. 2002; Wake et al. 2008; Shen et al. 2009). However, their small sample size and large uncertainties in the clustering in comparison with our sample make it difficult to draw any significant conclusions. In future studies,

larger samples of quasars and radio galaxies may provide new information about the clustering properties for both populations.

2.6 Acknowledgements

ERM wish to thank F. Mernier and E. Rigby for critical reading and F. Shankar for providing us the tracks in Fig. 13.

ERM acknowledges financial support from NWO Top project, No. 614.001.006. HR acknowledges support from the ERC Advanced Investigator program NewClusters 321271. Moreover, we would like to thank the referee for valuable suggestions on the manuscript. Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

Chapter 3

Deep LOFAR 150 MHz imaging of the Boötes field: Unveiling the faint low-frequency sky

Abstract

We have conducted a deep survey (with a central rms of $55\mu\text{Jy}$) with the LOw Frequency ARray (LOFAR) at 120-168 MHz of the Boötes field, with an angular resolution of $3.98'' \times 6.45''$, and obtained a sample of 10091 radio sources (5σ limit) over an area of 20 deg^2 . The astrometry and flux scale accuracy of our source catalog is investigated. The resolution bias, incompleteness and other systematic effects that could affect our source counts are discussed and accounted for. The derived 150 MHz source counts present a flattening below sub-mJy flux densities, that is in agreement with previous results from high- and low- frequency surveys. This flattening has been argued to

Retana-Montenegro and Röttgering, and 6 co-authors A&A, 620, A74 (2018)

be due to an increasing contribution of star-forming galaxies and faint active galactic nuclei. Additionally, we use our observations to evaluate the contribution of cosmic variance to the scatter in source counts measurements. The latter is achieved by dividing our Boötes mosaic into 10 non-overlapping circular sectors, each one with an approximate area of 2 deg^2 . The counts in each sector are computed in the same way as done for the entire mosaic. By comparing the induced scatter with that of counts obtained from depth observations scaled to 150MHz, we find that the 1σ scatter due to cosmic variance is larger than the Poissonian errors of the source counts, and it may explain the dispersion from previously reported depth source counts at flux densities $S < 1 \text{ mJy}$. This work demonstrates the feasibility of achieving deep radio imaging at low-frequencies with LOFAR.

3.1 Introduction

The most luminous radio sources are often associated with radio-loud active galactic nuclei (AGN) powered by accretion onto supermassive black holes (SMBHs), whose radio emission is generated by the conversion of potential energy into electromagnetic energy released as synchrotron radiation and manifesting itself as large-scale structures (radio jets and lobes). The less luminous radio-selected objects are mostly associated with accreting systems like radio-quiet AGNs or starburst galaxies. The radio-emission in star-forming systems has two components: a non-thermal synchrotronic component produced by cosmic rays originating from supernova shockwaves, and a thermal free-free component arising from the interstellar medium ionization by hot massive stars (Condon 1992). Star formation is also thought to be responsible at least for a fraction of radio emission in radio-quiet AGNs. (Padovani et al. 2011; Condon et al. 2012).

In recent years, many studies have confirmed a flattening in the (Euclidean normalized) radio counts below a few mJy (Smolčić et al. 2008; Padovani et al. 2009) first detected more than three decades ago (Windhorst et al. 1985; Kellermann et al. 1986). This flattening is due to an increasing contribution of faint radio sources at sub-mJy flux densities. The precise fraction associated with different objects is still under debate,

with studies showing a mixture of ellipticals, dwarf galaxies, high- z AGNs, and starburst galaxies (Padovani 2011; Smolčić et al. 2017a). The plethora of objects found suggests a complex interplay between star-formation (SF) and AGN activity in the universe.

Additional efforts are important to understand the physical processes that trigger the radio emission of the sub-mJy and microJy sources. Currently, this is partly hampered because the required sensitivity to detect fainter objects have been achieved in only a few small patches of the sky (Schinnerer et al. 2010; Condon et al. 2012; Miller et al. 2013; Vernstrom et al. 2016; Smolčić et al. 2017b).

The majority of deep surveys (Schinnerer et al. 2010; White et al. 2012; Miller et al. 2013; Vernstrom et al. 2016; Smolčić et al. 2017b) have been carried using radio telescopes operating at high-frequencies ($> 1.0\text{GHz}$). This situation is rapidly changing as the number of low-frequency radio surveys ($< 1.0\text{GHz}$) has increased in the last few years. Some survey examples include the VLA Low frequency Sky Survey (VLSS; Cohen et al. 2007), Murchison Widefield Array (MWA) Galactic and Extragalactic All-sky MWA survey (GLEAM; Wayth et al. 2015), and the LOFAR Two-metre Sky Survey (LoTSS, Shimwell et al. 2017a). However, several challenges such as strong radio interference and varying effects like ionospheric phase errors across the instrument field of view (FOV) make producing high-resolution, low-frequency radio maps a difficult task (Noordam 2004). The necessity to overcome these challenges and to fully exploit the science offered by low-frequency telescopes has spurred an invigorated interest by radio-astronomers in improving the low-frequency calibration and imaging techniques (e.g. Cotton et al. 2004; Intema et al. 2009; Kazemi et al. 2011; Smirnov 2011; van Weeren et al. 2016; Tasse et al. 2017).

The LOFAR Surveys Key Science Project (SKSP) is embarking on a survey with three tiers of observations: the LoTSS survey at Tier-1 level covers the largest area at the lowest sensitivity ($\gtrsim 100\ \mu\text{Jy}$) covering the whole 2π steradians of the northern sky. Deeper Tier-2 and Tier-3 programs aim to cover smaller fields with extensive multi-wavelength data up to a depth of tens and a few microJy, respectively (see Röttgering et al. 2011). Together these surveys will open the low-frequency electromagnetic spectrum for exploration, allowing unprecedented studies of the faint radio population across

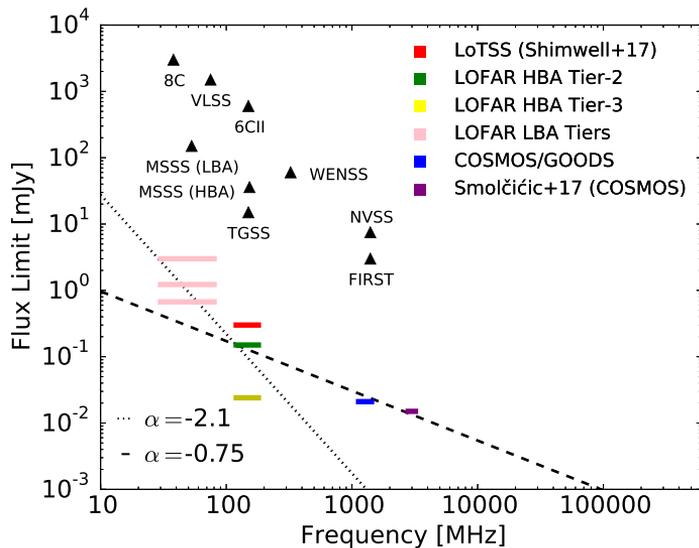


Figure 3.1: Comparison between two radio sources with the same flux, but different spectral indices. The black triangles denote the 5σ flux density limits for previous all-sky shallow low- and high- frequency surveys (Hales et al. 1988; Becker et al. 1995; Condon et al. 1998; Rengelink et al. 1997; Cohen et al. 2007; Heald et al. 2015; Intema et al. 2017), while color bars indicate the 3 different tiers for LOFAR surveys using the LOFAR Low band antennas (LBA) and High band antennas (HBA), and the deepest high-frequency surveys currently published (Schinnerer et al. 2010; Miller et al. 2013; Smolčić et al. 2017b). Sources steeper than $\alpha = -2.1$ will be detected at higher significance in the Tier2/Tier-3 surveys than in deep high-frequency surveys, while sources flatter than $\alpha = -0.75$ at detected at both low and high frequencies.

cosmic time and opening up new parameter space for searches for rare, unusual objects such as high- z quasars (Retana-Montenegro & Röttgering 2018) in a systematic way (see Fig. 3.1).

One of the regions for the Tier-2 and Tier-3 radio-continuum surveys is the Boötes field. This 9.2 deg^2 region is one of the NOAO Deep Wide Field Survey (NDWFS, Jannuzi & Dey 1999) fields, and has a large wealth of multi-wavelength data available including: X-rays (*Chandra*; Kenter et al. 2005), optical ($U_{spec}, B_W, R, I, z, Y$ bands; Jannuzi & Dey 1999; Cool 2007; Bian et al. 2013), infrared (J, H, K bands, *Spitzer*; Autry et al. 2003; Ashby et al. 2009; Jannuzi et al. 2010), and radio (60-1400 MHz; de Vries et al. 2002; Williams et al. 2013; van Weeren et al. 2014; Williams et al. 2016).

In this work, we present deep 150 MHz LOFAR observations of the Boötes field obtained using the facet calibration technique described by van Weeren et al. (2016). The data reduction and analysis for other deep fields using the kMS approach (Tasse 2014; Smirnov & Tasse 2015) and DDFACET imager (Tasse et al. 2017) will be presented in future papers (Mandal in prep.; Sabater in prep.; Tasse in prep.). This paper is structured as follows. In Sections 3.2 and 3.3, we describe the observations and data reduction, respectively. We present our image and source catalog in Section 3.4. We also discuss for the flux density scale, astrometry accuracy, and completeness and reliability. The differential source counts are presented and discussed in Section 3.5. The contribution of cosmic variance to the scatter in source counts measurements is also discussed in Section 3.5. Finally, we summarise our conclusions in Section 3.6. We assume the convection $S_\nu \propto \nu^{-\alpha}$, where ν is the frequency, α is the spectral index, and S_ν is the flux density as function of frequency.

3.2 Observations

The Boötes observations centered at 14h32m00s +34d30m00s (J2000 coordinates) were obtained with the LOFAR High band antenna (HBA). We combine 7 datasets observed from March 2013 (Cycle 0) to October 2015 (Cycle 4), which correspond approximately to a total observing time of 55 hours. When the LOFAR stations operate in the “HBA DUAL INNER” configuration at 150 MHz, LOFAR has a half-power beam width (HPBW) of ~ 5 degrees with an angular resolution of $\sim 5''$ (using only the central and remote stations located in The Netherlands). 3C196 is used as primary flux calibrator and was observed 10 minutes prior to the target observation. The nearby radio-loud quasar 3C295 was selected as secondary flux calibrator, and was observed for 10 minutes after the target. The observations from cycles 0 and 2 consist of 366 subbands covering the range 110-182 MHz. The subbands below 120 MHz and above 167 MHz generally present poor signal-to-noise (S/N). Therefore, in the following cycles, to obtain a more efficient use of the LOFAR bandwidth the frequency range was restricted to 120-167 MHz, resulting in only 243 subbands per observation. The total

time on target varies depending on the cycle. The two observations from Cycle 0 are 5 and 10 hours long, whereas Boötes was observed for 8 hours per observation in Cycles 2 and 4. The frequency and time resolution for the observations varies for each cycle. Table 3.1 presents the details for each one of the observations used in our analysis. Our observations include the dataset L240772 analyzed by Williams et al. (2016).

3.3 Data reduction

In this section, the data reduction steps of the LOFAR data processing are briefly explained. These steps are divided into three stages: the calibration into a non-directional and directional-dependent parts, and the combination of the final calibrated datasets. We refer the reader to the works of van Weeren et al. (2016) and Williams et al. (2016) for a more detailed explanation of the calibration procedure.

3.3.1 Direction independent calibration

First, we start by downloading the unaveraged data from the LOFAR Long Term Archive (LTA)². We follow the basic sequence of steps for the direction-independent (DI) calibration: basic flagging and RFI removal employing `A0f1agger` (Offringa et al. 2010, 2012); flagging of the contributing flux associated to bright off-axis sources referred as the A-team (Cyg A, Cas A, Vir A, and Tau A); obtaining XX and YY gain solution towards the primary flux calibrator using a 3C196 skymodel provided by V.N. Pandey; determining the clock offsets between core and remote stations using the primary flux calibrator phases solutions as described by van Weeren et al. (2016); measuring the XX and YY phase offsets for the calibrator; transferring of amplitude, clock values and phase offsets to the target field; averaging each subband to a resolution of 4 seconds and 4 channels (no averaging is done for cycle 0 data); initial phase calibration of the amplitude corrected target field using a LOFAR skymodel of Boötes. The final products from the DI calibration are fiducial datasets consisting of 10 subbands equivalent to 2 MHz bands. Each observation is composed of 23 or 21 bands depending on the number

²<http://lofar.target.rug.nl/>

Table 3.1: Summary of the LOFAR Boötes observations.

| Obs. ID | Amp. calibrator | Observation date | Frequency range [MHz] | Subbands (sb) | Ch. per sb | Ch. width [MHz] | Int. time [Seconds] | Total time [Hours] |
|---------|-----------------|------------------|--------------------------|---------------|------------|--------------------|------------------------|-----------------------|
| L240772 | 3C196 | 2014-08-10 | 110-182 | 000-365 | 8 | 24.41 | 2 | 8 |
| L243561 | 3C196 | 2014-09-15 | 110-182 | 000-365 | 8 | 24.41 | 2 | 8 |
| L374583 | 3C196 | 2014-09-24 | 120-169 | 244-487 | 16 | 12.21 | 1 | 8 |
| L400135 | 3C196 | 2015-10-10 | 120-169 | 244-487 | 16 | 12.21 | 1 | 8 |
| L401825 | 3C196 | 2015-10-21 | 120-169 | 244-487 | 16 | 12.21 | 1 | 8 |
| L133895 | 3C196 | 2013-05-13 | 110-182 | 000-365 | 4 | 48.83 | 5 | 5 |
| L131784 | 3C196 | 2013-05-07 | 110-182 | 000-365 | 4 | 48.83 | 5 | 10 |

of bands flagged due to RFI. We limit the frequency to the range 120-167 MHz to accomplish an uniform coverage in the frequency domain.

The DI calibrated bands are imaged at medium-resolution ($\sim 40'' \times 30''$) using `wsclean`³ (Offringa et al. 2014). From these images, we construct a medium-resolution skymodel that is subtracted from the visibility data. Later, these data are imaged at low-resolution ($\sim 110'' \times 93''$) to obtain a low-resolution skymodel. This two-stage approach allows to include extended emission that could have been missed in the medium-resolution image. Both medium- and low- resolution skymodels are combined to create the band skymodel. Finally, the band skymodel is used to subtract the sources from the UV data to obtain DI residual visibilities. This subtraction is temporarily, as these sources will be added later in the directional self-calibration process. This stage of the data processing is carried out using the `prefactor`⁴ tool.

³<https://sourceforge.net/projects/wsclean/>

⁴<https://github.com/lofar-astron/prefactor/>

3.3.2 Direction dependent calibration

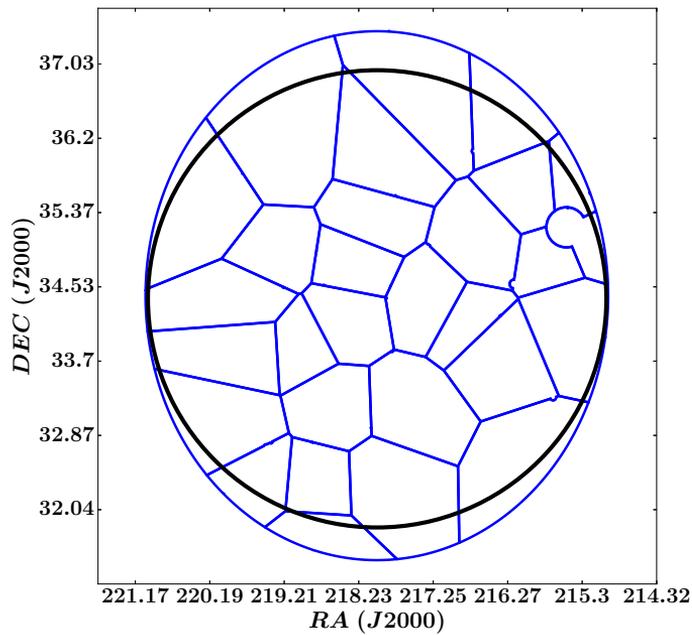


Figure 3.2: The spatial distribution of the facets in the Boötes field (blue solid lines). The large circle (solid black line) indicates the radial cutoff of 2.5 degrees used to apply the primary beam correction.

Direction-dependent (DD) effects such as the spatial and temporal variability of the LOFAR station beam response, and the ionospheric distortions must be considered to obtain high-fidelity low-frequency radio images. It is well known that these effects are responsible for artifacts and higher noise levels in low-frequency images (e.g. Yatawatta et al. 2013). A simple approach to correct these DD effects was originally proposed by Schwab (1984). If the variation of the DD effects across the field of view (FOV) is smooth, we can divide the FOV into a discrete number of regions or “facets”. Within each facet, there needs to be a bright source or group of closely spaced bright sources, which is designated as the facet calibrator. A self-calibration process can be performed on each facet calibrator. This yields a set of DD calibration solutions that are used to calibrate the whole facet. With the DD solutions applied an image of the facet is made

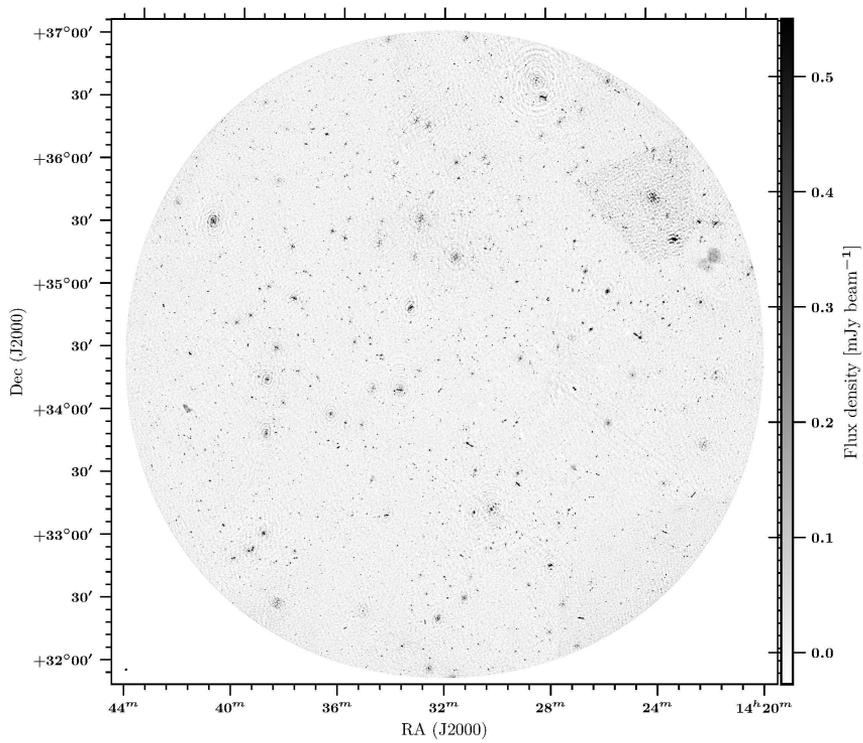


Figure 3.3: LOFAR 150 MHz mosaic of the Boötes field after beam correction. The size of the mosaic is approximately 20 deg^2 . The synthesised beam size is $3.98'' \times 6.45''$. The color scale varies from $-0.5\sigma_c$ to $10\sigma_c$, where $\sigma_c = 55 \mu\text{Jy}/\text{beam}$ is the rms noise in the central region.

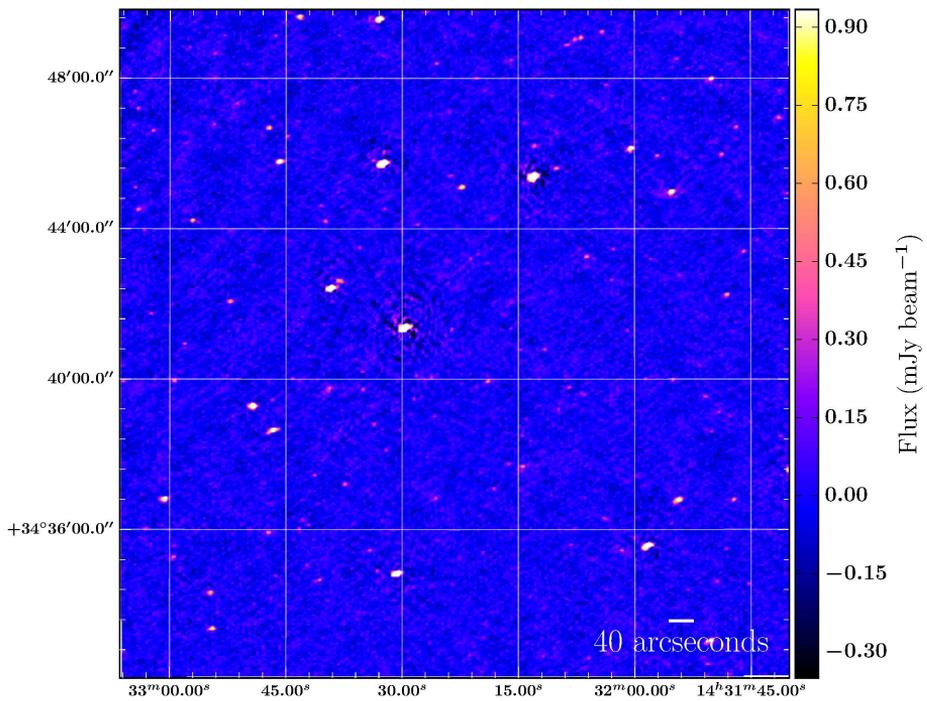


Figure 3.4: Map showing the central $400' \times 400'$ region of the mosaic center after primary beam correction. The synthesized beam size is $3.98'' \times 6.45''$. The color scale varies from $-6\sigma_l$ to $16\sigma_l$, where $\sigma_l = 55 \mu\text{Jy}/\text{beam}$ is the local rms noise.

and a model for the sources is created. Subsequently, this model is subtracted from the visibility data, and the next brightest facet is dealt with (Noordam 2004). By executing these steps in an iterative way, it is possible to correct the DD effects for all the facets in the FOV. Here, we adopt the DD calibration technique described by van Weeren et al. (2016) to process LOFAR HBA datasets. This procedure is now implemented in the `factor`⁵ pipeline.

In our data processing, we use the same facet calibrator distribution as Williams et al. (2016) with new boundary geometry (see Fig. 3.2). The range of the flux density for our facet calibrators is between 0.3 mJy and 2 Jy. To start the DD process, the corresponding facet calibrator, which was subtracted at the end of the DI calibration is added back to the UV data, and all the bands are phase-rotated in the direction of the calibrator. The self-calibration process comprises several cycles. In the first and second cycles, we solve for the phase-offsets and the total ionospheric electron content (TEC) terms (which introduces a frequency-dependent ionospheric distortion on the phases offsets) only on timescales of ~ 10 seconds. For the the third and fourth cycles, we initially solve only for phase+TEC. Finally, we obtain phase+amplitude solutions on large timescales (> 5 minutes for bright calibrators) to mainly capture the relative slow variations in the beam. The last self-calibration cycle can be iterated various times until convergence is achieved. This last iteration step helps to decrease the number of artifacts around bright facet calibrators.

The imaging of the facet starts when the sources not selected as facet calibrators are added back to the UV data and the DD solutions are applied. The facet is imaged in two stages with `wsclean` (Offringa et al. 2014). First, it is imaged at high resolution ($\sim 5''$) to include all the compact sources in a high-resolution facet skymodel. Secondly, the brightest sources from the high-resolution skymodel are subtracted, and the facet is imaged at low-resolution ($\sim 25''$) to obtain a skymodel that includes diffuse emission that can be missed during the high-resolution imaging step. Both high and low resolution models are combined into a new updated skymodel for the facet that is subtracted from the full data. This process does not only improve the DI residual visibilities by reducing

⁵<https://github.com/lofar-astron/factor>

the effective noise in the UV data as the source subtraction is performed now using the DD solutions, but also suppresses the effect of the presence of bright calibrators on the subsequent subtraction of fainter facets. The facets are processed in a serial sequence, which is ordered in descending order according to the facet calibrator flux density.

3.3.3 Combined facet imaging

The procedure to combine different observations is summarized in the following steps:

1. Shifting to a common phase center: For each facet, the astrometry ultimately depends on the precision of the calibration model of the facet calibrator. This implies that the astrometry can be shifted between different regions due to the differences in precision between the models of facet calibrators. This also explains the reason why the astrometry for the same facet is usually slightly shifted, compared to that of other observations. To account for the astrometry offsets between different observations, we phase-shift all the data corresponding to the same facet to a common phase center.
2. Normalizing imaging weights: The data from cycle 0 (4ch,5s) has been further time averaged in comparison with the data from cycles 2 and 4 (4ch,4s). Thus, the imaging weights of cycle 0 data are multiplied by a factor of 1.25 to account for the extra time averaging.
3. Facet imaging: The phase-shifted datasets from all the observations corresponding to a facet are imaged together with `wsclean`. We use a pixel size of $1.5''$, and a robust parameter of -0.7 to obtain a more uniform weighting between short and remote baselines.
4. Mosaicing and primary-beam correction: The resulting facets from the imaging step are mosaiced using `factor`. To apply the primary beam correction, we use a beam model created by `wsclean`. The correction is carried out by dividing the facet images by the regridded `wsclean` beam model. We impose a radial cutoff where the sensitivity of the phased array beam is 50 per cent of that at the pointing center (i.e. a radius of ~ 2.5 deg).

3.4 Images and sources catalog

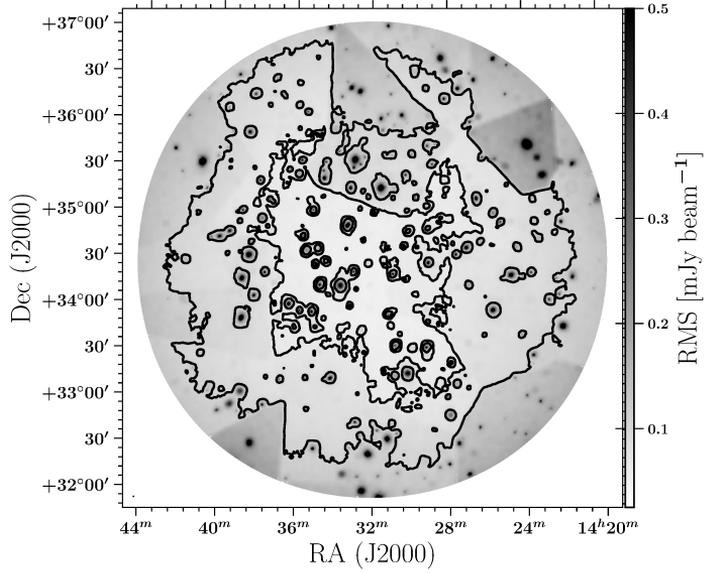


Figure 3.5: Noise map of the LOFAR 150 MHz mosaic of the Boötes field after primary beam correction. The color scale varies from $0.5\sigma_c$ to $9\sigma_c$, where $\sigma_c = 55 \mu\text{Jy}/\text{beam}$ is the rms noise in the central region. Contours are plotted at $70 \mu\text{Jy}/\text{beam}$ and $110 \mu\text{Jy}/\text{beam}$.

3.4.1 Final mosaic

The final mosaic has an angular resolution of $3.98'' \times 6.45''$ with $\text{PA} = 103^\circ$ and a central rms of $\sim 55 \mu\text{Jy}/\text{beam}$. The entire mosaic and the central region of the Boötes field are shown in Fig. 3.3 and Fig. 3.4, respectively.

3.4.2 Noise analysis and source extraction

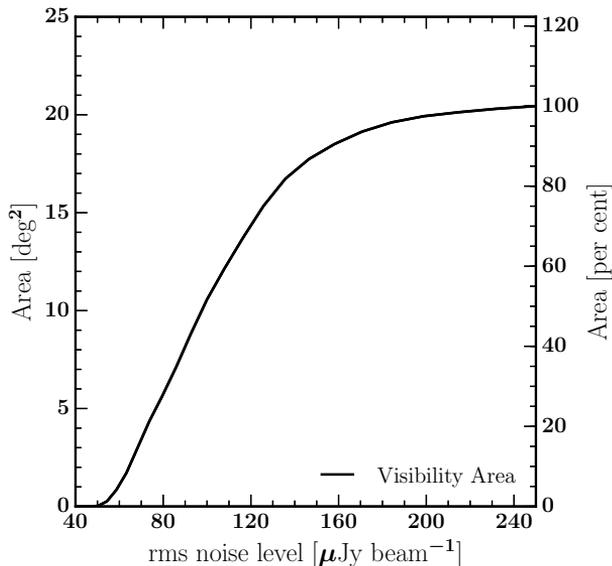


Figure 3.6: Visibility area of the LOFAR image of the Boötes field. The full area covered is 20 deg^2 .

We evaluate the spatial variation of the sensitivity of our mosaic using a noise map created by PyBDSF⁶ (the Python Blob Detection and Source Finder, formerly PyBDSM) (Mohan & Rafferty 2015). The noise map of the Boötes mosaic is shown in Fig. 3.5. The noise threshold varies from $\sim 55 \mu\text{Jy}/\text{beam}$ in the central region to $\sim 180 \mu\text{Jy}/\text{beam}$ at the mosaic edges. Around bright sources ($> 500 \text{ mJy}/\text{beam}$), the image noise can increase up to 5 times that of an unaffected region. This is caused by residual phase errors still present after DD calibration. The total area in which a source with a given flux can be detected, or visibility area, of our mosaic is displayed in Fig 3.6. As expected, the visibility area increases rapidly between $\sim 55 \mu\text{Jy}/\text{beam}$ to $\sim 250 \mu\text{Jy}/\text{beam}$, with approximately 90 per cent of the mosaic area having a rms noise less than $160 \mu\text{Jy}/\text{beam}$. Two facets located near the mosaic edge have relatively higher noise levels in comparison with adjacent facets. In these regions, the DD calibration fails as their facet calibrators have low flux densities ($S_{150\text{MHz}} < 1\text{mJy}$) resulting in

⁶<https://github.com/lofar-astron/PyBDSF>

amplitude and/or phase solutions with low S/N ratios. The application of these poor solutions to the data gives as result high-noise facets ($\sigma > 120 - 150 \mu\text{Jy}/\text{beam}$) in the mosaic.

The software package PyBDSF was used to build an initial source catalog within the chosen radial cutoff. The initial source catalog consists of 10091 sources detected above a 5σ peak flux density threshold. Of these 1978 are identified by PyBDSF with the source structure code “M” (i.e. sources with multiple components or complex structure), and the rest are classified as “S” (i.e. fitted by a single gaussian component). We inspected our mosaic and found 170 multi-component sources that are misclassified into different single sources by PyBDSF as their emission does not overlap. This includes the 54 extended sources identified by Williams et al. (2016). The components for such sources are merged together by 1) assigning the total flux from all the components as the total flux of the new merged source, 2) assigning the peak flux of the brightest component as the peak flux of the new merged source, and 3) computing the flux-weighted mean position of the components and assigning it as the position of the source. We list these merged sources as “Flag_merged” in the final source catalog.

We visually inspected the surroundings of bright objects to identify fake detections. A total of 119 objects are identified as artifacts and flagged “Flag_artifact” in our final catalog. These objects are excluded from our source counts calculations (see Section 3.6).

3.4.3 Astrometry

To check the positional accuracy, the LOFAR data is cross-correlated against the FIRST survey (Becker et al. 1995). We crossmatched the two catalogs using a matching radius of $2''$. In order to minimize the possibility of mismatching, we consider only LOFAR sources with the following criteria: i) a $S/N > 10$ in both LOFAR and FIRST maps (i.e. high S/N sources), and ii) an angular size less than $50''$ to select only compact sources with reliable positions. We find that the mean offsets in right ascension and declination for the cross-matched 989 LOFAR sources are $\langle\alpha\rangle = 0.012 \pm 1 \times 10^{-4}$ arcsec and $\langle\delta\rangle = 0.27 \pm 1 \times 10^{-4}$ arcsec, respectively. The standard deviations of the right

ascension and declination are $\sigma_{RA} = 0.57$ arcsec and $\sigma_{DEC} = 0.64$ arcsec, respectively. The examination of the offsets in the right ascension and declination directions shows that these have an asymmetrical distribution that differs between facets (see Fig. 3.7, left panels). We correct the positional offsets in both directions using the FIRST catalog for each facet independently. This is done by fitting a 2D plane to the offsets between the LOFAR and FIRST positions. The plane is $A_0(\alpha - \alpha_0) + B_0(\delta - \delta_0) + C_0 = 0$, where α and δ are the right ascension and declination of the LOFAR-FIRST sources, respectively, α_0 and δ_0 are the central right ascension and declination of the corresponding facet, and the constants A_0 , B_0 , and C_0 have units of arcseconds. This fitting provides the astrometry correction that is applied to all sources within the corresponding facet (see Fig. 3.7, right panels). We find a total of selected 1048 LOFAR/FIRST sources after the corrections are applied. The mean offsets for the corrected positions are $\langle\alpha\rangle = 0.009 \pm 1 \times 10^{-4}$ arcsec and $\langle\delta\rangle = 0.005 \pm 3 \times 10^{-4}$ arcsec, respectively. The standard deviations are $\sigma_{RA} = 0.42$ arcsec and $\sigma_{DEC} = 0.40$ arcsec, respectively. Fig. 3.8 shows the corrected positional offsets. As these offsets are typically smaller than the pixel scale in our mosaic, we do not apply any further corrections for positional offsets in our catalog.

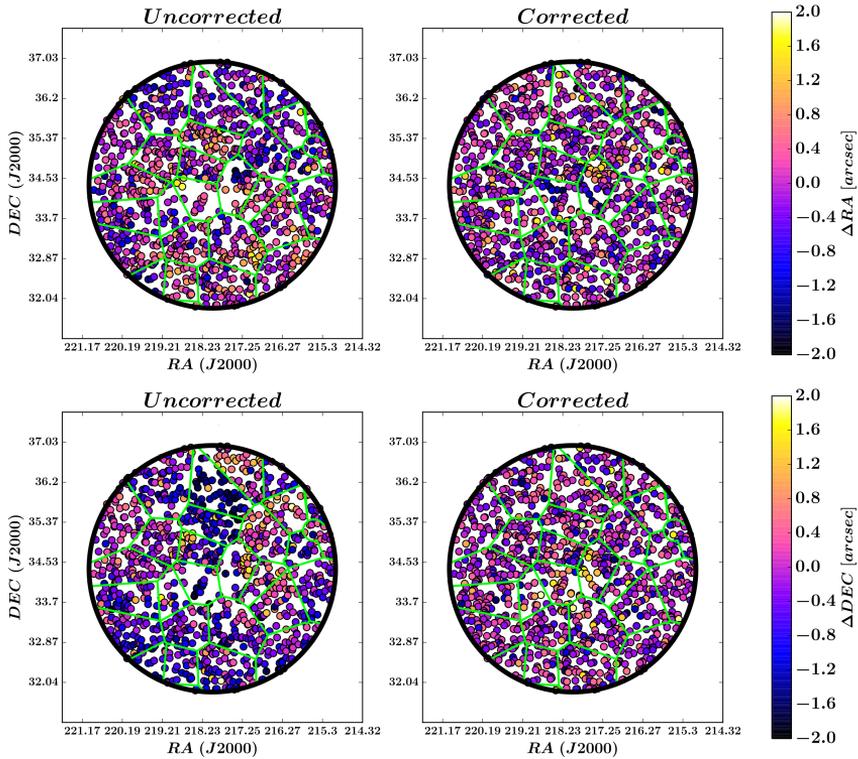


Figure 3.7: The spatial distribution of positional offsets uncorrected (left) and corrected (right) between high S/N and compact LOFAR sources and their FIRST counterparts in the right ascension (top) and declination (bottom) directions. The colorbar denotes the offsets for each object. We find 989 LOFAR/FIRST sources (left panels) using the uncorrected positions; when the astrometry corrections are applied a total of 1048 LOFAR/FIRST sources are found (right panels). The black circle indicates the radial cutoff used to apply the primary beam correction, while the green lines show the facet distribution in our Boötes mosaic.

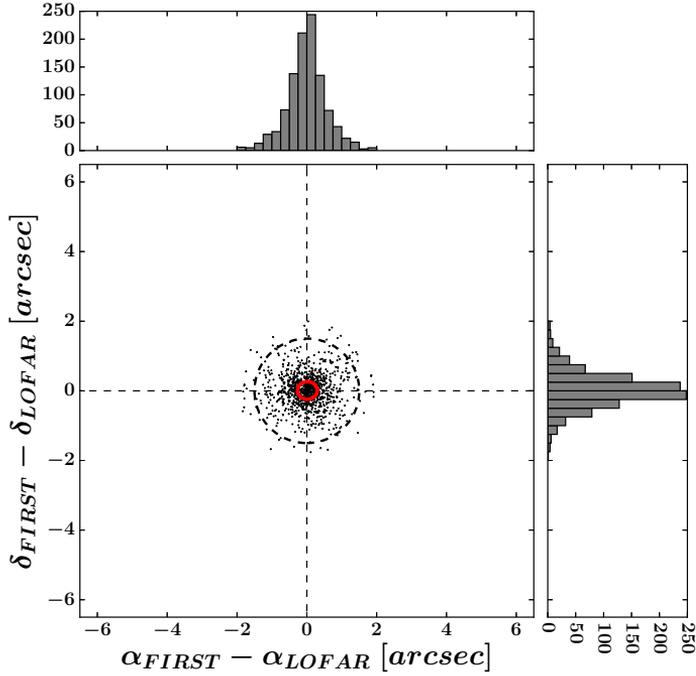


Figure 3.8: The corrected positional offsets between high S/N and compact LOFAR sources and their FIRST counterparts (see text for more details). The dashed lines denotes a circle with radius $r = 1.5''$, which is the image pixel scale. The ellipse (red solid line) centered on the right ascension and declination mean offsets indicates the standard deviation for both directions.

3.4.4 Bandwidth and time smearing

Two systematic effects that must be accounted for are bandwidth and time smearing. This combined smearing effect reduces the peak flux of a source, and simultaneously the source size is distorted or blurred in such way that the total flux is conserved, but the peak flux is reduced. The smearing effect depends on resolution, channel width, integration time, and increases with the source distance from the phase center. Williams et al. (2016) averaged their data to a resolution of 2 channels and 8 seconds, which yields a peak flux decrease of 21 per cent at 2.5 degrees from the pointing center according to the equations given by Bridle & Schwab (1999). In this work, the reduction in peak flux is less severe as our averaging factor is two times smaller in frequency and time. This results in a reduction of roughly 14 per cent at the same distance. This holds for all

the datasets that were not observed in cycle 0. For cycle 0 observations, the resolution available is 4 channel and 5 seconds. In this case, the peak flux underestimation is approximately 30 per cent at 2.5 deg from the pointing center. Following Bridle & Schwab (1999), we apply a weighted smearing correction that takes into account the frequency resolution and integration time of the data sets. The factor for Cycle 0 observations is $15/55 = 0.27$ (i.e. the ratio between the observing time obtained in Cycle 0 and the total observing time), and for the other cycles the factor is $40/55$ (i.e. its reciprocal 0.73). The smearing correction factor (≥ 1.0) depends on the distance of the source from the pointing center.

3.4.5 Flux density scale accuracy

To verify the flux density scale for our Boötes catalog and check its consistency with the Scaife & Heald (2012) flux scale, we compare our fluxes with the GMRT 150 MHz Boötes catalog by Williams et al. (2013). These authors obtained a mosaic with rms levels of 2 – 5mJy and an angular resolution of 25 arcsec. First, a representative sample of sources is chosen using the following criteria: i) a $S/N > 15$ in both LOFAR and GMRT maps (i.e. high S/N sources), ii) an angular size less than $50''$, and iii) no neighbors within a distance equal to the GMRT beam size or $25''$ (i.e. isolated sources). Secondly, we use a scaling factor of 1.078 to put the GMRT fluxes on the Scaife & Heald (2012) scale, according to the 3C196 calibration model (Williams et al. 2016). The crossmatching yields a total of 1250 LOFAR/GMRT sources. We find a mean flux ratio of $f_R = 0.88$ with a standard deviation of $\sigma_{f_R} = 0.15$, which indicates a systematic offset in our flux scale in comparison with the GMRT fluxes. Thus, we apply a correction factor of 12 per cent to our LOFAR fluxes. After correcting the fluxes, we find a mean flux ratio of $f_R = 1.00$ with a standard deviation of $\sigma_{f_R} = 0.12$ (see Fig. 3.9). Considering uncertainties on the flux scale such as: the accuracy of the fluxes on LOFAR images obtained using skymodels based on the Scaife & Heald (2012) is approximately of 10 per cent (e.g. Mahony et al. 2016; Shimwell et al. 2017a), the errors of the GMRT flux scale (Williams et al. 2016), and the differences in elevation between the calibrator and target, we conclude that a 15 per cent uncertainty in our flux

scale is appropriate. These global errors are added in quadrature to the flux uncertainties reported by PyBDSF in our final catalog.

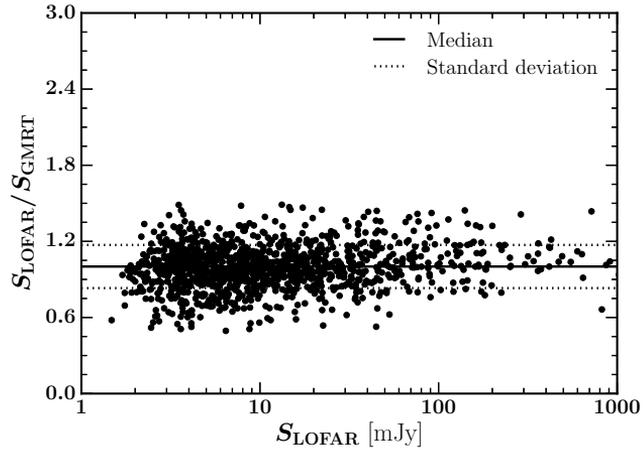


Figure 3.9: Total flux ratio for LOFAR sources and their GMRT counterparts. Only unresolved and isolated LOFAR sources with $S/N > 15$ are considered (see text for more details). The dashed lines correspond to a standard deviation of $\sigma_{f_R} = 0.12$, and the median ratio of 1.00 is indicated by a solid black line.

3.4.6 Resolved sources

We estimate the maximum extension of a radio source using the total flux S_T to peak flux S_P ratio:

$$S_T/S_P = \theta_{\text{maj}}\theta_{\text{min}}/b_{\text{min}}b_{\text{maj}}, \quad (3.1)$$

where θ_{min} and θ_{maj} are the source FWHM axes, b_{min} and b_{maj} are the synthesized beam FWHM axes. The correlation between the peak and total flux errors produces a flux ratio distribution with skewer values at low S/N , while it has a tail due to extended sources that extends to high ratios (Prandoni et al. 2000). If $S_T/S_P < 1$ sources are affected by errors introduced by the noise in our mosaic, we can derive a criterion for extension assuming that these errors affect $S_T/S_P > 1$ sources as well. The lower envelope (the curve that contains 90 per cent of all sources with $S_P < S_T$) is fitted

in the S_P/σ axis (where σ is the local rms noise). This curve is mirrored above the $S_P = S_T$ axis, and is described by the equation:

$$S_T/S_P = 1.09 + \left[\frac{2.7}{(S_P/\sigma)} \right]. \quad (3.2)$$

Using the upper envelope, we find that 4292 of 10091 (i.e. 42 per cent) of the sources in our catalog can be considered extended (see Fig. 3.10, right panel). These sources are listed as resolved in the final catalog (Section 3.4.8). However, still some objects classified by PyBDSF as made of multiple components are not identified by this criterion as resolved. Similarly, point sources could be located above the envelope by chance.

3.4.7 Completeness and reliability

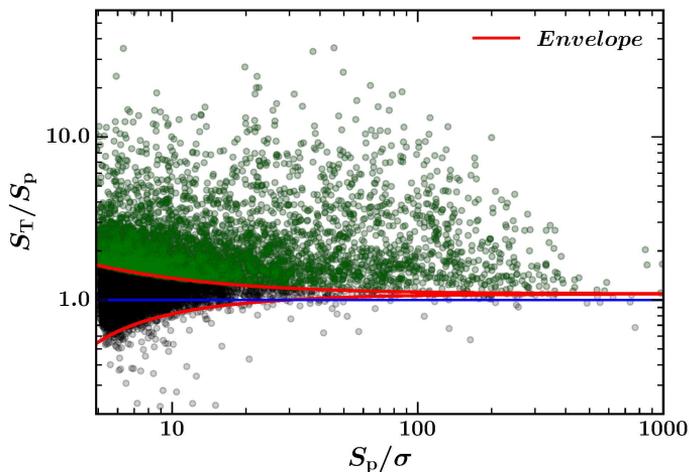


Figure 3.10: Ratio of the total flux density S_T to peak flux density S_P as a function of S/N ratio (S_P/σ) for all sources in our catalog. The red lines indicate the lower and upper envelopes. The blue line denotes the $S_T = S_P$ axis. Sources (green circles) that lie above the upper envelope are considered to be resolved.

The incompleteness in radio surveys is mainly an issue at low S/N ratios, where a significant fraction of the sources can be missed. This is consequence of the image noise on the source detection. For instance, at the detection threshold sources that are

located on random noise peaks are more easily detected than those located on noise dips (Prandoni et al. 2000).

The fraction of sources detected at 5σ in the mosaic is estimated through Monte-Carlo (MC) simulations. First, we insert artificial point sources into the residual map created by PyBDSF (see Section 3.4.2). We generate 30 random catalogs with an artificial source density of at least three times the real catalog. These artificial sources are placed at random locations in the residual map. The fluxes are drawn from a power-law distribution inferred from the real sources, with a range between 0.5σ and 30σ , where $\sigma = 55 \mu\text{Jy}/\text{beam}$. The source extraction is performed with the same parameters as for the real mosaic. To obtain a realistic distribution of sources, 40 per cent of the objects in our simulated catalogs are taken to be extended. In the MC simulations, the extended sources are modelled as objects with a gaussian morphology. Their major axis sizes are drawn randomly from values between one and two times the synthesized beam size, the minor axis sizes are chosen to have a fraction between 0.5 and 1.0 of the corresponding major axis size, and the position angles are randomly selected between 0° and 180° . We determine the completeness at a specific flux S_T by evaluating the integral distribution of the detected source fraction with total flux $> S_T$. The detected fraction and completeness of our catalog are shown in Fig. 3.11. Our results indicate that at $S_T > 1 \text{ mJy}$, our catalog is 95 percent complete, whereas at $S_T \lesssim 0.5 \text{ mJy}$ the completeness drops to about 80%.

In our facets, the presence of residual amplitude and phase errors causes the background noise to deviate from a purely Gaussian distribution. These noise deviations could be potentially detected by the source-finding algorithm as real sources. Assuming that the noise deviations can be equally likely negative or positive and real detections are due to positive peaks only, we run PyBDSF on the inverted mosaic as done in Section 3.4.2 to estimate the false detection rate (FDR) in our survey. This negative mosaic is created by multiplying all the pixels in the mosaic image by -1 . During our tests in the negative mosaic, we discovered that PyBDSF identifies a large number of artifacts around bright sources as “real” sources. This could potentially bias our FDR estimations. Therefore, we mask the regions around bright sources ($S_T > 200 \text{ mJy}$)

with circle of radius $25''$ to make certain that our FDR estimations are not dominated by artifacts. Excluding bright sources does not affect our FDR estimations, as FDR is generally relevant for fainter sources, whose noise deviations could be detected as real objects. The FDR is determined from the ratio between the number of false detections and real detections at a specific flux density bin. The reliability, $R = 1 - FDR$ at a given flux density S , is estimated by integrating the FDR over all fluxes $> S$. The FDR and reliability are plotted as a function of total flux density in Fig 3.12.

3.4.8 Source catalog

The final catalog contains 10091 sources detected above a 5σ flux density threshold and is made available online⁷. The astrometry, total and peak flux densities in the catalog are corrected as described in Sections 3.4.3, 3.4.4, and 3.4.5; respectively. The reported flux densities are on the Scaife & Heald (2012) flux density scale and their errors have the global uncertainties added in quadrature as described in Section 3.4.5. We list a sample from 13 rows of the published catalog in Table 3.2, where the columns are:

- (1) Source ID
- (2,4) source position (RA, Dec)
- (3,5) errors in source position
- (6,7) total flux density and error
- (8,9) peak flux density and error
- (10) combined bandwidth and time smearing correction factor for the peak flux density
- (11) local rms noise
- (12) source type (point source or extended)
- (13) PyBDSF source structure code (S/M)

Additionally, the catalog contains three flags not shown in Table 3.2. These flags follow the naming convention by Williams et al. (2016) as follows:

⁷<http://vizier.u-strasbg.fr/viz-bin/VizieR>

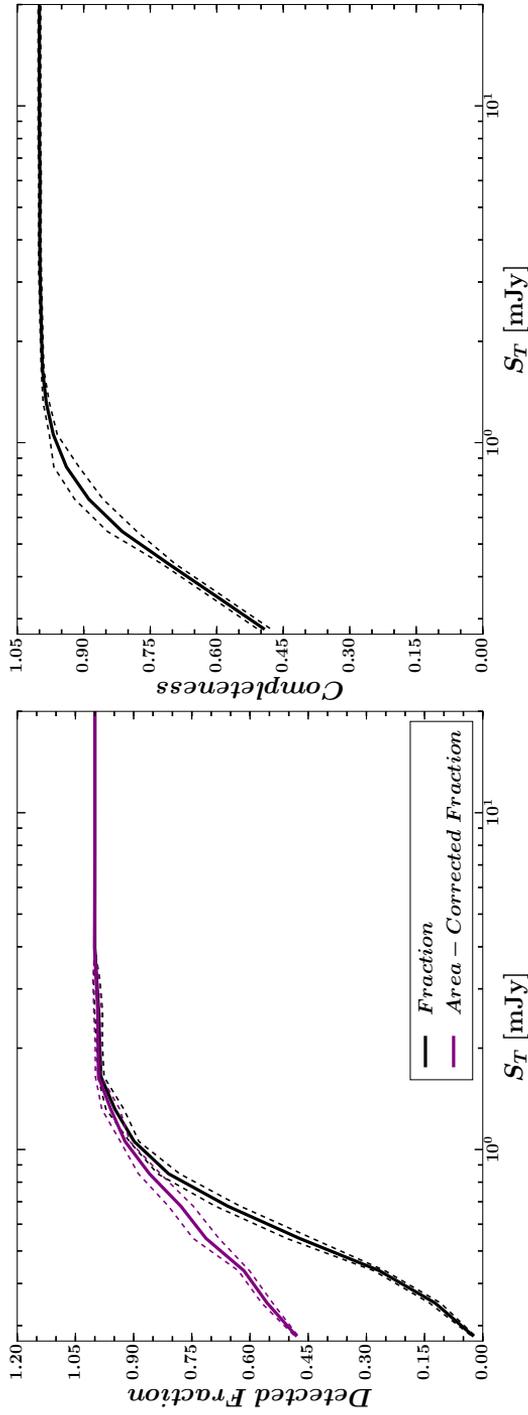


Figure 3.11: Left: The solid black line indicates the fraction of sources detected in our simulations. The solid purple line denotes the detected fraction corrected for the visibility area that is used in the source counts calculation. Right: The completeness function of our Boötes catalog as a function of flux density. Our catalog is 95 per cent complete at $S_{150\text{MHz}} > 1 \text{ mJy}$, while the completeness drops to about 80 per cent at $S_T \lesssim 0.5 \text{ mJy}$. The dashed lines in both plots represent 1σ errors estimated using Poisson statistics.

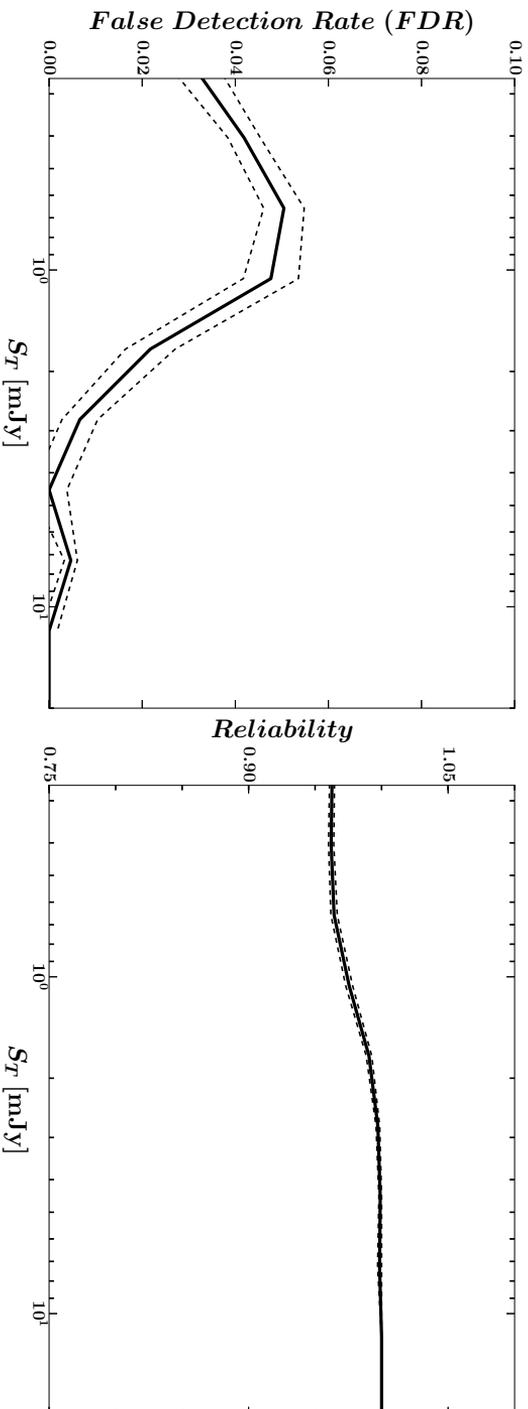


Figure 3.12: Left: False detection rate (FDR) as a function of flux density. For $S_T < 1$ mJy, the FDR is less than 5 per cent, while there are not false detections for $S_T > 5$ mJy. Right: The reliability function of our Boötēs catalog as a function of flux density. The dashed lines in both plots represent 1σ errors estimated using Poisson statistics.

- (13) Flag edge, when equals to 1 indicates an object that is located close to or in a facet edge, which could result in some flux loss.
- (14) Flag artifact, this flag indicate if an object is a calibration artifact: a value of “1” signifies a source that is probably an artifact, and “2” signifies that is surely an artifact.
- (15) Flag merged, when equal to 1 indicates a large diffuse source whose separate components are merged into a single one according to a visual inspection.

3.5 Source counts

3.5.1 Size distribution and resolution bias

Following Prandoni et al. (2001), we use estimate an upper limit Θ_{lim} for the angular size that a source of given flux can have before its peak flux falls below our detection threshold (5σ). This upper limit is defined as a function of the total flux density:

$$\Theta_{\text{lim}} = \max(\Theta_{\text{max}}, \Theta_{\text{min}}),$$

where Θ_{max} is obtained utilizing eq. 3.1 and Θ_{min} , the minimum angular size that is reliably resolved, can be derived combining eqs. 3.1 and 3.2. The constraint provided by Θ_{min} takes into account the finite size of the synthesized beam and ensures that Θ_{lim} does not become unphysical ($\Theta_{\text{max}} \rightarrow 0$ at low S/N ratios). Sources with sizes $> \Theta_{\text{max}}$ will remain undetected and the resulting catalog will be incomplete, whereas for sources with sizes $< \Theta_{\text{min}}$ the deconvolution is not reliable. This systematic effect is called resolution bias. The range of possible values for the Θ_{max} and Θ_{min} according to our rms levels are indicated by the green and yellow, respectively, shaded lines in Fig 3.13. To define the rms levels, we consider minimum and maximum noise values in our map. As shown in Fig. 3.6, 90 per cent of the total area has approximately $\sigma \lesssim 140 \mu\text{Jy}$. This value can thus be considered as representative of the maximum noise value. For the minimum noise value, we take the central rms noise in our map that is about $\sigma \sim 55 \mu\text{Jy}$. The (deconvolved) size distribution of our sources is shown in Fig. 3.13. As expected our sources tend to be smaller than the maximum allowed sizes.

A good knowledge of the angular size distribution of our LOFAR sources is critical

Table 3.2: A sample of ten rows from the source catalogue. See Section 3.4.8 for a description of the columns.

| Source ID | RA | σ_{RA} | DEC | σ_{DEC} | F_{total} | F_{peak} | F_{smear} | σ | Source type | PyBSF code |
|---------------------|--------|----------------------|-------|-----------------------|--------------------|-------------------|--------------------|------------|-------------|------------|
| (1) | (2) | (3) | (4) | (5) | (6,7) | (8,9) | (10) | (11) | (12) | (13) |
| | [deg] | [arcsec] | [deg] | [arcsec] | [mJy] | [mJy] | [mJy] | [mJy/beam] | | |
| J143941.32+340337.6 | 219.92 | 0.53 | 34.06 | 0.38 | 0.55 ± 0.15 | 0.66 ± 0.14 | 1.07 | 0.10 | P | S |
| J142880.28+323248.3 | 217.21 | 0.21 | 32.55 | 0.12 | 3.63 ± 0.58 | 2.27 ± 0.33 | 1.10 | 0.10 | E | M |
| J142948.72+325055.0 | 217.45 | 0.64 | 32.85 | 0.33 | 0.77 ± 0.16 | 0.56 ± 0.11 | 1.07 | 0.07 | P | S |
| J143105.69+341233.4 | 217.77 | 0.39 | 34.21 | 0.17 | 0.73 ± 0.15 | 0.67 ± 0.1 | 1.01 | 0.06 | P | S |
| J143212.23+340650.8 | 218.05 | 0.60 | 34.11 | 0.21 | 0.40 ± 0.10 | 0.38 ± 0.08 | 1.01 | 0.05 | P | S |
| J142280.39+334749.8 | 215.71 | 0.35 | 33.80 | 0.19 | 1.50 ± 0.28 | 1.29 ± 0.22 | 1.05 | 0.11 | E | M |
| J144016.42+354346.0 | 220.07 | 0.14 | 35.73 | 0.10 | 3.66 ± 0.60 | 3.19 ± 0.49 | 1.11 | 0.12 | P | S |
| J143847.45+351001.4 | 219.70 | 0.58 | 35.17 | 0.48 | 0.62 ± 0.14 | 0.49 ± 0.11 | 1.06 | 0.07 | P | S |
| J144101.02+344109.2 | 220.25 | 0.45 | 34.69 | 0.39 | 0.87 ± 0.21 | 0.77 ± 0.15 | 1.09 | 0.09 | P | S |
| J144030.46+354650.3 | 220.13 | 0.53 | 35.78 | 0.41 | 0.95 ± 0.25 | 0.92 ± 0.17 | 1.12 | 0.12 | P | S |

for a correct determination of the resolution bias in our survey. Particularly, at low-frequencies the sources can be more extended, and the size distribution can be different from that estimated in GHz surveys (Williams et al. 2016; Mahony et al. 2016). In Fig 3.13, we compare the median of the angular size for our sample (purple points) with the average of the two median size relations proposed by Windhorst et al. (1990, 1993) for 1.4GHz surveys:

$$\Theta_{\text{med},1} = 2 (S_{1.4\text{GHz}})^{0.3} \text{ arcsec},$$

$$\Theta_{\text{med},2} = \begin{cases} 2 (S_{1.4\text{GHz}})^{0.3} \text{ arcsec} & S_{1.4\text{GHz}} > 1 \text{ mJy} \\ 2 \text{ arcsec} & S_{1.4\text{GHz}} < 1 \text{ mJy}, \end{cases}$$

after scaling them to 150 MHz using a spectral index of $\alpha = -0.7$ (Smolčić et al. 2017b) (red solid lines). It is clear that our sources have larger median deconvolved angular sizes than those predicted by the Windhorst relations. A similar trend was found by Mahony et al. (2016) and Williams et al. (2016) in their analysis of LOFAR observations. These authors proposed to modify the Windhorst relations by increasing the normalization by factor of 2 (blue solid line) to obtain a better fit to the median angular sizes for their sources. A close examination to the median source sizes in our sample indicates that this modification indeed provides a good fit to our data. Therefore, we employ this relation to account for the resolution effects in our catalog.

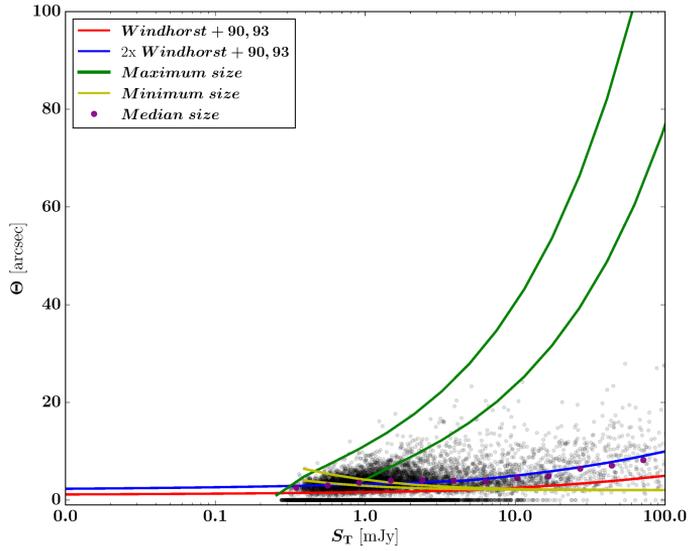


Figure 3.13: Angular size (deconvolved geometric mean) for LOFAR sources as function of their total flux density. The range of possible values for the maximum and minimum detectable angular sizes corresponding to the rms range in our mosaic ($55 - 140 \mu\text{Jy}$) are indicated by the green and yellow lines, respectively. All unresolved sources are located in the plane $\Theta = 0$, and the median source sizes for our sample are shown by purple points. The red line indicates the median of the Windhorst et al. (1990) functions, the blue line represents the same function increased by a normalization factor of 2.

To correct the source counts for the incompleteness due to the resolution bias we need to determine the *true* integral angular size distribution of radio sources as a function of the total flux density. Windhorst et al. (1990) reported an exponential form for the true angular size distribution:

$$h(\Theta_{\text{lim}}) = \exp \left[\left(b \left(\frac{\Theta_{\text{lim}}}{\Theta_{\text{med}}} \right)^a \right) \right], \quad (3.3)$$

with $a = -\ln(2)$ and $b = 0.62$. To determine the unbiased integral size distribution from our sample, we need to select sources in a total flux density range that is not affected by the resolution bias. For this purpose, we choose the flux density range $10 \text{ mJy} < S_T < 25 \text{ mJy}$. The reason for choosing this flux density range is two fold. First, the number of reliably deconvolved sources in this range is 93%, and second

to determine the integral size distribution with a large statistical sample that is close as possible to our 5σ detection threshold. In Fig. 3.14 (lef panel), we compare the integral size distribution (solid black line) for sources in our catalog with flux densities in the range $10 \text{ mJy} < S_{150\text{MHz}} < 25 \text{ mJy}$ with the 1.4GHz relations proposed by Windhorst scaled to 150 MHz using a spectral index of $\alpha = -0.7$. We find that the scaled Windhorst relations are a good representation of the integral size distribution for $\Theta \lesssim 5''$ sources, which correspond to a fraction of 80 per cent in our Boötes catalog. The resolution bias correction is defined as $c = 1/[1 - h(\Theta_{\text{lim}})]$ (Prandoni et al. 2001). Fig. 3.14 (right panel) shows the resolution bias correction as a function of the total flux density for the scaled Windhorst relations and the integral size distribution determined for our sample. We use the average of the Windhorst relations to apply the resolution bias correction to our catalog. Additionally, a 10 per cent uncertainty is added in quadrature to the errors in the source counts following Windhorst et al. (1990).

3.5.2 Visibility area

The varying noise present in our mosaic implies that objects with different flux densities are not distributed uniformly in the region surveyed. Thus, the contribution of each object to the source counts is weighted by the reciprocal of its visibility area (i.e. the fraction of the total area in which the source can be detected), as derived in Section 3.4.2. This correction allows us to account for different visibility areas within the same flux density bin.

3.5.3 Completeness and reliability

As can be seen in Fig. 3.4.7, the fraction of detected sources decreases towards fainter flux densities. Thus, a correction factor that accounts for the missed objects is required when calculating the source counts. For this purpose, we employ the detected fraction corrected for the visibility area (see Fig. 3.11) to account for the incompleteness in our source counts. Furthermore, we apply a factor to account for the reliability using the FDR derived in Section 3.4.7.

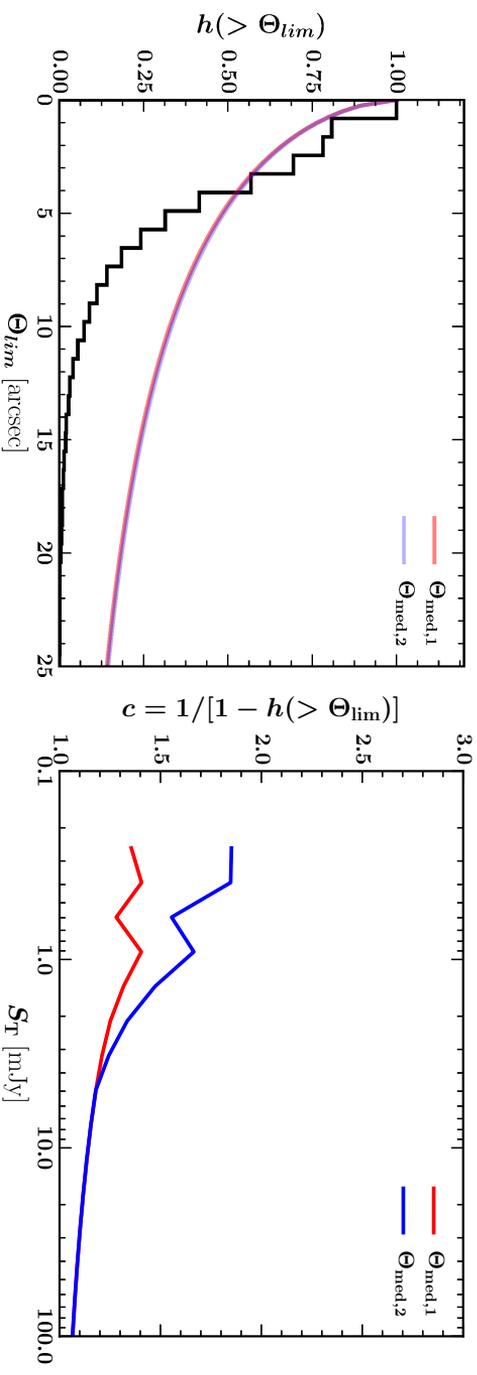


Figure 3.14: Left: Integral size distribution (black lines) for sources in our catalog with $10 \text{ mJy} < S_{150\text{MHz}} < 25 \text{ mJy}$. The red and blue lines represent the Windhorst et al. (1990) and Windhorst et al. (1993) relations scaled to 150MHz and increased by a normalization factor of 2 for the corresponding median angular sizes. Right: The resolution bias $c = 1/[1 - h(\Theta_{lim})]$ as a function of the total flux density. The color legends are the same as in the left panel.

3.5.4 Multiple-component sources

In Section 3.4.2, we carried out a visual inspection to identify resolved sources that have been misclassified into different single components by our source extraction software. However, for sources that are resolved out and split out into multiple-components and do not show signs of physical connection, establishing that their components are part of a same source is not trivial. Consequently, these components are still listed as separate sources in our catalog. This must be taken into account when computing the source counts to ensure these multi-component sources are only counted once. For this purpose, we employ the algorithm by Magliocchetti et al. (1998), to identify the missed double sources in our catalog. First, the separation between a component and its nearest neighbor, and the total flux density of the two components are compared. The components are considered as part of a double source if their flux ratio f is in the range $0.25 \leq f \leq 4$, and satisfies the separation criterion scaled to 150MHz using a spectral index of $\alpha = -0.7$:

$$\Theta_0 < 100 \sqrt{\left(\frac{S_T}{20}\right)},$$

where Θ_0 is in arcseconds and S_T is the summed flux of the two components, otherwise the components are considered independent single sources. We identify 633 sources (i.e. 6 per cent of the catalog) as doubles following the Magliocchetti et al. (1998) criterion.

3.5.5 Differential source counts

The normalized 150Hz differential radio-source counts derived from our LOFAR Boötes observations between our 5σ flux density threshold of $275 \mu\text{Jy}$ and 3Jy are shown in Fig. 3.16. Vertical error bars indicate the uncertainties obtained by propagating the errors on the correction factors to the \sqrt{n} Poissonian errors (Gehrels 1986) from the raw counts. Horizontal error bars denote the flux bins width.

For comparison purposes, previous 150 MHz source counts by Intema et al. (2017) and Franzen et al. (2016), as well as the Boötes counts obtained by Williams et al. (2016) are shown in Fig. 3.16. Additionally, we show previous results from deep fields at

1.4GHz (Padovani et al. 2015), 3GHz (Smolčić et al. 2017b) and the compilation by de Zotti et al. (2010) scaled to 150 MHz using a spectral index of $\alpha = -0.7$ (Smolčić et al. 2017b).

Our source counts are in fairly good agreement with previous low- and high- frequency surveys. At $S_{150\text{MHz}} > 1\text{mJy}$, there is a very good consistency for the source counts derived from the various surveys. The situation is different for the fainter flux bins ($S_{150\text{MHz}} < 1\text{mJy}$), where there is a large dispersion between the results from the literature. In the range $S_{150\text{MHz}} \leq 1.0\text{mJy}$, our source counts are consistent with those derived by Williams et al. (2016), and also they closely follow the counts reported by Smolčić et al. (2017b). In the flux density bins $S_{150\text{MHz}} \leq 0.4\text{mJy}$, the drop in the source counts may be the result of residual incompleteness. Our data confirms the change in the slope at sub-mJy flux densities previously reported in the literature by high- (Katgert et al. 1988; Hopkins et al. 1998; Padovani et al. 2015) and low- (Williams et al. 2016; Mahony et al. 2016) frequency surveys. This change can be associated to the increasing contribution of SF galaxies and radio-quiet AGNs at the faintest flux density bins (Smolčić et al. 2008; Padovani et al. 2009, 2011; Smolčić et al. 2017a).

3.5.6 Cosmic variance

The differences between source counts at flux densities $< 1.0\text{mJy}$ for multiple independent fields are generally larger than predicted from their Poissonian fluctuations (Condon 2007). These differences may result from either systematics uncertainties such as the calibration accuracy, primary beam correction, and bandwidth smearing, or different resolution bias corrections adopted in the literature, or cosmic variance introduced by the large scale structure. The combination of large area coverage and high sensitivity of our Boötes observations offers an excellent opportunity to investigate the effect of cosmic variance in the source counts from different extragalactic fields. For this purpose, we divide the 20deg^2 Boötes mosaic into 10 non-overlapping circular sectors, each one with an approximate area of 2deg^2 and on average containing more than 900 sources. Fig. 3.15 shows the spatial distribution of the circular sectors in the Boötes mosaic.

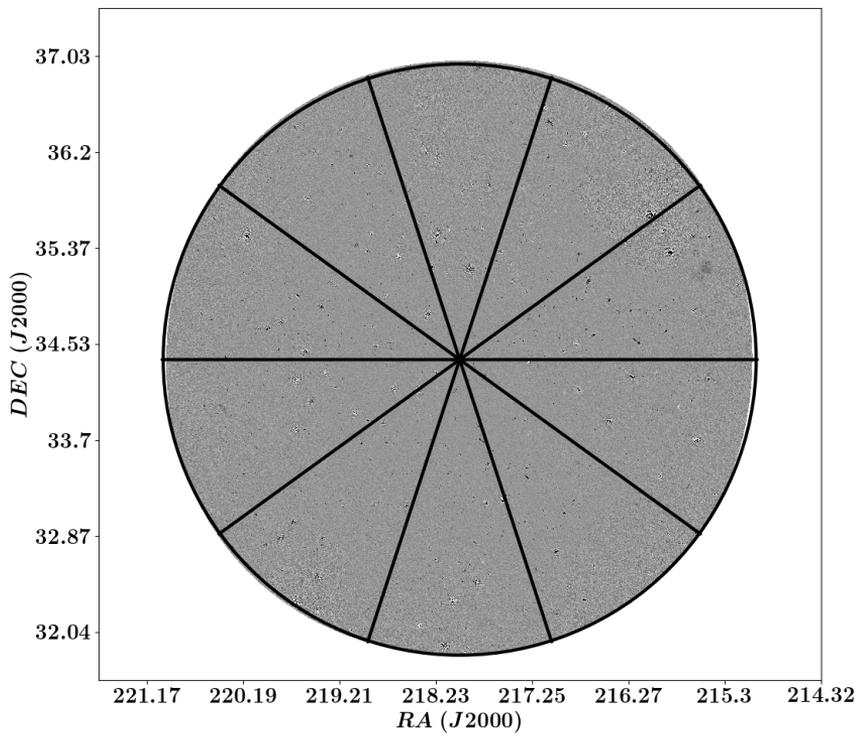


Figure 3.15: The spatial distribution of the circular sectors in the Boötes mosaic used to test the effect of cosmic variance in our source counts. Each circular sector has an approximate area of 2 deg^2 .

The purple shaded region in Fig. 3.16 shows the 1σ scatter due to cosmic variance in our source counts. The counts in each circular sector are computed in the same way as done for the entire mosaic. The comparison of the shaded region with the counts derived from deep observations scaled to 150MHz suggests that the 1σ scatter due to cosmic variance is larger than the Poissonian errors of the source counts, and it may explain the dispersion from previously reported depth source counts at flux densities $S < 1$ mJy. This confirms the results of Heywood et al. (2013) who reached a similar conclusion by comparing the scatter of observed source counts with that of matched samples from the S3-SEX simulation by Wilman et al. (2008).

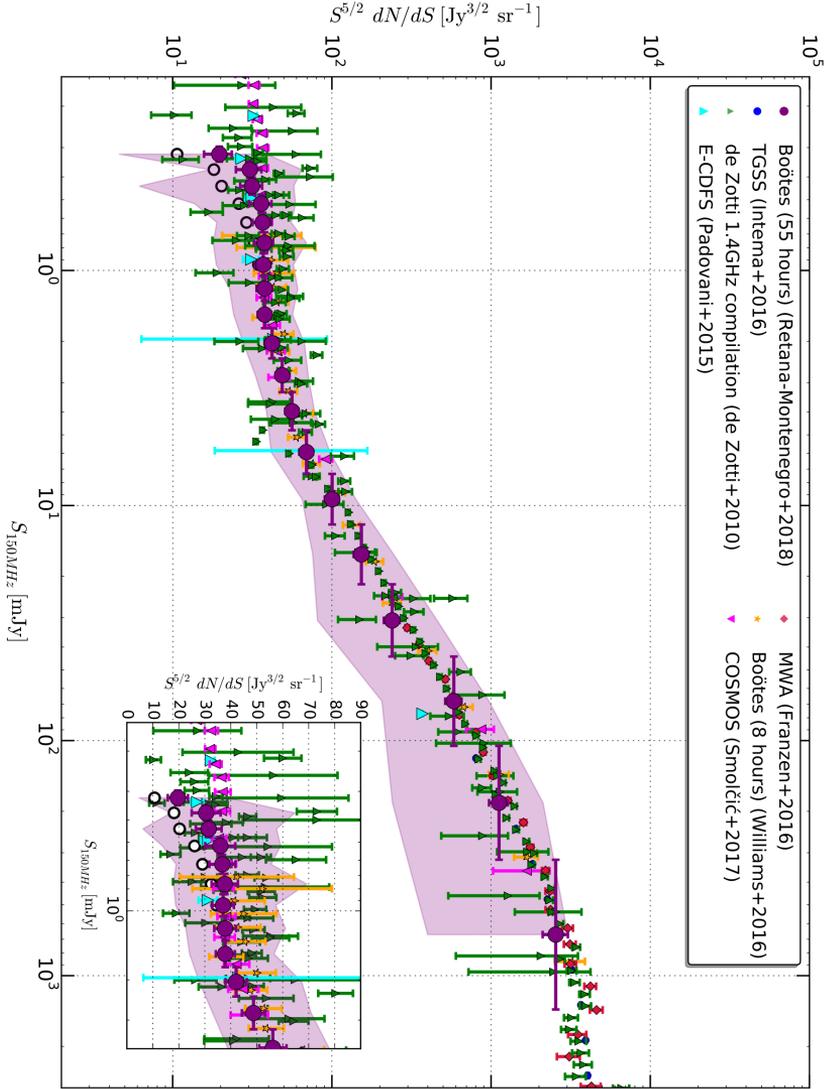


Figure 3.16: Normalized 150MHz differential radio-source counts derived from our LOFAR Boötes observations between $275 \mu\text{Jy}$ and 2 Jy (purple points). Vertical error bars are calculated assuming Poissonian statistics and horizontal error bars denote the flux bins width. Open black circles show the counts uncorrected for completeness and reliability. The purple shaded area displays the 1σ range of source counts derived from 10 non-overlapping circular sectors. For comparison, we overplot the source counts from recent deep and wide low-frequency surveys (Franzen et al. 2016; Intema et al. 2017), as well the source counts derived by Williams et al. (2016) in the Boötes field. In addition, the results of previous deep surveys carried out at 1.4GHz (de Zotti et al. 2010; Padovani et al. 2015); and 3GHz (Smolčić et al. 2017b) are scaled to 150 MHz using a spectral index of $\alpha = -0.7$ (Smolčić et al. 2017b). The inset shows the source counts in the range $0.080 \text{ mJy} \leq S_{150\text{MHz}} \leq 4 \text{ mJy}$.

3.6 Conclusions

We have presented deep LOFAR observations at 150 MHz. These observations cover the entire Boötes field down to an rms noise level of $\sim 55 \mu\text{Jy}/\text{beam}$ in the inner region, with a synthesized beam of $3.98'' \times 6.45''$. Our radio catalog contains 10091 entries above the 5σ detection over an area of 20 deg^2 . We investigated the astrometry, flux scale accuracy and other systematics in our source catalog. Our radio source counts are in agreement with those derived from deep high-frequency surveys and recent low-frequency observations. Additionally, we confirm the sharp change in the counts slope at sub-mJy flux densities. The combination of large area coverage and high sensitivity of our Boötes observations suggests that the 1σ scatter due to cosmic variance is larger than the Poissonian errors of the source counts, and it may explain the dispersion from previously reported depth source counts at flux densities $S < 1 \text{ mJy}$.

Our LOFAR observations combined with the Boötes ancillary data will allow us to perform a photometric identification of most of the newly detected radio sources in the catalog, including rare objects such as high- z quasars (Retana-Montenegro & Röttgering 2018). Future spectroscopic observations will provide an unique opportunity to study the nature of these faint low-frequency radio sources.

3.7 Acknowledgements

ERM acknowledges financial support from NWO Top project No. 614.001.006. HJAR and RJvW acknowledge support from the ERC Advanced Investigator program NewClusters 321271. PNB is grateful for support from the UK STFC via grant ST/M001229/1. LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy. The Open University is incorporated by Royal Charter (RC 000391), an exempt charity in England & Wales and a charity registered in Scotland (SC 038302). The Open University is authorized and regulated by the Financial Conduct Authority.

Chapter 4

On the Selection of High- z Quasars Using LOFAR Observations

Abstract

We present a method to identify candidate quasars which combines optical/infrared color selection with radio detections from the Low Frequency ARray (LOFAR) at 150MHz. We apply this method in a region of 9 square degrees located in the Boötes field, with a wealth of multi-wavelength data. Our LOFAR imaging in the central region reaches a rms noise of $\sim 50\mu\text{Jy}$ with a resolution of $5''$. This is so deep that we also routinely, ‘radio-quiet’ quasars. We use quasar spectroscopy from the literature to calculate the completeness and efficiency of our selection method. We conduct our analysis in two redshift intervals, $1 < z < 2$ and $2 < z < 3$. For objects at $1.0 < z < 2.0$, we identify 51% of the spectroscopic quasars, and 80% of our candidates are in the spectroscopic sample; while for objects at $2.0 < z < 3.0$ these numbers are 62% and 30%,

Retana-Montenegro and Röttgering, *Front. Astron. Space Sci*, 5, 5 (2018)

respectively. We investigate the effect of the radio spectral index distribution on our selection of candidate quasars. For this purpose, we calculate the spectral index between 1400MHz and 150MHz, by combining our LOFAR-Boötes data with 1.4GHz imaging of the Boötes field obtained with the Westerbork Synthesis Radio Telescope (WSRT), which has a rms noise of $\sigma \sim 28\mu\text{Jy}$ with a resolution of $13'' \times 27''$. We find that 27% of the candidate quasars are detected at 1400 MHz, and that these detected objects have a spectral index distribution with a median value of $\alpha = -0.73 \pm 0.07$. Using a flux density threshold of $S_{150\text{MHz}} = 1.50\text{mJy}$, so that all the $\alpha > -1.0$ sources can be detected in the WSRT-Boötes map, we find that the spectral index distribution of the 21 quasars in the resulting sample is steeper than the general LOFAR-WSRT spectral index distribution with a median of $\alpha = -0.80 \pm 0.06$. As the upcoming LOFAR wide area surveys are much deeper than the traditional 1.4GHz surveys like NVSS and FIRST, this indicates that LOFAR in combination with optical and infrared will be an excellent fishing ground to obtain large samples of quasars.

4.1 Introduction

In recent years, large spectroscopically confirmed quasar samples have become available (Croom et al. 2005; Schneider et al. 2010; Pâris et al. 2017). These quasar samples enabled statistical studies related to many topics, including the relation between the black holes (BHs) and their host galaxies (Kauffmann et al. 2003), BH growth across cosmic time (McLure & Dunlop 2004), and the quasar environments (Ross et al. 2009; Retana-Montenegro & Röttgering 2017). With the next generation of wide-field surveys such as Pan-STARRS (Kaiser et al. 2002, 2010), Dark Energy Survey (Flaugher 2005), and the future Large Synoptic Survey Telescope (Tyson 2002), such studies will be extended to the fainter quasars. A challenge in properly exploiting these surveys is the identification of quasars without spectroscopic observations.

Quasar surveys historically made use of the ultraviolet-excess (UVX) of the typical quasar spectrum (Sandage et al. 1965; Richards et al. 2002). This translates into a set of optical and near-infrared color cuts chosen to separate quasars from stars.

However, for $z > 2$ quasars this selection begins to fail as one approaches the flux limit, due to photometric errors broadening the stellar locus, and quasar and stellar color distributions blending. The necessity to increase the efficiency of quasar surveys has led to the development of new selection techniques (MacLeod et al. 2010; Yèche et al. 2010; Palanque-Delabrouille et al. 2011; Bovy et al. 2011; Kirkpatrick et al. 2011).

A way to separate high- z quasars from stars is to complement optical/infrared color cuts with a radio detection. By imposing a radio detection the stellar contamination is reduced significantly, as radio stars are very rare (Kimball et al. 2009). This approach has been successful in discovering quasars that otherwise might have been missed using typical color selection (McGreer et al. 2009; Bañados et al. 2015) such as red and dusty quasars (Glikman et al. 2004, 2012, 2013) and rare high- z quasars (Hook et al. 2002; McGreer et al. 2006; Zeimann et al. 2011).

LOFAR is a new European radio interferometer operating at frequencies 15-240MHz (van Haarlem et al. 2013) and represents a milestone in terms of radio survey speed compared to existing telescopes. The LOFAR Surveys Key Science Project aims to carry out a tiered survey. At Tier-1 level, the LOFAR Two-metre Sky Survey (LoTSS, Shimwell et al. 2017b, 2018) aims to cover the whole northern sky down to $\sim 100\mu\text{Jy}$ rms. Deeper tiers cover smaller areas in fields with extensive multi-wavelength data (Röttgering et al. 2011) with the LOFAR Boötes field the first of these deep fields to reach Tier-2 depth (Retana-Montenegro et al. 2018b). These surveys will open the low-frequency electromagnetic spectrum for exploration, allowing unprecedented studies of the radio population across cosmic time and opening up new parameter space for searches for rare, unusual objects such as high- z radio quasars in a systematic way. Perhaps, one of the most tantalizing prospects are the 21cm absorption line measurements using LOFAR along sight lines towards $z > 6$ radio quasars.

One of the possibilities to increase the efficiency in the selection of quasars is by combining optical/infrared quasar selection techniques with LOFAR radio data. With its high sensitivity, LOFAR should detect significantly more quasars in comparison with previous shallower radio surveys. Here we describe a method to select candidate quasars that combines optical/infrared color cuts with LOFAR radio detection.

4.1.1 Method Overview

- Optical color cuts to select Ly α break objects, and to separate quasars from stars.
- Mid-infrared color cuts to identify the presence of AGN-heated dust, and to reduce the contamination from low- z star-forming galaxies.
- Imposing a LOFAR 5σ detection. This point guarantees that stellar contamination in the sample is negligible.
- Fitting the UV/optical to MIR SEDs of the candidate quasars sample to quasar templates. This allows us to select the best candidates and further eliminate nonquasar contaminants from the sample.

4.1.2 Optical selection

4.1.2.1 Selection of Ly α break objects

The use of color selection to identify high- z objects was first proposed more than four decades ago (Meier 1976a,b). Since then this approach has been applied successfully to select quasars up to $z \gtrsim 6$ (Fan et al. 2001; Willott et al. 2007). The multi-color selection for finding high- z quasars usually employs at least 3 bands: one containing the Ly α emission line, one blueward (the dropout band), and one redward. This translates into a set of colors that can be used to locate the Ly α emission line.

4.1.2.2 Separating quasars and stars

Although, quasars are starlike in appearance their radiation mechanisms are different to those of stellar sources. Stars have approximately single-temperature black-body spectral energy distributions (SEDs) (Bisnovatyi-Kogan 2001), whereas energetic processes sculpt the distinctive power-law SEDs of quasars (Davidson & Netzer 1979). These differences in the SEDs imply that stars and quasars occupy different regions in colorspace (Fan 1999). Thus, in principle, optical color cuts can be chosen to reject the majority of stars.

4.1.3 Mid-infrared selection

Although, stellar contamination is reduced using the previous points, some contamination will still remain from other objects like compact low- z star-forming galaxies. These star-forming systems present a optical red colors mimicking those of quasars, which is likely caused by a strong Balmer break or dust-extincted continuum. Here, we impose the color cuts proposed by Lacy et al. (2007) and Donley et al. (2012) to the Spitzer/IRAC photometry to reduce contamination by star-forming galaxies in our quasar sample.

4.1.4 LOFAR detection

With increasing redshift the Ly α emission moves through and out of the blueward optical bands, resulting in quasars having similar colors to stars. Thus, a selection method based only on color cuts becomes less efficient at higher redshifts, as quasars occupy regions that overlap with those occupied by a significant fraction of stellar sources. This is worst at $2.2 < z < 3.0$, where the optical colors of quasars become indistinguishable from those of stars (Fan 1999; Richards et al. 2002, 2006). An alternative approach to improve the quasar selection in these regions is the incorporation of information provided by radio surveys (Richards et al. 2002; Ross et al. 2012). The number of radio stars with faint optical fluxes is very small (Kimball et al. 2009), therefore, by imposing a radio detection the stellar contamination becomes negligible in our sample.

4.1.5 Visual inspection

We carry out a visual inspection of the imaging data for each candidate quasar. This process allows us to reject contaminants such as low- z galaxies and objects with photometry contaminated by nearby bright objects. We accept candidate quasars with the following characteristics: stellar optical morphology; no bright neighbors in close proximity; and no blending issues. The radio counterparts in the LOFAR map are also examined to reject artifacts or misclassified radio-lobes.

4.1.6 Fitting the UV/optical to MIR spectral energy distributions of the candidate quasar sample

Our selection method exploits a variety of quasar observational properties to identify them in our survey data. We apply color cuts that diminish the fraction of stars and star-forming galaxies in our samples. However, these procedures do not completely eliminate confusion with other types of objects. Therefore, as a final confirmation we fit quasar templates to their SEDs. We build SEDs spanning from the optical to the MIR range to identify the candidate quasars. These SEDs are fitted to the quasar templates from the AGN template library presented by Salvato et al. (2009).

The SED fits are inspected visually. We look for the following unequivocal features in the SEDs of quasars: i) the strong break by absorption at 1215\AA (rest-frame), ii) the $\text{Ly}\alpha$ emission line, and iii) a rising or flat power-law in the IRAC bands. We examine each SED to assess the overall quality of the fit. In this way, we are able to eliminate nonquasar contaminants.

4.2 Results

4.2.1 Selecting candidate quasars in the NDWFS-Botes field

In this section, we apply the selection method using the Boötes ancillary data and our Tier-2 LOFAR catalog following the points aforementioned.

4.2.1.1 Data

The 9.3 deg^2 region in Boötes covered by the NOAO Deep Wide Field Survey (NDWFS, Jannuzi & Dey 1999) has optical data available on the U_{spec} , B_w , R , I , and Z bands. All these filters are standard except the U_{spec} and B_w , which have better efficiency and a more uniform wavelength coverage than the U-Bessel and B-Johnson filters, respectively. Additionally, the Boötes field has multi-wavelength coverage spanning from X-rays to radio wavelengths. In the X-rays and UV regimes, it has been observed with the Chandra (Kenter et al. 2005) and GALEX (Martin et al. 2005) satellites. At

infrared wavelengths, it was part of the NEWFIRM survey (J,H,K bands; Autry et al. 2003) and Spitzer Deep, Wide-Field Survey (SDWFS) with IRAC (Ashby et al. 2009). Finally, in the radio regime, the Boötes region has been observed at 1.4GHz with the VLA (Becker et al. 1995) and WSRT (de Vries et al. 2002), and at 150MHz with GMRT (Williams et al. 2013) and LOFAR (Williams et al. 2016). In this work, we use the deep 150MHz LOFAR imaging presented by Retana-Montenegro et al. (2018b), with a noise level of $1\sigma \sim 50\mu\text{Jy}$ with a spatial resolution of $5''$. We use AB magnitudes for all bands in our analysis. We assume the convention $S_\nu \propto \nu^{-\alpha}$, where ν is the frequency, α is the spectral index, and S_ν is the flux density as function of frequency.

4.2.1.2 Candidate quasars selection

To test our quasar selection method, we utilize spectroscopy data from the AGES survey (Kochanek et al. 2012). While the spectroscopic sample spans the range $0 < z < 5.8$, we limit our selection to the intervals $1.0 \leq z \leq 2.0$ and $2.0 \leq z \leq 3.0$. The reason for using these two redshift intervals is twofold. First, quasars in these intervals provide a good test for our selection method. Secondly, there are more spectroscopic confirmed quasars for the redshift intervals considered as compared to those available at $z > 3.0$.

Quasars at $1.0 \leq z \leq 2.0$ are frequently selected using the excess of ultra-violet flux in the u -band, which results in a bluer $u - g$ color as compared to that of stars with the same visual color (e.g. the $g-r$ color) (Richards et al. 2002). However, the NDWFS-Boötes bandpass system ($U_{spec}, B_w, R, I, Z_{Subaru}$) does not include a g filter found in other photometric systems such as the SDSS filter set (u, g, r, i, z) (Fukugita et al. 1996). But instead the non-standard U_{spec} ($\lambda_c = 3590$, FWHM=540) and B_w ($\lambda_c = 4111$, FWHM=1275) filters had been used. The main disadvantage of the U_{spec} and B_w filter combination is the significant wavelength overlap between the two filters. This implies that quasars at $1.0 \leq z \leq 2.0$ can not be efficiently selected using their $U_{spec} - B_w$ colors.

To optimally define the color cuts for candidate quasars at $1.0 < z < 2.0$, we generate a library of synthetic quasar spectra following the procedure described in detail

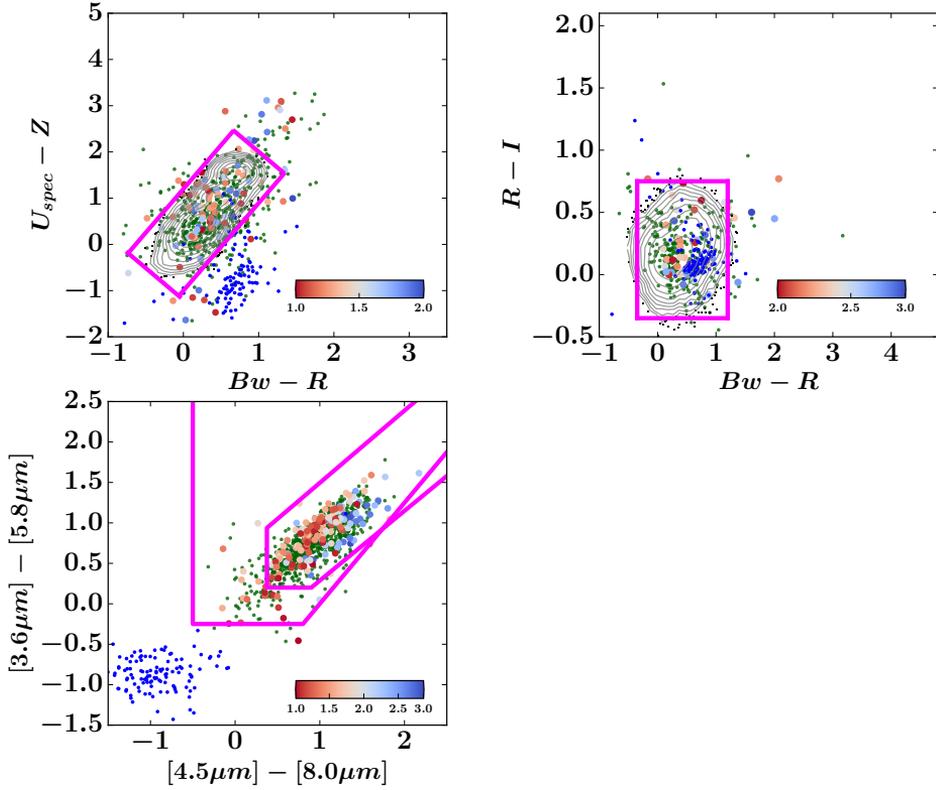


Figure 4.1: Optical and mid-infrared colors for spectroscopic quasars in the Boötes field. The LOFAR detected quasars are plotted as redshift color-coded points according to the color bar at the lower right. The corresponding redshift bin is indicated by the colorbar legend. The dark green points represent spectroscopic quasars in the corresponding redshift that are undetected in our LOFAR observations, while the blue circles mark the location of stars. The solid magenta lines delimit the regions used to select the quasars in each color space. The gray contours denotes the density for 18000 simulated quasars in the respective redshift bins.

by . These spectra are convolved with the NDWFS-Boötes filter curves to calculate the colors for the selection of $1.0 < z < 2.0$ quasars. Based on the colors derived, we adopt the color cuts shown by magenta lines in the first panel of Fig. 4.1. These color cuts are:

$$y \geq 1.89 \times x - 1.0 \wedge y \leq 1.89 \times x + 1.20$$

$$\wedge y \geq -1.37 \times x - 1.20 \wedge y \leq -1.37 \times x + 3.38,$$

where $y = B_w - R$ and $x = U_{spec} - Z$.

Based on the colors obtained from simulated quasar spectra, we derive the color cuts to select $2.0 \leq z \leq 3.0$ quasars. The color cuts adopted for the selection are the following: $-0.35 \leq R - I \leq 0.75$ and $-0.35 \leq B_w - R \leq 1.2$.

To reduce contamination from low- z star-forming galaxies in our quasar samples we adopt in both redshift bins the color cuts proposed by Donley et al. (2012):

$$w \geq 0.08 \wedge z \geq 0.15,$$

$$z \geq 1.21 \times w - 0.27 \wedge z \leq 1.21 \times w + 0.27$$

and Lacy et al. (2007):

$$w > -0.1, z > -0.2,$$

$$z \leq 0.80 \times w + 0.5,$$

where $w = \log_{10} \left(\frac{S_{5.8\mu m}}{mS_{3.6\mu m}} \right)$ and $z = \log_{10} \left(\frac{S_{8.0\mu m}}{S_{4.5\mu m}} \right)$.

Having defined the color cuts, the next task is to crossmatch the catalogs to find radio counterparts of the optical sources. We initially search for radio sources that lie within a radius of $2''$ from the optical source that fulfill our color cuts with a 5σ detection in our LOFAR catalog. For each one of these objects, we inspect its images in at least 4 bands. During this examination, we require that our candidate quasars are unresolved, not close to a bright neighbor and not blended. Additionally, we examine the morphology of the radio counterparts to prevent imaging artifacts or radio-lobes being incorrectly matched to optical sources. This examination of the radio maps ensures that only robust radio counterparts are matched to optical sources. Finally, we performed SED fitting to these sources with the photometric redshift code EAZY (Brammer et al. 2008). This allow us to assess the overall quality of the quasars photometry and to

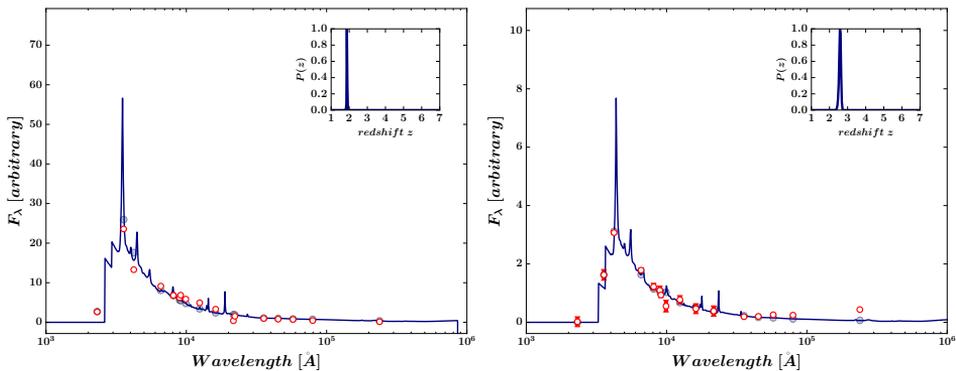


Figure 4.2: Typical examples of the spectral energy distribution for two candidate quasars identified using our selection method. In each case the best-fit quasar template (as derived from the EAZY calculation) is also plotted. Red circles are the photometric points and the blue circles indicate the predicted photometry by the best-fit template. The photometric redshifts for objects are 1.87 and 2.57, respectively. The probability density distributions (PDFs) for each object are shown in the small inset. These PDFs strongly suggest that these objects are located at high- z . The $\text{Ly}\alpha$ line in the two candidate quasars is clearly identified as an abrupt break in the quasar SED between the NUV-GALEX band and U_{spec} and B_w filters, respectively.

filter out contaminants from our sample. Fig. 4.2 shows two candidate quasars SEDs from our sample.

An important aspect to consider is the accuracy of the photometric redshifts. An inaccurate photometric redshift may cause the rejection of a quasar candidate. In Fig. 4.3, we compare the EAZY z_{photo} and z_{spec} in the range $1.0 < z < 3.0$ for Boötes spectroscopic quasars with a signal-to-noise greater than 5σ . The objects that are catastrophic outliers (i.e. objects with a difference between the photometric and spectroscopic redshift larger than the 3σ uncertainty for the photometric redshift) based on the one-to-one relation are found using an iterative 3σ -clipped standard deviation. The fraction of catastrophic outliers is around 3.1%. After catastrophic outliers are eliminated, we compute the standard dispersion $\delta z = (z_{\text{photo}} - z_{\text{spec}})/(1 + z_{\text{spec}})$ (Ilbert et al. 2006), and the normalized median absolute deviation (NMAD), defined as $\text{NMAD}(\delta z) = 1.48 \times \text{Median}(\delta z)$ (Hoaglin et al. 1983). We find $\delta z = 0.15$ and $\text{NMAD} = 0.12$. For comparison, Salvato et al. (2011) obtained more accurate photometric redshifts for

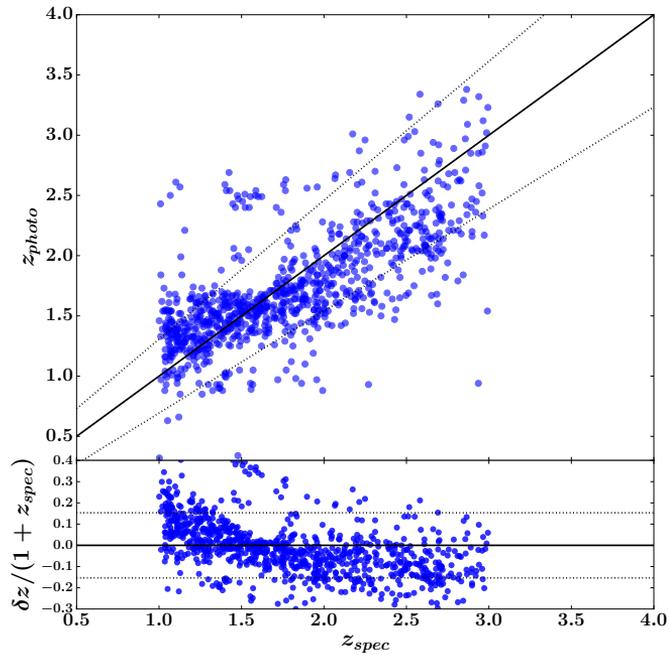


Figure 4.3: *Top panel:* Comparison between photometric and spectroscopic redshifts for 929 quasars in the Boötes field at $1.0 < z < 3.0$. The solid line represents the one-to-one $z_{phot} = z_{spec}$ relation, and the dotted lines correspond to $z_{phot} = z_{spec} \pm \sigma \times (1 + z_{spec})$. *Bottom panel:* Standard dispersion between photometric and spectroscopic redshifts as function of the spectroscopic redshift. The solid and dotted lines are the same as in the top panel.

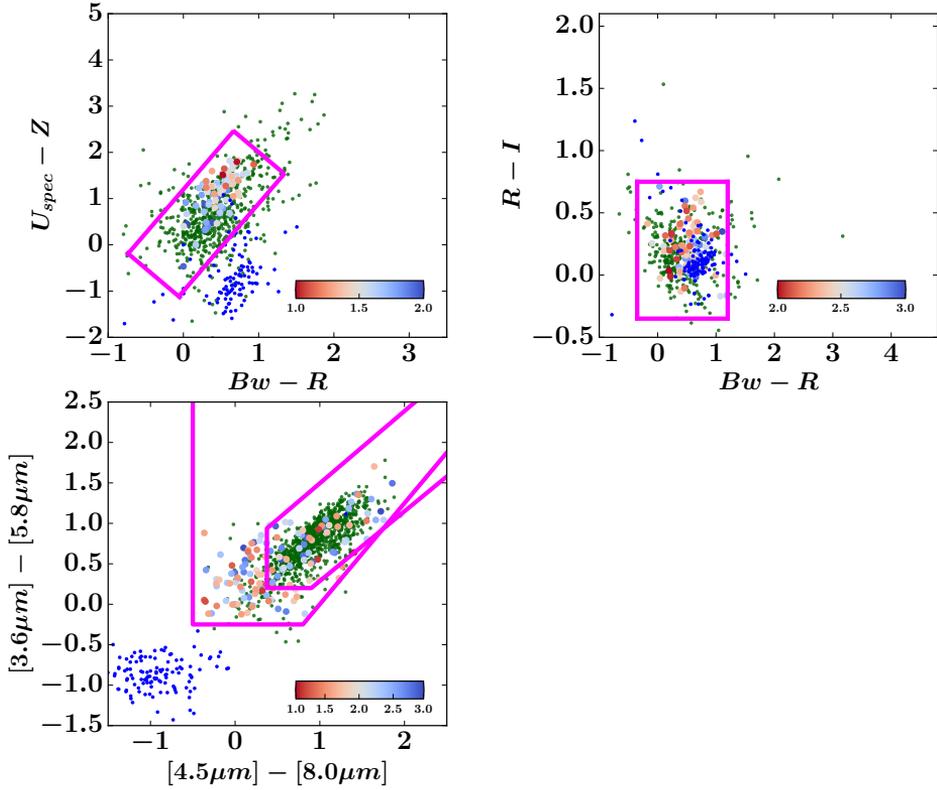


Figure 4.4: Optical and mid-infrared colors for the candidate quasars identified within our selection regions (solid magenta lines). The color-scale indicates the photometric redshift for the candidate quasars. The dark green points represent all the spectroscopic quasars (both undetected and detected by LOFAR) in the Boötes region, while the blue circles mark the location of stars. The corresponding redshift bin is indicated by the colorbar legend.

COSMOS quasars with $\text{NMAD}=0.015$ using 30 bands, while Assef et al. (2010) found $\delta z = 0.18$ for point-source AGNs in Boötes. Therefore, we conclude that fraction of candidates quasars rejected with inaccurate z_{photo} is small in comparison with the total number of candidates in the final sample.

The optical and MIR colors of the spectroscopic quasars detected by LOFAR in the Boötes field are shown in Fig. . The colors of the spectroscopic quasars are generally consistent with the proposed color cuts. Fig. 4.4 shows the colors for the 154 candidate quasars selected using our method.

4.2.1.3 Performance of the selection method

In order to assess the performance of our selection method, we calculate the completeness and efficiency for our samples.

We define the completeness C as the number of spectroscopic quasars selected as candidates compared to the *total* number of spectroscopic quasars (Hatziminaoglou et al. 2000; MacLeod et al. 2011):

$$C = \frac{\text{no. of selected spectroscopic quasars}}{\text{total no. of spectroscopic quasars}} \times 100.$$

Similarly, the efficiency E , i.e., the number of spectroscopic quasars selected as candidates compared to the number of objects selected as candidate quasars, is defined as:

$$E = \frac{\text{no. of selected spectroscopic quasars}}{\text{total no. of candidate quasars}} \times 100.$$

At $1.0 < z < 2.0$, our selection method identifies 59 of the 116 radio quasars with spectroscopic confirmation, resulting in a completeness of 51%. In the range $2.0 < z < 3.0$, 25 of 40 quasars pass our selection, which results in a completeness of 62%. For the entire redshift interval considered, we obtain a completeness of 54%.

With our method, we find 74 quasars candidates at $1.0 < z < 2.0$, which corresponds to an efficiency of 80%. In the range $2.0 < z < 3.0$, 84 candidate quasars are identified, which gives $E = 30\%$. For the full range, we find an efficiency equal to $E = 53\%$.

4.2.1.4 Effect of the radio spectral index distribution on the candidate quasar selection

In this section, we investigate the effect of the radio spectral index distribution on our selection of candidate quasars. We therefore combine our LOFAR data with the deep 1.4 GHz radio survey of the Boötes field obtained using the Westerbork Synthesis Radio Telescope (WSRT) telescope (de Vries et al. 2002). The WSRT-Boötes obser-

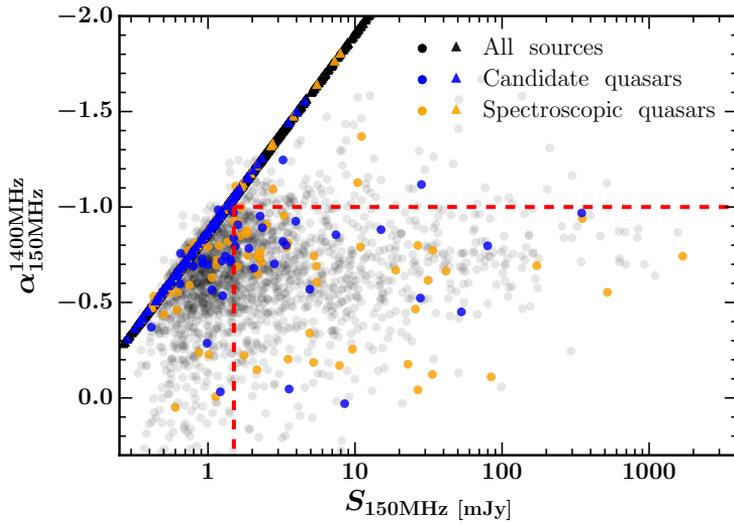


Figure 4.5: The spectral index between 1400MHz and 150MHz for sources in the Boötes field as a function of 150MHz flux density. The candidate quasars, spectroscopic quasars, and all the sources in LOFAR catalog are shown by blue, orange and black markers, respectively. The circles denote 5σ detections in the LOFAR and WSRT catalogs, while the triangles indicate upper limits on the spectral indexes assuming a 5σ WSRT detection ($S_{1.4\text{GHz}} = 0.140\text{mJy}$) for these objects. The red dashed lines indicate the region with $S_{150\text{MHz}} > 1.5\text{mJy}$ and $\alpha > -1.0$.

vations reach a rms noise of $1\sigma \sim 28\mu\text{Jy}$, with an angular resolution of $13'' \times 27''$. To compare the LOFAR and WSRT maps, we must take into account that there are incompleteness effects due to the different noise levels between the two observations. Therefore, we compare the LOFAR and WSRT observations using a flux density threshold of $S_{150\text{MHz}} = 1.5\text{mJy}$. For a spectral index of -0.7 (Smolčić et al. 2017b), this threshold is approximately equivalent to a noise level of 11σ in the WSRT-Boötes map, and ensures all the $\alpha > -1.0$ sources with a signal-to-noise greater than 5σ will be detected in the WSRT-Boötes map. The spectral index distribution for the 1998 sources in the LOFAR-WSRT sample has a median of $\alpha = -0.65 \pm 0.016$.

Using these cuts, in the overlapping area between the LOFAR and WSRT maps, we find that 42 of 154 candidate quasars are detected at 1.4 GHz. The detected objects have a spectral index distribution with a median value of $\alpha = -0.73 \pm 0.07$ (see Fig. 4.5). Using the flux density threshold of $S_{150\text{MHz}} \geq 1.50\text{mJy}$, we find that the spectral index distribution of the 21 candidate quasars in this sample is steeper than the general LOFAR-WSRT spectral index distribution with a median of $\alpha = -0.80 \pm 0.06$. The 21 candidate quasars detected at 1400MHz with $S_{150\text{MHz}} < 1.5\text{mJy}$ are characterized by a steeper spectral index distribution compared to the LOFAR full sample with a median value of $\alpha = -0.71 \pm 0.05$. For the remaining 112 candidates undetected by WSRT, we derive an upper limit for their spectral indices assuming a 5σ WSRT detection ($S_{1.4\text{GHz}} = 0.140\text{mJy}$). The median upper limit of the distribution of spectral indexes for these objects is $\alpha_{\text{upp}} < -0.75$. In the WSRT footprint, there are 70 of 139 spectroscopic quasars detected by WSRT. These detected quasars have a steeper distribution of spectral indices compared to the LOFAR-WSRT full sample with a median of $\alpha = -0.70 \pm 0.06$.

4.3 Limitations

The application of the selection method described in this work is dependent on the availability of LOFAR imaging and ancillary data. Fortunately, the dedicated LOFAR Tier-2 program selects extra-galactic fields with extensive multi-wavelength data to

maximize the scientific exploitation of the LOFAR imaging. The ongoing LoTSS survey aims to map the observable northern sky, which has been observed previously in the optical (SDSS York et al. 2000 and Pan-STARRS Kaiser et al. 2002, 2010) and MIR (WISE, Wright et al. 2010) wavelengths. These LOFAR datasets will allow us to extend the identification of candidate quasars to a larger survey volume and to smaller regions with extensive multi-wavelength data.

4.4 Summary

We have examined the identification of high- z candidate quasars with LOFAR observations as an additional tool. The motivation for our method was to compile large samples of candidate quasars and to improve the efficiency of spectroscopic programs targeting these objects. Our selection method adopts color cuts between near-infrared and optical wavelengths to obtain a list of candidate quasars, while minimizing the contamination by stars and star-forming galaxies. Second, a LOFAR detection is required to further reduce the stellar contamination in our sample. We also carried out a visual inspection of candidate quasar SEDs to discard nonquasar contaminants. We used the LOFAR Tier-2 Boötes observations as an example of the application of our method and examined its completeness and efficiency in various redshift intervals. We also investigated the effect of the radio spectral index distribution on our selection of candidate quasars. For this purpose, we calculated the spectral index between 1400 MHz and 150 MHz, by combining our LOFAR data with WSRT-Boötes imaging. We found that the candidate quasars have a steep distribution of spectral indexes with a median value of $\alpha = -0.73 \pm 0.07$.

In conclusion, this work demonstrates that our selection method combining radio detections from LOFAR with optical/infrared color cuts will provide an excellent approach for obtaining large samples of quasars.

4.5 Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.6 Author Contributions

ERM reduced the LOFAR Boötes data and carried out the source selection, as well as writing most of the text. HR contributed with ideas to the text writing.

4.7 Funding

ERM acknowledges financial support from NWO Top project, No. 614.001.006. HR acknowledges support from the ERC Advanced Investigator program NewClusters 321271.

Chapter 5

The luminosity function of LOFAR radio-selected quasars at $1.4 \leq z \leq 5.0$ in the NDWFS-Boötes field

Abstract

We present an estimate of the luminosity function (LF) of LOFAR radio-selected quasars at $1.4 < z < 5.0$ in the 9.3 deg^2 NOAO Deep Wide-field survey (NDWFS) of the Boötes field. Selection is based on optical and mid-infrared photometry used to train three different machine learning (ML) algorithms (Random forest, SVM, Bootstrap aggregation). Objects identified as quasars by the ML algorithms are required to be detected at 5σ significance in deep radio maps. Optical imaging comes from the Sloan Digital Sky Survey and the Pan-STARRS1 3π survey, mid-infrared photometry is taken from the Spitzer Deep, Wide-Field Survey, and radio data is obtained from deep LOFAR

Retana-Montenegro and Röttgering, submitted to A&A, (2019)

imaging of the NDWFS-Boötes field. The sample comprises 134 objects, including both photometrically-selected candidate quasars (51) and spectroscopically confirmed quasars (83). The depth of our LOFAR observations allow us to detect the radio-emission of quasars that otherwise would be classified as radio-quiet. Around 65% of the quasars in the sample are fainter than $M_{1450} < -24.0$, a regime where the luminosity function of quasars selected through their radio emission (radio-selected quasars, RSQs) has not been investigated in detail. It is demonstrated that in cases where mid-infrared wedge-based AGN selection is not possible due to a lack of appropriate data, the selection of quasars using ML algorithms trained with optical/infrared photometry in combination with LOFAR data provides an excellent approach for obtaining samples of quasars. The LF of RSQs can be described by pure luminosity evolution at $z < 2.4$, and a combined luminosity and density evolution at $z > 2.4$. The faint-end slope, α , becomes steeper with increasing redshift. This trend is consistent with previous studies of faint quasars ($M_{1450} < -22.0$). We demonstrate that RSQs show an evolution that is very similar to the exhibited by faint quasars. We find evidence that supports a scenario where RSQs produce only a few per cent of the photons required to ionize the intergalactic medium at $z > 3$. By comparing the spatial density of RSQs with that of the *total* (radio-detected plus radio-undetected) faint quasar population at similar redshifts, we find that RSQs may compose up to 30% of the whole faint quasar population. This fraction within uncertainties is constant with redshift. Finally, we discuss how the compactness of the RSQs radio-morphologies and their steep spectral indices could provide valuable insights how quasar and radio activity are triggered in these systems.

5.1 Introduction

A good determination of the quasar luminosity function (QLF) is important to understand several aspects of the cosmological evolution of supermassive black holes (SMBHs). These aspects include: i) the build-up of black hole (BH) demography; ii) the integrated UV contribution from quasars to the ionization of the intergalactic medium; iii) the accretion history of BHs across cosmic time; iv) triggering and fueling

quasar mechanisms and their co-evolution with host galaxies.

The cosmological evolution of quasars has been studied in detail over a wide range of optical luminosities at $z < 3$ (e.g. Richards et al. 2006; Croom et al. 2009b; Ross et al. 2013). These studies show that the comoving space density of quasars evolves strongly, with lower-luminosity quasars peaking in their space density at lower redshift than higher-luminosity quasars. This result is interpreted as a downsizing evolutionary scenario for SMBHs in which very massive BHs were already in place at very early times, whereas less massive BHs evolve predominantly at lower redshifts. These results provide valuable benchmarks to BH formation models (e.g. Volonteri & Rees 2005; Somerville et al. 2008).

At $z > 3$, the shallow flux limits of current quasar surveys restricts our understanding of BH growth to the brightest optical objects. As a result, there is little information on how faint quasars evolve. The BH downsizing behavior in the early universe is not well understood, and the role of faint quasars in the cosmic reionization of hydrogen remains poorly constrained. While studies considering only the brightest quasars found that their contribution to cosmic reionization is not significant (e.g. Haardt & Madau 2012), other authors which take into account faint quasars claim that potentially they can produce the high emissivity rate required to ionize the intergalactic medium (e.g. Glikman et al. 2011; Giallongo et al. 2015). For a good picture of the quasar phenomena at high- z , it is important to study significant numbers of these low-luminosity objects at high redshifts.

The selection of quasars at $z > 2.2$ is challenging, particularly for low-luminosity quasars with optical magnitudes close to the flux limit. In this regime, the photometric errors broaden the stellar locus, and color distributions of quasars and stars are hard to distinguish using color selection. One approach to circumvent this difficulty, is to build quasar samples using optical/infrared color selection combined with a radio detection. The main advantages of radio selection over typical color selection is that stellar contamination is reduced significantly as radio stars are very rare (Kimball et al. 2009). The application of this technique has therefore led to the discovery of quasars outside the typical color boxes used to select them (see Hook et al. 2002; McGreer et al. 2009;

Zeimann et al. 2011; Bañados et al. 2015).

One caveat of using radio selection is that the quasars selected may not be representative of the entire quasar demographics. In fact, the majority of studies of the luminosity function of radio-selected quasars (RSQs) (Shaver et al. 1996; Vigotti et al. 2003; McGreer et al. 2009; Carballo et al. 2006; Tuccillo et al. 2015) include only radio-loud quasars (RLQs) with luminosities $L_{1.4GHz} \gtrsim 1 \times 10^{26}$ W/Hz, that are selected using shallow all-sky radio and optical surveys such as FIRST (Becker et al. 1995) and SDSS (York et al. 2000), respectively. Interestingly, the works presented by McGreer et al. (2009) and Tuccillo et al. (2015) found that the luminosity function of RSQs shows a flattening of the bright-end that is similar to the whole quasar population at $3.5 < z < 4.4$ (Richards et al. 2006). Moreover, the analysis by Cirasuolo et al. (2005) suggests a decrement of the space density of faint RSQs from $z \simeq 1.8$ to $z = 2.2$ by a factor of 2. An issue is the origin of the radio emission in radio-quiet quasars (RQQs). It is still a matter of debate whether it is linked to star-forming activity occurring in the host galaxy (Kimball et al. 2011; Padovani et al. 2011; Condon et al. 2013; Bonzini et al. 2013) or non-thermal processes near the SMBH (Prandoni et al. 2010; Herrera Ruiz et al. 2016).

The number of known quasars has increased dramatically in the course of the last two decades (e.g. Croom et al. (2009a); Pâris et al. (2018)). This number will continue increasing with forthcoming facilities such as the Large Synoptic Survey Telescope (LSST, Ivezić et al. 2019), WFIRST (Spergel et al. 2015), and Euclid (Laureijs et al. 2011), expected to deliver millions of quasars (e.g. Ivezić et al. 2014). One of the major challenges is the identification of quasars without spectroscopic observations, which are costly in terms of telescope time for such large samples. Several machine learning (ML) techniques have been proposed to photometrically create large samples of quasars, including artificial neural networks (Yèche et al. 2010; Tuccillo et al. 2015), random forest (Schindler et al. 2017), support vector machine (Gao et al. 2008; Peng et al. 2012; Jin et al. 2019), extreme deconvolution (Bovy et al. 2011; Myers et al. 2015), bayesian selection (Kirkpatrick et al. 2011; Peters et al. 2015; Yang et al. 2017), and bootstrap aggregation (Timlin et al. 2018).

Many of these quasars will be detected by the next generation of radio-surveys (Norris 2017). Particularly, low-frequency radio telescopes such as the Low Frequency Array (LOFAR, van Haarlem et al. 2013) open a new observational spectral window to study the evolution of quasar activity. The LOFAR Surveys Key Science Project (LSKSP, Röttgering et al. 2011) aims to map the entire Northern Sky down to $\lesssim 100 \mu\text{Jy}$, while for extragalactic fields, greater than a few square degrees in size and with extensive multi-wavelength data, the target rms noise is of a few tens of μJy . In this paper, we investigate the evolution of the luminosity function of RSQs. For this purpose, we take advantage of the deep optical, infrared and LOFAR data available for the NOAO Deep Wide-field survey (NDWFS) Boötes field (Jannuzi & Dey 1999).

This paper is organized as follows. In Section 5.2, we present the surveys used in this work. In Section 5.3, we briefly discuss the classification algorithms employed to compile our quasar sample. We explain the method utilized to compute the photometric redshifts for our photometrically-selected candidate quasars in Section 5.4. In Section 5.4.5, we compare the LOFAR and wedge-based mid-infrared selection for objects classified as quasars by the ML classification algorithms. In Section 5.5, we present the luminosity function of our RSQs, and compare our results with previous works from the literature in Sections 5.6.1 and 5.6.2. The comoving spatial density of RSQs is studied in Section 5.6.3. The contribution of RSQs to photoionization of the intergalactic medium (IGM) is discussed in Section 5.6.4. Section 5.7 discusses how the compactness of the RSQs radio-morphologies and their steep spectral indices could provide insights into how quasar and radio activity are triggered. Section 5.7.3 discusses the possible location of RSQs in different spectroscopic parameter spaces. Finally, we summarize our conclusions in Section 5.8. For simplicity, we henceforth refer to photometrically-selected candidate quasars as photometric quasars and spectroscopically confirmed quasars as spectroscopic quasars. Also, we refer to published samples of quasars with $M_{1450} > -22.0$ (Siana et al. 2008; Glikman et al. 2011; Masters et al. 2015; Yang et al. 2018; McGreer et al. 2018; Akiyama et al. 2018) as faint quasars. Through this paper, we use a Λ cosmology with the matter density $\Omega_m = 0.30$, the cosmological constant $\Omega_\Lambda = 0.70$, the Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. We

assume a definition of the form $S_\nu \propto \nu^{-\alpha}$, where S_ν is the source flux, ν the observing frequency, and α the spectral index. To estimate radio and optical luminosities, we adopt optical and radio indices $\epsilon_{\text{opt}} = 0.5$ and $\alpha = 0.7$, respectively. All the magnitudes are expressed in the AB magnitude system (Oke & Gunn 1983).

5.2 Data

In this section, we introduce the datasets that will be utilized for the selection of quasars, and for the estimation of photometric redshifts for objects without spectroscopy.

5.2.1 NOAO Deep Wide-field survey

The NOAO Deep Wide-field survey (NDWFS) is a deep imaging survey that covers approximately two 9.3 deg^2 fields (Jannuzi & Dey 1999). One of these regions, the Boötes field has a large wealth of data available at a range of observing windows including: X-ray (*Chandra*; Kenter et al. 2005), UV/optical (*NUV*, *U_{spec}*, *B_W*, *R*, *I*, *Z_{Subaru}* bands; Jannuzi & Dey 1999; Martin et al. 2005; Cool 2007; Bian et al. 2013), infrared (*Y*, *J*, *H*, *K*, *K_s* bands, *Spitzer*; Ashby et al. 2009; Jannuzi et al. 2010), and radio (150-1400MHz; de Vries et al. 2002; Williams et al. 2013, 2016; Retana-Montenegro et al. 2018a). We use the Spitzer/IRAC- $3.6 \mu\text{m}$ matched photometry catalog presented by Ashby et al. 2009. This catalog contains 677522 sources detected at 5σ limiting magnitudes measured in a $4''$ diameter (aperture-corrected) of 22.56, 22.08, 20.24, and 20.19 at 3.6 , 4.5 , 5.8 , and $8.0 \mu\text{m}$, respectively. The $3.6 \mu\text{m}$ and $4.5 \mu\text{m}$ magnitudes are converted to AB units using the relations: $[3.6 \mu\text{m}]_{AB} = [3.6 \mu\text{m}]_{Vega} + 2.788$ and $[4.5 \mu\text{m}]_{AB} = [4.5 \mu\text{m}]_{Vega} + 3.255^2$. To select our RSQs, we use the deep 150MHz LOFAR observations of the Boötes field presented by Retana-Montenegro et al. (2018a). The image obtained covers more than 20 deg^2 and was based on 55 hours of observations. The central rms noise of the mosaic is $55 \mu\text{Jy}$ with an angular resolution of $3.98'' \times 6.45''$. The final radio catalog contains 10091 sources detected above a 5σ peak flux density threshold. There are 170 extended sources in the catalog, whose com-

²http://irsa.ipac.caltech.edu/data/COSMOS/gator_docs/scosmos_irac_colDescriptions.html

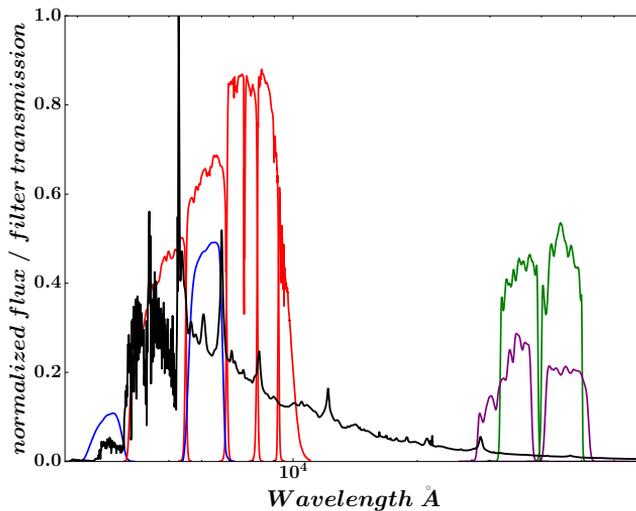


Figure 5.1: The transmission curves of the filters used in this work. *Blue lines*: SDSS-*u* and SDSS-*r* filters; *red lines*: the Pan-STARRS1 filter set: g_{PS} , r_{PS} , i_{PS} , z_{PS} , y_{PS} ; *green lines*: *Spitzer-IRAC* [3.6 μm] and [4.5 μm] bands; *purple lines*: WISE W1 and W2 bands; **and** the solid black line shows a simulated quasar spectrum from our library at $z = 3.4$ (See Section 5.5.2).

ponents were merged according to a visual inspection. This reduces the possibility of missing sources without detected cores in the LOFAR mosaic. A total of 5646 LOFAR sources are found in the *Spitzer/IRAC*-3.6 μm matched catalog using a matching radius of 2".

5.2.2 SDSS, Pan-STARRS1, WISE, and Spitzer surveys

The Sloan Digital Sky Survey (SDSS, York et al. 2000) is a multi-filter imaging and spectroscopic survey conducted with the 2.5m wide-field Sloan telescope (Gunn et al. 2006) located at the Apache Point observatory in New Mexico, USA. The SDSS-DR14 (Abolfathi et al. 2018) provides photometry for 14955 deg^2 in five broad-band optical filters (u, g, r, i, z ; Fukugita et al. 1996). The magnitude limits (95% completeness for point sources) in the five filters are $u = 22.0$, $g = 22.2$, $r = 22.2$, $i = 21.3$, and $z = 20.5$ mag, respectively.

We also use optical/near-infrared imaging from the 1.8m Pan-STARRS1 telescope

(Hodapp et al. 2004) located on the summit of Haleakala on the Hawaiian island of Maui, which provides five band photometry ($g_{PS}, r_{PS}, i_{PS}, z_{PS}, y_{PS}$). The Pan-STARRS1 first and second data releases (Chambers et al. 2016) are dedicated to the 3π survey, which observed, for almost four years, the sky north of -30° declination reaching 5σ limiting magnitudes in the $g_{PS}, r_{PS}, i_{PS}, z_{PS}, y_{PS}$ bands of 23.3, 23.2, 23.1, 22.3, 21.3, respectively. Pan-STARRS1 provides deeper imaging in overlapping optical bands (except the SDSS $-u$ band) and has the near-IR filter y_{PS} . SDSS has the u band covering wavelengths between 3000-4000 Å, which contains the Lyman alpha emission at redshifts $1.3 \lesssim z \lesssim 2.2$. This makes the SDSS $-u$ band important for the selection of $z \lesssim 2.2$ quasars. For these reasons, we combine the SDSS $-u$ band with the Pan-STARRS1 filter set ($g_{PS}, r_{PS}, i_{PS}, z_{PS}, y_{PS}$) to have wavelength coverage from 3000 Å to 10800 Å (see Figure 5.1).

As a first step to obtaining mid-infrared photometry for the spectroscopic quasars, we combine the observations from all deep Spitzer-IRAC surveys including: XFLS (Lacy et al. 2005), SERVS (Mauduit et al. 2012), SWIRE (Lonsdale et al. 2003), S-COSMOS (Sanders et al. 2007), SDWFS (Ashby et al. 2009), SHELA (Papovich et al. 2016), and SpIES (Timlin et al. 2016). We follow the same procedure described by Richards et al. (2015) to combine all the Spitzer-IRAC observations. The final catalog contains over 6.2 million Spitzer-IRAC sources. In cases where an IRAC source has been observed multiple times, we use only the deepest IRAC observation.

The mid-infrared photometry for spectroscopic quasars outside the footprint of Spitzer-IRAC surveys comes from observations by NASA's Wide Infrared Survey Explorer (WISE, Wright et al. 2010). WISE mapped the sky at 3.4, 4.6, 12 and 22 μm (known as W1, W2, W3, and W4). After the cryogenic fuel of the satellite was exhausted in 2010, WISE continued its observations as part of the post-cryogenic NEOWISE mission phase using only its two shortest bands (W1 and W2). We use the SDSS/unWISE forced photometry catalog by Lang (2014). This catalog provides forced photometry of custom WISE coadds, at the positions of over 400 million SDSS primary sources. This approach provides WISE flux measurements for sources blended in WISE coadds, but resolved in SDSS images and non-detected objects below the "official" WISE magnitude

limits (i.e. ALLWISE; Cutri 2013). We use only the W1 and W2 bands, as the other two bands are shallower, and thus have lower detection rates.

We retrieve the SDSS, Pan-STARRS1, and WISE photometry from the SDSS database via CasJobs³ and the Mikulski Archive for Space Telescopes (MAST) with PS1 CasJobs⁴. We make sure that the objects in our samples have clean photometry, by excluding sources with the SDSS bad photometry flags described in Richards et al. (2015). However, we opt to keep objects with the flag BLENDED, as high-*z* quasars could be flagged as BLENDED in some instances despite being isolated objects (e.g. McGreer et al. 2009). Only PRIMARY sources are selected from the SDSS data. The flags that describe the quality of the Pan-STARRS1 sources are taken from Table 2 in Magnier et al. (2016). For SDSS and Pan-STARRS1, we use PSF magnitudes, and adopt the *w1mag* and *w2mag* columns from the unWISE catalog as the WISE measurements for the W1 and W2 bands, respectively. These WISE magnitudes are converted to AB units using the relations: $W1_{AB} = W1_{Vega} + 2.699$ and $W2_{AB} = W2_{Vega} + 3.339$ ⁵. We consider only WISE sources meeting the following criteria: $w1_prochi2 \leq 2.0$ && $w2_prochi2 \leq 2.0$ (to avoid sources with low-quality profile fittings), and $w1_profracflux \leq 0.1$ && $w1_profracflux \leq 0.1$ (to exclude sources with fluxes severely affected by bright neighbors). The SDSS *cMODEL*MAG⁶ magnitudes are also retrieved to investigate the separation of point and extended sources (Section 5.4.4). The SDSS magnitudes in the *u* filter, originally in inverse hyperbolic sine magnitudes (Lupton et al. 1999), are converted to the AB system using $u_{AB} = u_{SDSS} - 0.04$ (Fukugita et al. 1996). The WISE-W1 and WISE-W2 photometry is converted to the IRAC 3.6 μm and 4.5 μm bands, respectively, using the transformations derived by Richards et al. (2015). We crossmatch the WISE and Spitzer-IRAC catalogs using a radius of 2". If the crossmatch is positive, we keep only the IRAC measurement. The SDSS, Pan-STARRS1 and IRAC magnitudes are corrected for Galactic extinction using the prescription by Schlafly & Finkbeiner (2011). Figure 5.1 shows the transmission curves of the filters utilized in this work.

³<http://skyserver.sdss.org/CasJobs/>

⁴<https://mastweb.stsci.edu/ps1casjobs/>

⁵<http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/>

⁶<https://www.sdss.org/dr14/algorithms/magnitudes/>

5.2.3 Spectroscopic quasars with optical and mid-infrared photometry

To efficiently discover new quasars using ML techniques, requires the compilation of a large and representative sample of spectroscopic quasars (e.g. (Richards et al. 2015; Pasquet-Itam & Pasquet 2018; Jin et al. 2019)). For this purpose, we use the Million Quasars (Milliquas) catalog v6.2 2019⁷ by Flesch (2015). This catalog contains more than 600000 type-I quasars and active galactic nuclei (AGN) from the literature, and is updated on a regular basis. The majority of quasars included in the Milliquas catalog were discovered as part of SDSS/BOSS (Schneider et al. 2010; Pâris et al. 2018), LAMOST (Yao et al. 2019), ELQS (Schindler et al. 2017), 2QZ (Croom et al. 2004), 2SLAQ (Croom et al. 2009a), and many other surveys (e.g. Papovich et al. 2006; Trump et al. 2009; Kochanek et al. 2012; Maddox et al. 2012; McGreer et al. 2013). We consider only quasars with magnitudes measured for each band, to maximize the use of the multi-dimensional color information available.

5.3 Classification

In this section, we explain how the training and target (objects to be classified) samples are compiled, and the different algorithms used for the classification of quasars in the NDWFS-Boötes field. We also assess the performance of the classification algorithms by calculating their efficiency and completeness.

5.3.1 Training sample

A critical success factor for any ML technique to classify astronomical sources is the use of an appropriate training sample to identify new objects in the target sample. The training sample must have measurements in the relevant filters to identify the characteristic spectral features (e.g. colors) of the sources of interest (e.g. quasars) in order to map their parameter space. At the same time, the training sample has to be representative of the target data. This means not only including a significant number of

⁷<http://quasars.org/>

the sources of interest, but also the other types of astronomical objects expected to be part of the target sample (i.e. stars and galaxies). In particular for quasars, the training samples require several thousands of these objects to robustly extract their color trends as a function of redshift (Yèche et al. 2010; Richards et al. 2015; Nakoneczny et al. 2019; Pasquet et al. 2019). Unfortunately, there are only 2042 quasars with $0.126 < z < 6.12$ in the NDWFS-Boötes field, with only 1259 of these quasars having redshifts larger than 1.4. The Boötes quasars in this catalog are drawn mainly from the AGN and Galaxy Evolution Survey (AGES, Kochanek et al. 2012), but other quasar surveys (Cool et al. 2006; McGreer et al. 2006; Glikman et al. 2011; Pâris et al. 2018) have been used as well. These objects are included in the Milliquas catalog (Flesch 2015). However, this sample is too small and sparse, to properly map the parameter space of quasars in the NDWFS-Boötes field. Instead, of relying only on the NDWFS-Boötes quasar sample to identify new quasars, we create a training sample using as starting point the Milliquas catalog presented in Section 5.2.3.

We restrict the redshift range of our analysis to $1.4 < z < 5.0$ for the following reasons. First, the host galaxies of some $z < 1$ quasars can be detected in the NDWFS images. This implies that the contribution of the host galaxy component to the overall quasar spectra has to be considered for these sources. This makes it difficult to separate low- z galaxies and quasars using morphological classification based on standard photometric criteria. Secondly, low- z contaminants such as star-forming, blue, and emission-line galaxies can mimic the colors of high- z quasars due to their 4000 \AA breaks (Smith et al. 1993; Richards et al. 2002). Therefore, we set $z = 1.4$ as the limit for the lower redshift interval in our analysis. This choice is a compromise between reducing contamination by galaxies and excluding good candidate quasars from the sample, but ensures that the degree of contamination due to galaxies across the redshift interval considered is as low as possible. Thirdly, at redshifts $z > 5$ the number of known quasars is significantly low compared with lower redshifts. Thus, we set $z = 5.0$ as the upper redshift limit for our analysis. In the training sample, we include spectroscopic quasars with redshifts slightly larger than the redshift limits of our analysis. This helps to reduce the possibility of quasars located at the edges of the redshift intervals not

being identified by the classification algorithms. Therefore, we include spectroscopic quasars with $1.2 < z < 5.5$ in our training sample.

The first step to create our training sample is to crossmatch the entire Milliquas catalog with the SDSS, Pan-STARRS1, WISE, and IRAC surveys as described in Section 5.2.2. We make sure that the quasars in the training sample have clean photometry, by excluding quasars with SDSS, IRAC, and WISE bad photometry flags (see Section 5.2.2). We limit the spectroscopic quasar sample to magnitudes $16.0 \leq i_{PS} \leq 23.0$ as this range contains 99.9% of all optical/mid-infrared quasars in the training sample. In total, our sample contains 328956 spectroscopic quasars with $1.2 < z < 5.5$. These quasars have clear photometry and are detected in the optical/near-infrared ($u, g_{PS}, r_{PS}, i_{PS}, z_{PS}, y_{PS}$), and mid-infrared ($3.6 \mu\text{m}, 4.5 \mu\text{m}$) bands considered in our analysis. These objects are assigned the label “quasar” in our training sample. This label assignment could be seen as trivial, but it is fundamental because the algorithms introduced in Section 5.3.3 require labels to categorize new unlabeled data in the target sample.

The rest of the training sample that comprises non-quasar objects is compiled as follows. First, we need to consider that the target sample does not only contain $1.2 \leq z \leq 5.5$ quasars, but ones with redshifts lower than 1.2 as well. Thus, to ensure that the training sample is representative, we include quasars with redshifts $0 \leq z < 1.2$ from the Milliquas catalog. There is a total of 71830 quasars that are selected as described in Section 5.2.2, and are assigned the “non-quasar” label. Secondly, stars and galaxies are expected to be part of the target sample. We include them in the training sample by randomly drawing objects with clean photometry (see Section 5.2.2) from the SDSS database with CasJobs, and excluding sources located in the NDWFS-Boötes region. These objects have $m_{AB} < 15$ in all the bands to avoid saturated pixels, and must have been classified spectroscopically as stars or galaxies by the SDSS pipeline. If the source is matched within a radius of $2''$ to a known spectroscopic quasar it is excluded. We do not apply any morphological criteria to the sources added to the training sample. The drawing process is repeated until a total of 1098858 “non-quasar” objects are selected. This size is chosen to have, in combination with the spectroscopic quasar sample, a

Table 5.1: The performance of the classification algorithms for the quasar training sample. The statistics are indicated for each subsample.

| Algorithm | Completeness | Efficiency |
|--|--------------|------------|
| | % | % |
| Random Forest | 87.83 | 96.69 |
| Support Vector Machines | 84.54 | 94.31 |
| Bootstrap Aggregation on K-Nearest Neighbors | 83.66 | 94.81 |

Notes: All experiments used the same training and test samples.

Table 5.2: Properties of the training and target samples.

| Sample | Number of objects | i_{PS} magnitude |
|-------------------------------|-------------------|---------------------------------|
| | | [AB] |
| Training sample | 5.0×10^6 | $15 \leq i_{\text{PS}} \leq 24$ |
| $1.2 \leq z \leq 5.5$ quasars | 328956 | $16 \leq i_{\text{PS}} \leq 23$ |
| $0 < z < 1.2$ quasars | 71830 | $16 \leq i_{\text{PS}} \leq 23$ |
| Target sample | 287218 | $15 \leq i_{\text{PS}} \leq 24$ |

total of a million and a half of objects in the training sample. The inclusion of galaxies, stars, and $z < 1.2$ quasars in the training sample is important because it helps the classification algorithms to delimit the color space of “quasars” and “non-quasars”. The details of the training sample are summarized in Table 5.2.

5.3.2 Target sample

The target sample is the Spitzer/IRAC- $3.6 \mu\text{m}$ matched catalog presented in Section 5.2.1. As mentioned in Section 5.3.1, there is deep optical photometry available for the NDWFS-Boötes field in the $U_{\text{spec}}, B_w, R, I, Z_{\text{Subaru}}$ bands, however, the number of quasars with photometry in these bands is small. Fortunately, the entire NDWFS-Boötes field is located inside the SDSS/Pan-STARRS1 footprint, and thus photometry from these surveys is available for the target sample. We obtain SDSS and Pan-STARRS1 photometry for all the objects in the NDWFS-Boötes catalog following the same procedure as for the training set. We remove sources with $m_{AB} < 15$ in all the bands (SDSS, Pan-STARRS1, IRAC) regardless of their flags to avoid saturated pixels. We keep only sources with the SDSS/Pan-STARRS1 good photometry flags. We have also considered using the IRAC flags (SExtractor flags indicating possible blending issues in the source

extraction) but found that around 60% of the $z > 1.4$ spectroscopic quasars in Boötes are flagged. Considering that the removal of such a significant fraction of quasars from the analysis could affect our conclusions, we decide not to remove flagged IRAC sources from the target sample at this point. In Sections 5.3.3.5 and 5.4.3, we confirm that including sources marked by the IRAC flags does not result in a deterioration of the quality of the classification, and/or determination of the photometric redshifts. Finally, the details of the target sample are summarized in Table 5.2.

5.3.3 Classification algorithms

In supervised ML, classification algorithms rely on labeled input data (training sample) to produce an inferred function, which can be used to categorize new unlabelled data (target sample). If there is a strong correlation between the input data and the labels a robust inferred function can be obtained. This usually results in a better performance of the ML classification algorithms. In this work, our aim is to identify new quasars in the NDWFS-Boötes field in the target sample using supervised ML classification algorithms. For quasars, the obvious choice is to use their colors for classification purposes (e.g. Richards et al. 2002, 2009, 2015; Timlin et al. 2018). Therefore, we use the color indices ($u - g_{PS}$, $u - r_{PS}$, $g_{PS} - r_{PS}$, $r_{PS} - i_{PS}$, $i_{PS} - z_{PS}$, $z_{PS} - y_{PS}$, $y_{PS} - [3.6 \mu\text{m}]$, $[3.6 \mu\text{m}] - [4.5 \mu\text{m}]$) of the training and target samples as inputs to the classification algorithms. The algorithms used in our analysis (Random Forest, Support vector machine, and Bootstrap aggregation) are selected because of their extensive use in previous studies of quasars (e.g. Gao et al. 2008; Peng et al. 2012; Carrasco et al. 2015; Timlin et al. 2018; Jin et al. 2019). Each one of these algorithms is briefly explained below, and are part of the open-source scikit-learn⁸ Python library.

5.3.3.1 Random forest

A random forest (RF, Breiman 2001) ensemble is composed of random decision trees, with each one created from a random subset of the data. The outputs of the decision trees are combined to make a consensus prediction. The final RF classification of an

⁸<https://scikit-learn.org>

unlabeled instance is determined using the majority vote of all decision trees.

5.3.3.2 Support vector machines

The support vector machines (SVM, Cortes & Vapnik 1995) is a discriminative classifier for two-group problems. The basic idea is to find an optimal hyperplane in an N -dimensional space (N -the number of features) that distinctly categorizes the data points. In two dimensions, the hyperplane is a line dividing the parameter space in two parts wherein each group is located on either side. For unlabeled instances, the SVM classifier outputs an optimal hyperplane which is used to classify them.

5.3.3.3 Bootstrap aggregation on K-nearest neighbors

Bootstrap Aggregation (Breiman 1996), also called bagging, is a method for generating multiple versions from a training set, by sampling the original sample uniformly and with replacement. Subsequently, each one of these subsets is used to train the K-Nearest Neighbors algorithm (KNN, Altman 1992). In the KNN algorithm, an unlabelled object is classified by a simple majority vote of its neighbors, with the object being assigned the label most common among its k nearest neighbors (k is a positive integer). In the case $k = 1$, the label assigned is of that of the single nearest neighbor. For each bagging subset, we use a value of $k = 50$. Finally, the results of the bagging subsets are aggregated by averaging to obtain a final classification.

5.3.3.4 Performance

We assess the performance of the classification algorithms with the quasar training sample presented in Section 5.3.1, by calculating their completeness and efficiency.

The completeness C (Hatziminaoglou et al. 2000; Retana-Montenegro & Röttgering 2018) is defined as the ratio between the number of quasars correctly identified as quasars, and the total number of quasars in the sample:

$$C = \frac{\text{Number of identified quasars}}{\text{Total number of quasars}} \times 100. \quad (5.1)$$

Table 5.3: The performance of the classification algorithms for the quasar training sample. The statistics are indicated for each subsample.

| Algorithm | Completeness | Efficiency |
|--|--------------|------------|
| | % | % |
| Random Forest | 93.09 | 90.50 |
| Support Vector Machines | 89.36 | 81.63 |
| Bootstrap Aggregation on K-Nearest Neighbors | 93.36 | 85.97 |

Notes: All experiments used the same training and test samples.

The efficiency E is defined as the ratio between the number of quasars correctly identified as quasars, and the total number of objects identified as quasars:

$$E = \frac{\text{Number of identified quasars}}{\text{Total no. of objects identified as quasars}} \times 100. \quad (5.2)$$

As is common practice in supervised ML, the training sample described in Section 5.3.1 is separated into validation and test samples for cross-validation (CV) purposes, in order to calculate the performance on observations of the classification algorithms. The validation set is used as the training sample, while the CV test sample, which contains spectroscopic quasars, is employed to report the completeness and efficiency of the classification algorithms. As the CV test set, we choose a random 25% subsample of the full training set. The remaining 75% of the training data is the validation sample, and is used to train the algorithms.

In Table 5.3, we present the completeness and efficiency for all the classification algorithms. While the differences are small, performance of the SVM algorithm is the worst, while RF has the highest completeness and efficiency values.

5.3.3.5 Classification results

In this section, we discuss the results of the application of the classification algorithms to our target sample. To identify the maximum number of quasars, we combined the results of the three classification algorithms. While this step increases the sample completeness, it also increases the amount of contamination on the sample despite each

algorithm having low degree of contamination (see Table 5.3). In Sections 5.3.3.6 and 5.4.4, we take additional steps to eliminate contaminants from our quasar sample. At this point, the classification algorithms identify 39160 objects as quasar candidates. Of these, 36699 lack spectroscopic observations, and 2470 sources have been classified spectroscopically. By crossmatching our sample with the Milliquas catalog, we find that 1374 are already known quasars. From these quasars, 1042 have redshifts larger than $z > 1.4$. This corresponds to a completeness of 87% for the sample of Boötes spectroscopic quasars. Additionally, we check the AGES catalog (Kochanek et al. 2012) and NED⁹ database to find that 1096 objects have been classified spectroscopically as either galaxies or galaxies. The confirmed stars and galaxies are removed from the sample.

5.3.3.6 Radio data

The contamination by the stars and galaxies in quasar samples is inevitable, as they occupy similar regions in the color space used to train the classification algorithms. As already discussed, an efficient way to eliminate stellar sources from quasar samples is to include information from radio surveys (Richards et al. 2002; Retana-Montenegro & Röttgering 2018). For this purpose, we use the LOFAR observations of the NDWFS-Boötes field by Retana-Montenegro et al. (2018a) introduced in Section 5.2.1. We crossmatch our quasar sample and the LOFAR catalog using a radius of $2''$, and inspect their stacked B_w, R, I images with radio contours overlaid to ensure that the match between optical and radio counterparts is correct, and to eliminate likely contaminants still present in our sample. To eliminate extended sources, we apply a morphology cut as a function of redshift to the photometric quasars after their photometric redshifts are calculated in Section 5.4. At this point, our RSQs sample contains 83 spectroscopic quasars and 172 photometric quasars with 5σ LOFAR detections. Figure 5.2 shows the colors of the training sample and the quasar samples in Boötes. In general, the colors of the photometric quasars agree well with those of the $z > 1.4$ quasars in the training and NDWFS-Boötes quasar samples.

⁹<https://ned.ipac.caltech.edu/>

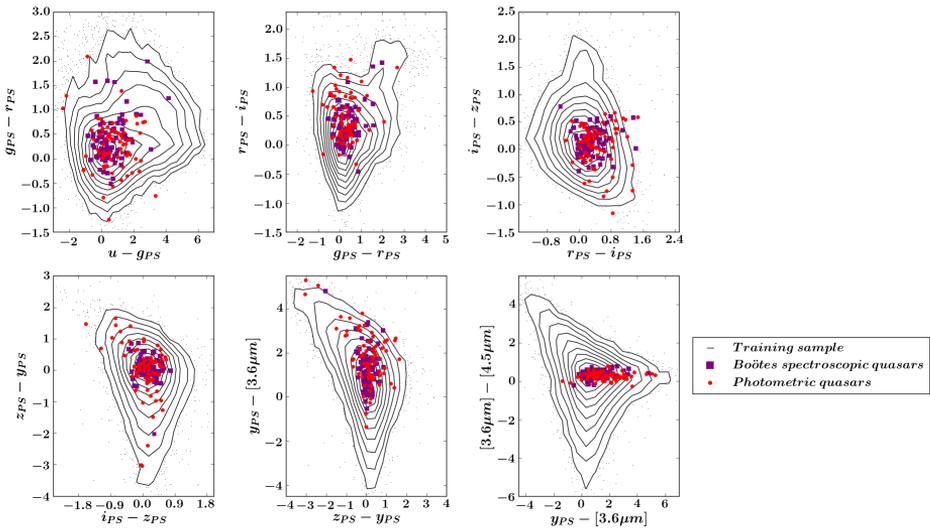


Figure 5.2: Comparison between the colors of the training and NDWFS-Boötes (photometric and spectroscopic quasars) samples. Black contours and points denote the spectroscopic quasars with $z > 1.4$ of the training sample, while purple squares indicate all the spectroscopic quasars in Boötes with $z > 1.4$. Photometric quasars in our sample are denoted by red circles. The training sample is employed to identify quasars in the target sample (see Section 5.3), and to determine their photometric redshifts with the NW kernel regression method (see Section 5.4.1).

5.4 Photometric redshifts

5.4.1 Nadaraya-Watson kernel regression

Our sample contains only 83 spectroscopic quasars. For the photometric quasars, we estimate their photometric redshifts z_{phot} using the Nadaraya-Watson (NW) kernel regression estimator (Nadaraya 1964; Watson 1964). The NW estimator is part of the family of kernel regression methods, in which the expectation of the variable Y relative to a variable X does not depend on all the X -values, as in traditional regression methods, but on sets of locally weighted values. A bandwidth scale parameter determines the amount of local averaging that is performed to obtain the estimate of Y . The NW estimator has been widely applied for nonparametric classification and regression (e.g.

Li & Racine 2011; Campbell et al. 2012; Qiu 2013), and to derive photometric redshifts (Wang et al. 2007; Timlin et al. 2018).

The NW estimate \hat{y} is a weighted average of the observed variable y_i calculated utilizing nearby points around the test point x_0 . The estimate is calculated using the following equation:

$$\hat{y}(x_0) = \sum_{i=1}^N w_i(x_i, x_0) y_i, \quad (5.3)$$

where

$$w_i(x_i, x_0) = \frac{K(x_i, x_0)}{\sum_{i=1}^N K(x_i, x_0)}, \quad (5.4)$$

is the normalized kernel built using the local information from the training sample, and N is the number of objects in the training sample. The kernel weighting function $K(x_i, x_0)$ is chosen to have a Gaussian form:

$$K(x_i, x_0) = \exp\left(-\frac{1}{2h^2} \|x_i - x_0\|^2\right), \quad (5.5)$$

where h is the bandwidth scale that defines the region of parameter space in which the data is averaged, and $\|x_i - x_0\|$ is the euclidean distance between the data points from the training and test samples. A more detailed discussion about the NW estimator can be found in Härdle (1990) and Wu & Zhang (2006).

In this work, the training sample is composed of spectroscopic quasars, and the distance is calculated between the colors of the spectroscopic quasars and photometric quasars. Finally, for each photometric quasar its photometric redshift is calculated considering all the spectroscopic redshifts of the training sample by using the equation (Wang et al. 2007; Timlin et al. 2018):

$$z_{photo} = \sum_{i=1}^N w_i z_{spec,i}. \quad (5.6)$$

5.4.2 Quasar training sample

To provide a training sample on which to use the NW estimator, a spectroscopic quasar catalog is necessary. For this purpose, we use the quasar catalog introduced in Section 5.2.3. In total, our training set contains 328956 quasars with $1.2 \leq z \leq 5.5$ to determine photometric redshifts using the NW estimator (see Table 5.2).

5.4.3 Redshift estimation

The distance between the colors indices ($u-g_{PS}$, $u-r_{PS}$, $g_{PS}-r_{PS}$, $r_{PS}-i_{PS}$, $i_{PS}-z_{PS}$, $z_{PS}-y_{PS}$, $y_{PS}-[3.6\ \mu\text{m}]$, $[3.6\ \mu\text{m}]-[4.5\ \mu\text{m}]$) of the spectroscopic quasars from the training sample and the corresponding photometric objects are used as inputs to build the kernel matrix as given in eq. 5.5. An important decision in building the kernel is the choice of the bandwidth scale h . With smaller values of h nearby data points weight more, while larger values of h result in an increasing contribution of distant data points. This h value is slightly higher than the $h = 0.05$ used in previous estimations of photometric redshifts using the NW estimator (e.g. Wang et al. 2007; Timlin et al. 2018). Finally, the kernel functions determined using eq. 5.6 are used as weights in the computation of the photometric redshift.

The quality of the photometric redshifts determined with the NW estimator is investigated utilizing two quasar samples. The first sample is composed of 1193 quasars from the quasar training set described in Section 5.2.3 that are located in the NDWFS-Boötes field. The second sample is a subsample selected randomly from the training set. We measure the performance of the photometric redshifts in the samples using the following statistics:

- mean of the difference between photometric and spectroscopic redshifts, $\langle \Delta z \rangle = \langle (z_{\text{phot}} - z_{\text{spec}}) \rangle$, unclipped;

Table 5.4: The performance of the photometric redshifts for the quasar training sample. The statistics are given for each subsample.

| Sample | Number of quasars | $\langle \Delta z \rangle$ | $\sigma(\Delta z)$ | $R_{0,3}$ | $\langle \Delta z_{\text{norm}} \rangle$ | $\sigma(\Delta z_{\text{norm}})$ | $R_{0,1}^{\text{norm}}$ | $R_{0,2}^{\text{norm}}$ |
|----------------------------------|-------------------|----------------------------|--------------------|-----------|--|----------------------------------|-------------------------|-------------------------|
| Boîtes (All) | 1193 | -0.060 | 0.35 | 0.76 | -0.013 | 0.10 | 0.78 | 0.94 |
| Boîtes ($\text{err} \leq 0.2$) | 535 | -0.015 | 0.22 | 0.87 | -0.0008 | 0.071 | 0.87 | 0.97 |
| Boîtes ($\text{err} \leq 0.3$) | 689 | -0.033 | 0.25 | 0.84 | -0.006 | 0.080 | 0.84 | 0.97 |
| Boîtes ($\text{err} \leq 0.5$) | 881 | -0.04 | 0.26 | 0.82 | -0.007 | 0.084 | 0.82 | 0.96 |
| 20% Training sample | 6573 | -0.006 | 0.28 | 0.81 | 0.005 | 0.09 | 0.82 | 0.95 |

Notes: The experiments all used the same training sample that does not include any spectroscopic quasars from the NDWFS-Boîtes field, with the exception of the experiment where the Boîtes quasars are included in the training sample.

- standard deviation of the Δz , $\sigma(\Delta z)$;
- fraction of quasars with $|\Delta z| < 0.3$, $R_{0.3}$;
- mean of normalized Δz , $\langle \Delta z_{\text{norm}} \rangle = \langle (z_{\text{phot}} - z_{\text{spec}})/(1 + z_{\text{spec}}) \rangle$, unclipped;
- standard deviation of the normalized bias, $\sigma(\Delta z_{\text{norm}})$;
- fraction of quasars with $|\Delta z_{\text{norm}}| < 0.1$ ($|\Delta z_{\text{norm}}| < 0.2$), $R_{0.1}^{\text{norm}}$ ($R_{0.2}^{\text{norm}}$).

These results are summarized in Table 5.4. We tested practically all the values in the range $0.01 \leq h \leq 0.5$, and found that $h = 0.09$ gives the best performance and least scatter for the Boötes spectroscopic quasars (see Table 5.4). Figure 5.3 shows the photometric versus spectroscopic redshifts for the spectroscopic sample of Boötes quasars. At low- z , the dispersion of the photometric is slightly higher in comparison with the high- z estimations. This is expected as at low- z there are less spectral features for the NW method to exploit and predict trends on the training sample.

Figure 5.4 presents the normalized histogram of the bias $\Delta z = z_{\text{phot}} - z_{\text{spec}}$ between photometric and spectroscopic redshifts for the different experiments listed in Table 5.4. The first row lists the result for the sample of Boötes quasars. This shows that the majority of Boötes quasars have good redshift estimations. The number of outliers is small with 76% of the quasars having photometric redshifts that do not differ more than $|\Delta z| \leq 0.3$ from their spectroscopic redshifts. A good comparison for the redshift accuracy of the Boötes quasars can be obtained using a validation sample that is randomly selected from a training sample. This validation sample contains 65573 quasars, and its size is 20% of the total training sample. The result for this sample is listed in the last row of Table 5.4. The Boötes sample performs worse than the 20% of the training sample in terms of bias, scatter, and fractions of quasars with correct redshifts. In order to better understand the reasons for the difference in performance between the two samples we examine the performance of the NW method as a function of the photometric errors in the SDSS/Pan-STARRS1 bands. By restricting the Boötes sample only to quasars with photometric errors in the SDSS/Pan-STARRS1 bands smaller than $err \leq [0.2, 0.3, 0.5]$, respectively, its performance gets closer to that of the 20%

training sample, and this further improves for the case when the photometric errors are smaller or equal to 0.3 (Table 5.4, third-fourth rows).

Finally, we compare the performance of the NW regression kernel with other photometric-redshift algorithms. For instance, Richards et al. (2001) and Weinstein et al. (2004) reported that 70% and 83% of their predicted photometric redshifts are correct within $|\delta z| \leq 0.3$. These authors used empirical methods that used the color-redshift relations to derive photometric redshifts using early SDSS data. Yang et al. (2017) considered the asymmetries in the relative flux distributions of quasars to estimate quasar photometric redshifts obtaining an accuracy of $R_{0.2}^{\text{norm}} = 87\%$ for SDSS/WISE quasars. Jin et al. (2019) employed the SVM and XGBoost (Chen & Guestrin 2016) algorithms to achieve an average accuracy of $R_{0.2}^{\text{norm}} \simeq 89\%$ for the photometric redshifts of their Pan-STARRS1/WISE quasars. Schindler et al. (2017) used SDSS/WISE adjacent flux ratios to train the RF and SVM methods to obtain average results of $R_{0.2}^{\text{norm}} \simeq 93\%$ for their photometric redshifts. In summary, despite the fact that all these algorithms employed different samples and redshift intervals, their accuracy is consistent with the results obtained in this work using the NW regression kernel.

5.4.4 Final quasar sample

We have used the NW regression algorithm to assign a photometric redshift to each photometric quasar detected by LOFAR. To eliminate potential contamination by low- z galaxies, we restrict our quasar sample only to point sources. The SDSS photometry pipeline¹⁰ classifies an object as point-like (star) or extended (galaxy) source based on the difference between its PSF and cMODEL MAG magnitudes¹¹. Various methods have been proposed to perform the morphological star/galaxy separation with photometric data by adding Bayesian priors to the aforementioned magnitude differences (Scranton et al. 2002), and using decision tree classifiers (Vasconcellos et al. 2011). Here, we employ a criterion derived directly from the Boötes spectroscopic quasars by considering

¹⁰<https://www.sdss.org/dr14/algorithms/classify/>

¹¹<https://www.sdss.org/dr14/algorithms/magnitudes/>

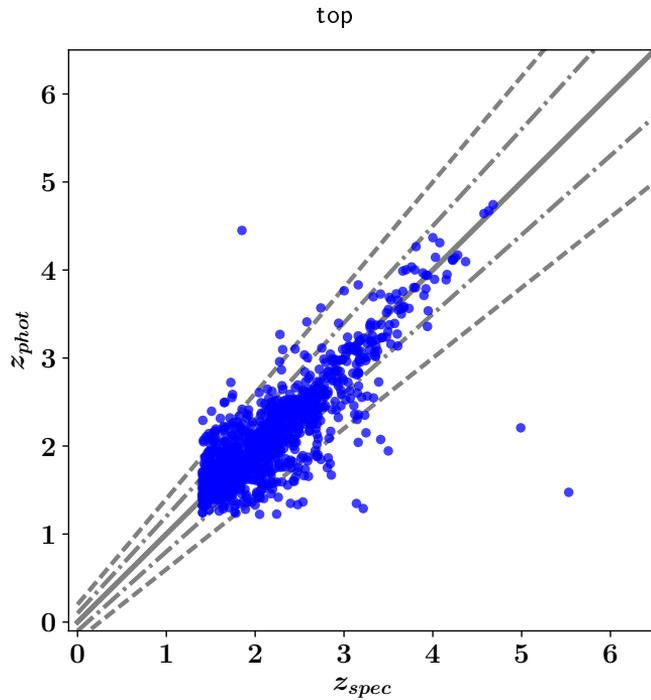


Figure 5.3: Photometric z_{photo} versus spectroscopic z_{spec} redshift for spectroscopic quasars in the NDWFS-Boötes region. The photometric redshifts are obtained using the NW kernel regression. The grey line indicates the one-to-one $z_{NW} = z_{spec}$ relation, and the dashed-dotted and dashed lines indicate the $z_{NW} - z_{spec} = \pm 0.10 \times (1 + z_{spec})$ and $z_{NW} - z_{spec} = \pm 0.20 \times (1 + z_{spec})$ relations, respectively.

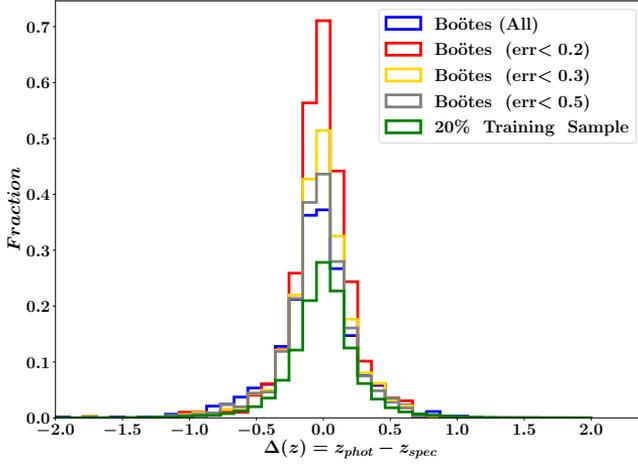


Figure 5.4: Normalized histogram of the bias $\Delta z = z_{\text{phot}} - z_{\text{spec}}$ between photometric and spectroscopic redshifts for different samples. The photometric redshifts are obtained using the NW kernel regression. Around 76% of the spectroscopic quasars in the Boötes field have photometric redshifts that are correct within $|\Delta z| = 0.3$ (see Table 5.4).

the difference Δ_{mag} between the PSF and cMODEL_{MAG} magnitudes in the SDSS-*r* band as a function of redshift. Spectroscopic quasars are binned according to their redshifts to calculate the magnitude difference as the quantile that contains 95% of the quasars in the corresponding bin. The redshift intervals match those utilized to derive the luminosity function in Section 5.5.4. Objects with magnitude differences less than the Δ_{mag} value in their corresponding redshift bin are considered to be point sources, and are included in our RSQs sample. Figure 5.6 shows the magnitude differences as a function of redshift. Finally, we restrict our RSQs sample to magnitudes $16.0 \leq i_{\text{PS}} \leq 23.0$ to avoid extrapolation beyond the range of the quasar training sample (see Table 5.2). The resulting catalog consists of 17924 objects, with 104 sources having a LOFAR counterpart within a radius of $2''$.

The final RSQs catalog resulting from our selection contains 51 photometric quasars and 83 spectroscopic quasars. In Figure 5.7, we compare the colors of the photometric quasars with those of the training sample and Boötes spectroscopic quasars as func-

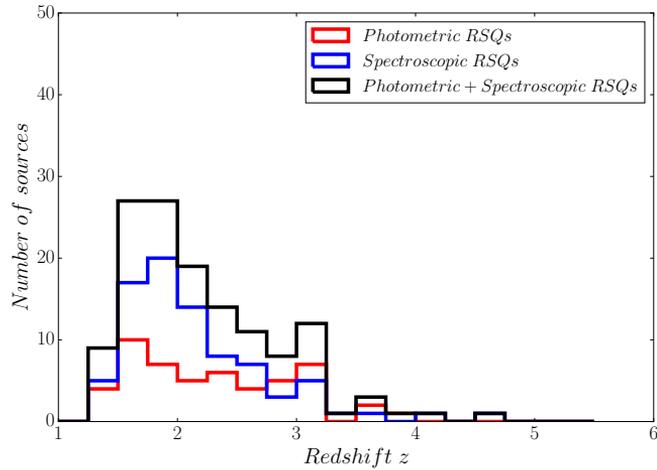


Figure 5.5: The redshift distribution of photometric (red) and spectroscopic (blue) RSQs. Also, the combined redshift distribution (black) of photometric and spectroscopic RSQs is plotted.

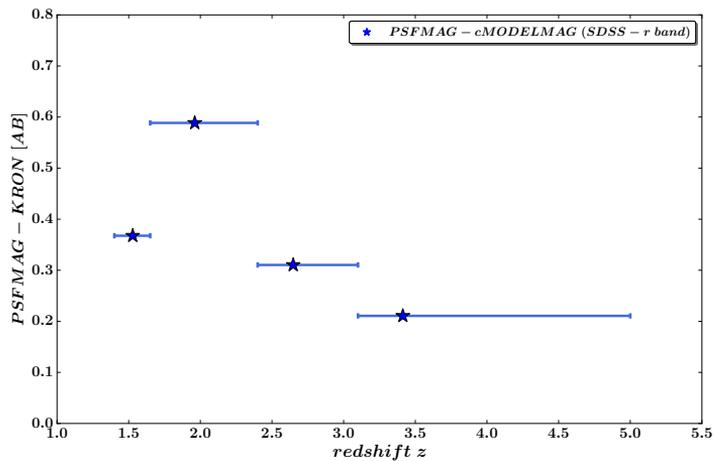


Figure 5.6: The difference between the PSF and cMODEL MAG magnitudes in the r band to separate point-like and extended sources as a function of redshift. The difference values are calculated considering the quantile that contains 95% of the quasars in the corresponding redshift bin. The redshift intervals match those used to derive the luminosity function in Section 5.5.4.

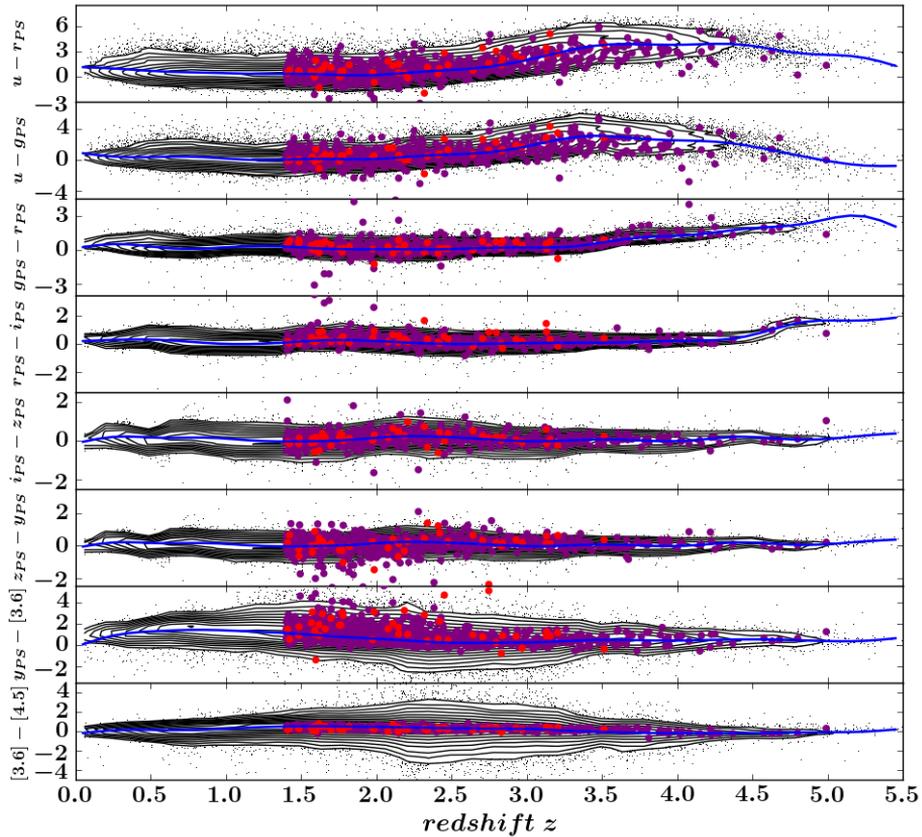


Figure 5.7: Quasar colors versus redshift for different quasar samples between $z = 1.4$ and $z = 5.5$. *Red points*: RSQs with photometric redshifts; *purple points*: Boötes spectroscopic quasars; *black contours* and *points*: the color distributions of the quasar training sample from Section 5.2.3; *blue lines*: mean color–redshift relations derived from the quasar training sample.

tions of redshift. The color-redshift spaces occupied by photometric RSQs are in good agreement with those of Boötes spectroscopic quasars and the quasar training sample. Figure 5.5 displays the redshift distribution of photometric and spectroscopic quasars. At $z \lesssim 2.8$, the majority of the quasars in our sample are spectroscopic. Considering that 77% of the photometric quasars have photometric errors smaller than $err \leq 0.5$, we conclude that the accuracy of their photometric redshifts is similar or slightly worse to that of the Boötes sample with photometric errors that are $err \leq 0.5$ (see Table 5.4). The i_{PS} -band magnitude and 150 MHz flux distributions of the RSQs sample are presented in Figure 5.8, while the absolute magnitude and radio luminosity are displayed in Figures 5.9 and 5.10, respectively. The absolute magnitudes are calculated using the K-correction discussed in Section 5.5.2. The majority of RSQs (134 in total) in our sample are unresolved or resolved into single-component radio sources in the LOFAR-Boötes mosaic with a resolution of $\sim 5''$. In our sample, only 11 quasars show present radio-morphologies consistent with a core-jet structure. Appendix 5.A presents a selection of false-color RGB images of spectroscopic and photometric RSQs from our final sample.

5.4.5 LOFAR and wedge-based mid-infrared selection of quasars

In this section, we compare the LOFAR and wedge-based mid-infrared selection for objects classified as quasars by the ML classification algorithms. Firstly, we use the mid-infrared color cuts proposed by Lacy et al. (2007) and Donley et al. (2012) to identify the presence of AGN-heated dust in our photometric RSQs. Figure 5.11 shows the mid-infrared colors for different quasar samples in the NDWFS-Boötes field. The mid-infrared colors of the photometric RSQs are in good agreement with those of spectroscopic quasars, and the majority of both spectroscopic and RSQs are located within the region delimited by the Donley et al. (2012) color cuts. To be exact, 89.93% (70.59%) of

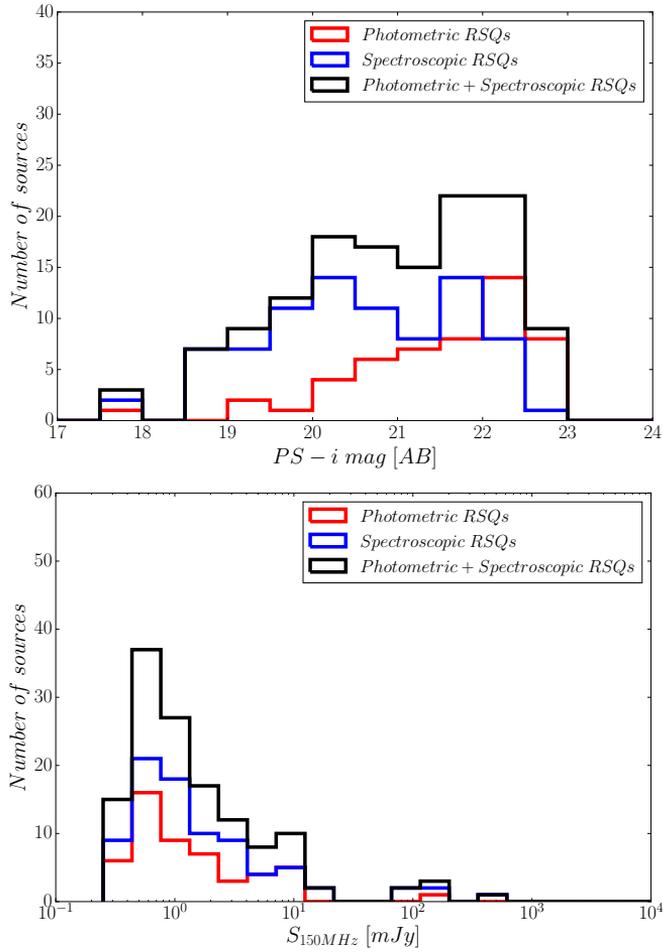


Figure 5.8: The i_{PS} (top) and total flux $S_{150\text{MHz}}$ (bottom) distributions of photometric (red) and spectroscopic (blue) RSQs. Also, the combined redshift distribution (black) of photometric and spectroscopic RSQs is plotted. The method employed to select the quasars is described in Section 5.3.3.

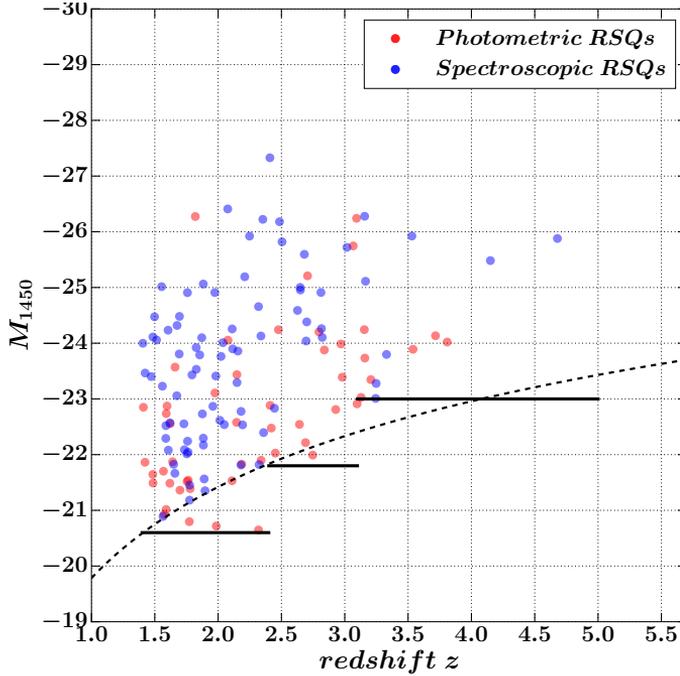


Figure 5.9: The absolute magnitude M_{1450} versus redshift for photometric (red) and spectroscopic (blue) RSQs in our sample. To minimize incompleteness due to incomplete M_{1450} bins while retaining the maximum numbers of quasars for estimating the luminosity function, we consider only RSQs with $M_{1450} \leq [-20.6, -21.8, -23.0]$ at $1.4 \leq z < 2.4$, $2.4 \leq z < 3.1$, and $3.1 \leq z < 5.0$; respectively. The dashed line denotes the magnitude limit $i_{\text{PS}} = 23.0$. This limit is calculated assuming a quasar continuum described by a power-law with slope $\alpha = -0.5$ with no emission line contribution or Ly_α forest blanketing.

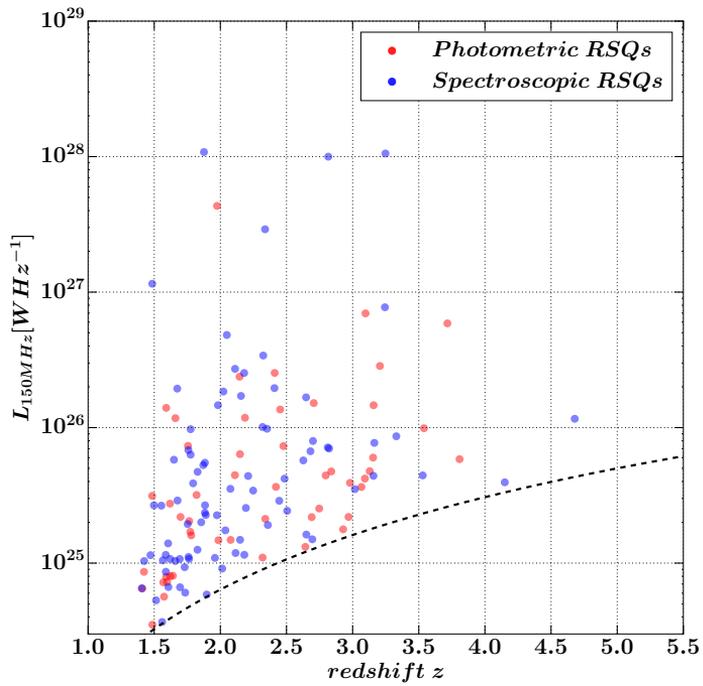


Figure 5.10: The 150 MHz rest frame absolute luminosity density versus redshift for photometric (red) and spectroscopic (blue) RSQs in our sample. The solid line denotes the 5σ flux limit ($275\mu\text{Jy}$) of the Boötes observations presented by Retana-Montenegro et al. (2018a).

the 1192 (51) spectroscopic quasars (photometric RSQs) with redshifts $z > 1.4$ are located within the region delimited by the Donley et al. (2012) color cuts, 98.15% (94.12%) reside in the region common to the Lacy et al. (2007) and Donley et al. (2012) color cuts, and only 1.85% (5.88%) are located outside the boundaries delimited by the aforementioned wedge-based mid-infrared color cuts. Considering the following points regarding our photometric RSQs: i) these objects are identified as quasar by the ML algorithms; and ii) their optical and mid-infrared colors are similar to those of spectroscopic quasars. We are confident therefore that our sample of photometric RSQs is composed mainly of real quasars and the number of contaminants is minimal. Moreover, these points show that utilizing ML algorithms trained using optical/infrared photometry and combined with LOFAR data is a very efficient and robust way to identify quasars.

To investigate the wedge-based mid-infrared selection of quasars without radio detections, we first need to establish the nature of these objects as robust quasars candidates or contaminants (i.e. stars or galaxies). For this purpose, we follow a classification method based on the goodness-of-fit, χ^2 , estimated through the fitting of the object's photometric data to a given type of spectral energy distribution (SED) template (quasar, galaxy, and stellar) (e.g. Ilbert et al. 2006; Masters et al. 2015). This fitting assigns to each object three χ^2 values: χ^2_{QSO} , χ^2_{GAL} , and χ^2_{STAR} , which corresponds to the fitting of the object photometry against the quasar, galaxy, and stellar templates, respectively. We perform these fittings using the photometric redshift code EAZY (Brammer et al. 2008), and utilize only the SDSS, Pan-STARRS1, and IRAC 3.6 μm /4.5 μm bands as this was the photometry used to classify them originally as quasars by the ML algorithms. This also represents a scenario where there is SDSS, Pan-STARRS1, and LOFAR coverage, but shallow or incomplete mid-infrared data to perform a wedge-based mid-infrared selection of faint quasars (e.g. Stern et al. 2005; Messias et al. 2012). The three template libraries used in this analysis are as follows. For the galaxy library, we use the latest version of the Brammer et al. (2008) SED templates that include nebular emission lines; while for the star library we select the Chabrier et al. (2000) SED templates. For the quasar template set, we use

a compilation of quasar templates from the literature. This includes the composite quasar templates by Cristiani & Vio (1990), Vanden Berk et al. (2001) and Gavignaud et al. (2006), respectively; and the type-1 quasar templates by Salvato et al. (2009) (pl_l22491_10_TQSO1_90, pl_QSOH, pl_QSO_DR2, pl_TQSO1). To account for dust extinction in the galaxies and quasar hosts, we modify the galaxy and quasar templates using the Calzetti et al. (2000) starburst and Pei (1992) Small Magellanic Cloud (SMC) extinction laws. Additionally, and only for quasar templates, we use the Czerny et al. (2004) quasar extinction curve. The extinction is applied to each template according to a grid $E(B - V) = [0.0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5]$. For all the SED libraries, the zero points are calculated using the standard procedure of fitting the observed SEDs of a sample of spectroscopic objects, fixing their redshift to the known spectroscopic redshift (see Ilbert et al. 2006 for more details). The spectroscopic samples used to determine the zero points are the quasars, galaxies, and stars of the training sample presented in Section 5.3.1. Due to the large number of templates in each template library, only single template fits are considered. From the χ^2 distribution of spectroscopic quasars in Boötes, we define empirical χ^2 cuts to separate quasars from stars and galaxies. Figure 5.12 displays the comparison of the χ^2 values resulting from the quasar, galaxy, and star template fitting to the photometry of spectroscopic and photometric quasars in the NDWFS-Boötes field. We find that the empirical cuts $\chi_{\text{STAR}}^2 \geq 22$, $\chi_{\text{QSO}}^2 \leq \chi_{\text{STAR}}^2 \times 0.33 + 8.33$ and $\chi_{\text{GAL}}^2 \geq 2.5$ select the majority of quasars and reject an important fraction of likely stars and galaxies. Based on the cuts described before, we select 957 total of 1193 spectroscopic quasars; and 4935 of 17829 photometric quasars. From these, 1423 photometric quasars are located in the region delimited by the Donley et al. (2012) color cuts. The analysis of the mid-infrared selection is limited to the Donley et al. (2012) wedge, as it is expected that the majority of quasars will be located within this region.

Next, we compare the number of photometric quasars selected by the Donley et al. (2012) wedge to the expected number of quasars. The expected number of quasars can be determined using the following expression:

$$N_{\text{QSO}} = A \times \iint \Phi^*(M_{1450}^*, z) V_c(z) dz dM_{1450}, \quad (5.7)$$

where A is the survey area, $\Phi^*(M_{1450}^*, z)$ is the quasar luminosity function, and $V_c(z)$ is the comoving volume. We estimate the number of total (radio detected and undetected) faint quasars using the results by Yang et al. (2018) and eq. 5.7. These authors studied the luminosity function of faint quasars between $0.5 < z < 4.5$ in a 1.0 deg^2 field within the VIMOS VLT Deep Survey (Le Fèvre et al. 2013). For redshifts $z > 3.5$, Yang et al. (2018) did not fit the luminosity function due to the low number of quasars in their sample, therefore, our analysis is limited to the redshift range $1.4 \lesssim z \lesssim 3.5$. According to Yang et al. (2018), the expected number of faint quasars at $1.4 \lesssim z \lesssim 3.5$ in a 9.3 deg^2 region like the NDWFS-Boötes field is approximately $N_{\text{QSO}} \sim 1060$. However, in the redshift range $1.4 \leq z \leq 3.5$ there are more than 1100 spectroscopic quasars, and 1380 photometric quasars fulfill the Donley et al. (2012) color cuts. Counting, only the photometric quasars and not known spectroscopic quasars, the number is higher than the expected number of quasars. This indicates some degree of contamination in the sample of photometric quasars selected using the wedge-based mid-infrared selection. Most likely these contaminants are compact galaxies or AGNs that mimic the colors of quasars, as the majority of stars are expected to be located outside the wedge define by the Donley et al. (2012) color cut. Additionally, comparing these numbers against those of the LOFAR selection, it is likely that there are more contaminants in the wedge-based mid-infrared selected sample than in the LOFAR selected sample. For a case like this, the addition of Euclid (Laureijs et al. 2011) near-infrared imaging (JH bands) in the ML classification process will be useful in eliminating some of these likely contaminants.

In summary, the results of this section demonstrate that in the cases where a lack of deep and complete mid-infrared coverage needed to perform a wedge-based mid-infrared selection of quasars, ML algorithms trained with optical/infrared photometry combined with LOFAR data provide an efficient and robust approach for obtaining samples of quasars.

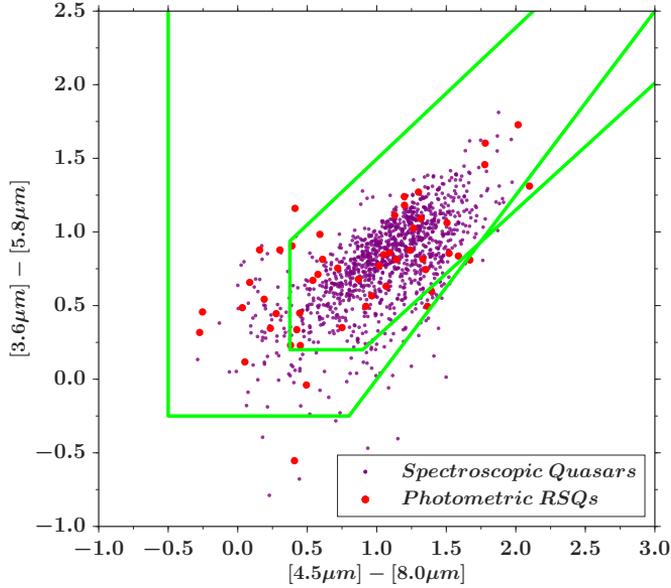


Figure 5.11: Mid-infrared colors for photometric and spectroscopic quasars in the Boötes field. The photometric RSQs are plotted as red circles, while the spectroscopic quasars are indicated by purple circles. The light green lines are the mid-infrared color cuts proposed by Lacy et al. (2007) and Donley et al. (2012).

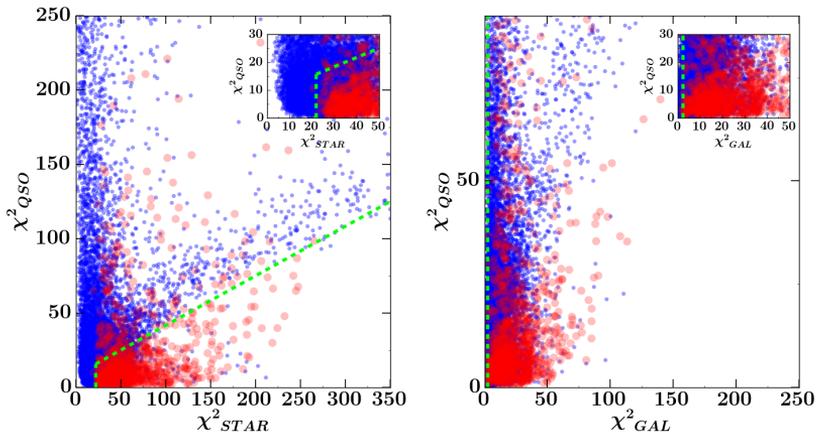


Figure 5.12: Comparison of the χ^2 values for quasar, galaxy, and star template fitting to the photometry of spectroscopic quasars and candidate quasars in the NDWFS-Boötes field.

5.5 Luminosity Function of radio-selected quasars

In the following subsections, we describe the steps required to compute the luminosity function of RSQs.

5.5.1 Selection completeness and accuracy of photometric redshifts

An important step in measuring the luminosity function of quasars is to account (and correct) for the different sources of incompleteness that could bias the quasar counts. In our analysis, we need to consider the completeness of our sample selection and the accuracy of the photometric redshifts. The selection completeness, $P_{\text{comp}}(i_{\text{PS}}, z)$, is the fraction of quasars that were successfully identified as quasars by the classification algorithms, as function of magnitude i_{PS} and redshift. $P_{\text{comp}}(i_{\text{PS}}, z)$ is derived from the data themselves as follows. First, the spectroscopic quasar sample introduced in Section 5.3.1 is binned according to magnitude i_{PS} and redshift z , in bins of size $\Delta i_{\text{PS}} = 0.5$, and $\Delta z = 0.3$, respectively. The quasars in each bin are separated into two subsamples. The first subsample is created by randomly sampling without replacement all quasars in the bin, while the second subsample includes all the quasars that were not sampled. The sizes of the first and second subsamples are 20% and 80% of quasars in the $i_{\text{PS}} - z$ bin, respectively. Having done this for all bins, the corresponding samples are combined to create final training (80%) and target (20%) samples. The final result is the uniform and randomly sampled separation of the training sample into two subsamples in the $i_{\text{PS}} - z$ plane. The main advantage of this binning scheme is that it provides an unbiased and efficient way to map locally the selection completeness obtained using the classification algorithms as a function of magnitude and redshift. The second subsample is used as the training sample for the classification algorithms, while the first subsample has the role of target sample and it is employed to derive the selection completeness.

In addition to correcting for selection incompleteness in our sample, we also need to correct for the accuracy of the photometric redshifts determined using the NW regression

method. For this purpose, we determine the expected number of spectroscopic quasars to have photometric redshifts correctly and incorrectly assigned within a redshift bin using the NW method. This is done following a similar approach to determining the selection completeness. First, the spectroscopic quasar sample introduced in Section 5.3.1 is divided into the same redshift bins used to derive the luminosity function in Section 5.5.4. In each bin, the quasars contained in that bin are randomly separated to create samples with sizes of 20% and 80% of all quasars in the bin, respectively. The corresponding samples from all the bins are combined to create final training (80%) and target (20%) samples. The training sample is used to train the NW regression method, while the target sample is utilized to determine the expected number of spectroscopic quasars with correctly and incorrectly assigned redshifts within the boundaries of the redshift bins. The ratio between the number of spectroscopic quasars with correctly and incorrectly assigned redshifts, $f_{\text{photo-z}}$, provides an estimate of the excess of photometric quasars with incorrectly assigned photometric redshifts within a redshift bin. This ratio is used as a correction factor for each photometric quasar within the corresponding redshift bin. The derived correction factors have a median factor of $f_{\text{photo-z}} \simeq 1.0$, with the $1.65 < z < 2.4$ redshift bin having the smallest value with $f_{\text{photo-z}} = 0.90$.

5.5.2 Simulated Quasar Spectra

In order to calculate the the K-correction (see Section 5.5.3), we construct a synthetic quasar library that is an accurate representation of the quasar demography. The variety in the quasar spectral features (UV continuum slope, emission-line EW, and intervening HI absorbers along the line of sight) determine the range in quasar colors. It is important that these spectral features are taken into account to obtain reliable simulated quasar spectra. These spectra are later incorporated into a synthetic quasar library that allow us to compute the K-correction for a given redshift. We explain below the procedure followed to build our synthetic quasar library

Following several authors (Møller & Jakobsen 1990; Fan 1999; Richards et al. 2006; McGreer et al. 2013), the synthetic quasar spectra are built using a Monte-Carlo (MC) approach. We perform MC simulations to generate quasar spectra adopting a broken

power-law ($f_\lambda \propto \lambda^{-\alpha_\lambda}$) for the UV continuum at 1100 \AA . The slope values are drawn from a Gaussian distribution, with mean values of $\langle \alpha_\lambda \rangle = -1.7$ for $\lambda < 1100 \text{ \AA}$ (Telfer et al. 2002), and $\langle \alpha_\lambda \rangle = -0.5$ for $\lambda > 1100 \text{ \AA}$ (Vanden Berk et al. 2001), both with standard deviations of $\sigma = 0.30$. We bin BOSS quasars by their luminosity to obtain the distribution for the parameters (wavelength, EW, FWHM) of emission lines. This allows us to recover the intrinsic emission line mean and dispersion as function of luminosity, as well as reproducing empirical trends such as the Baldwin effect (Baldwin 1977). We again assume gaussianity when the emission line features are added to the quasar continuum. For each template spectrum, the intergalactic absorption that gives rise to the Ly $_\alpha$ forest is included by creating sightlines in a MC fashion adopting the prescription of neutral absorbers by Bershady et al. (1999). The spectrum is then convolved with our filter passbands to obtain the colors for each synthetic quasar. A Gaussian error is added to the photometry of the mock quasars with a σ derived from the photometric errors of the real magnitudes that match the simulated ones. This error is combined in quadrature with the photometric calibration errors of Pan-STARRS1 (Tonry et al. 2012). In Figure 5.1, we show a synthetic spectrum from our quasar library.

5.5.3 K-correction

Usually, the luminosity functions of quasars are expressed in the absolute magnitude at rest-frame 1450 \AA , M_{1450} , which provides a good measurement of the quasar continuum in a region without strong emission lines (e.g. Richards et al. 2006; Croom et al. 2009b; Masters et al. 2012). To derive M_{1450} for the Boötes quasars, we use the apparent magnitude m_X in a fiducial filter as a proxy:

$$M_{1450} = m_X - 5 \log(d_L/10) - K_X, \quad (5.8)$$

where $d_L(z)$ is the luminosity distance in parsecs, and K_X is the K-correction which allows us to convert the magnitudes of distant objects in a given bandpass filter into an equivalent measurement into their rest-frame. Using our synthetic quasar library described in Section 5.5.2, the K-correction can be determined from the difference

between apparent magnitudes m_X and m_{1450}

$$K_X = m_X - m_{1450} - 2.5 \log(1 + z), \quad (5.9)$$

with m_{1450} calculated using a top-hat filter of width 50 \AA . Figure 5.13 displays the K-correction obtained for five different filters, and the expected result from a quasar that has only a power-law continuum and no emission line contribution or Ly_α forest blanketing. The K-correction curves between $1.0 \leq z \leq 6.0$ are obtained by calculating their average value in redshift bins of size $\Delta z = 0.1$. At $z \gtrsim 3.7$, the difference between the r_{PS} and i_{PS} bands becomes more significant as the Ly_α line moves or exits the filters. The same situation occurs at $z \gtrsim 4.7$, but for the i_{PS} and z_{PS} bands. Therefore, we estimate M_{1450} using K-corrections selected to minimize any bias caused by the redshifting of the Ly_α emission line. For $z < 3.7$ quasars, we use a K-correction based on the r_{PS} band, while for the intervals $3.7 \leq z \leq 4.7$ and $z > 4.7$ K-corrections based on the i_{PS} and z_{PS} bands, respectively, are employed.

5.5.4 Quasar Luminosity function

We construct the luminosity functions for all quasars in our radio-matched sample using the classical $1/V_{\text{max}}$ method (Schmidt 1968) for flux limited samples. The main advantage of this method is that an assumption about the underlying model of the luminosity function is not required. The estimator adopted to compute the comoving quasar density in a certain luminosity bin is:

$$\begin{aligned} \Phi(L) = & \frac{1}{\Delta L} \sum_{i=1}^n (F(S_{150\text{MHz}}) \times P_{\text{comp}}(i_{\text{PS}}, z) \\ & \times P_{\text{comp}}(i_{\text{PS}}, z) \times f_{\text{photo-z}}(z) \times V_{\text{max},i})^{-1} \end{aligned} \quad (5.10)$$

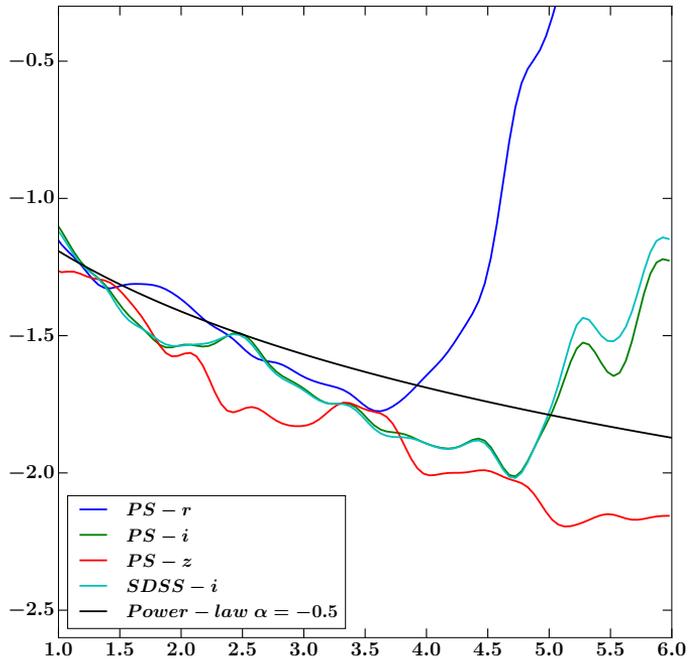


Figure 5.13: K-correction for different filters determined using our simulated quasar spectra. The Pan-STARRS1 r_{PS} and i_{PS} filters are indicated by blue and green, while the red and cyan are the expected K-corrections for the Pan-STARRS1 z_{PS} and SDSS- i bands. The solid black line is the K-correction assuming a power-law with slope $\alpha = -0.5$ with no emission line contribution or Ly_α forest blanketing. At $z \gtrsim 3.7$, the difference between the r_{PS} and i_{PS} bands becomes more significant as the Ly_α line moves in or out of the filters. The same situation occurs at $z \gtrsim 3.7$, but for the i_{PS} and z_{PS} bands.

where n is the number of quasars in the luminosity bin, $V_{\max,i}$ is the maximum comoving volume in which a quasar would be observable and included in our sample, ΔL is the luminosity bin width, $F(S_{150\text{MHz}})$ is the radio-catalog completeness of the LOFAR-Boötes mosaic (Retana-Montenegro et al. 2018a). $P_{\text{comp}}(i_{\text{PS}}, z)$ is the selection completeness, and $f_{\text{photo-z}}(z)$ is the accuracy of the NW photometric redshifts derived in Section 5.5.1. Since our quasar sample is built using a radio-optical survey, we calculate V_{\max} using the maximum redshift at which the flux of a quasar with a certain luminosity lies above the corresponding flux limit (Cirasuolo et al. 2005; Tuccillo et al. 2015), $z_{\max} = \min(z_{\max}^{\text{R}}, z_{\max}^{\text{O}})$, where z_{\max}^{R} and z_{\max}^{O} are the maximum redshifts according to the radio and optical flux limits.

We model the quasar luminosity function as a double power-law in absolute magnitude M_{1450} (Pei 1995)

$$\Phi(M_{1450}) = \Phi^*(M_{1450}^*) \times \left(10^{0.4(\alpha+1)(M_{1450}-M^*)} + 10^{0.4(\beta+1)(M_{1450}-M^*)} \right)^{-1}, \quad (5.11)$$

where M_{1450}^* is the break magnitude, Φ^* the normalization constant, α is the faint-end slope, and β is the bright-end slope. We split the quasar sample into four different redshift intervals between $1.4 < z < 5.0$ with a M_{1450} bin size equal to $\Delta M_{1450} = 1.2 \text{ mag}$. The redshift intervals and M_{1450} bin size are selected to obtain the highest S/N in the luminosity function calculations. Due to the small number of quasars in each luminosity bin ($N_{\text{QSO}} < 50$), the error bars are calculated assuming the low-statistics limit, using the 84.13% confidence Poisson upper limits and lower limits from Gehrels (1986). Finally, we use the M_{1450} lower limits indicated in Figure 5.9 to avoid incompleteness effects in the calculation of the luminosity function.

Figure 5.14 shows the luminosity function measurements in four subpanels with one for each redshift interval. We use a total of 83 spectroscopic and 51 photometric quasars to estimate the quasar luminosity function. The resulting binned luminosity functions are tabulated in Table 5.5. The luminosity function in the range $1.65 < z < 2.4$ is plotted as a reference in all the subpanels. This reference indicates that the space

Table 5.5: Binned luminosity functions for RSQs between $1.4 < z < 5.0$.

| Redshift range | z_{median} | M_{1450}^* bin center | $\log(\Phi)^a$ | σ_{low}^b | σ_{upp}^b | N^c |
|------------------|---------------------|-------------------------|----------------|-------------------------|-------------------------|-------|
| $1.4 < z < 1.65$ | 1.57 | -24.8 | -7.03 | 0.42 | 1.81 | 3 |
| | | -23.6 | -6.50 | 1.89 | 5.01 | 6 |
| | | -22.4 | -6.38 | 2.90 | 5.79 | 11 |
| | | -21.2 | -6.30 | 3.09 | 7.53 | 7 |
| $1.65 < z < 2.4$ | 1.92 | -26.0 | -7.36 | 0.22 | 0.77 | 4 |
| | | -24.8 | -6.52 | 1.99 | 4.54 | 8 |
| | | -23.6 | -6.62 | 1.82 | 3.08 | 18 |
| | | -22.4 | -6.33 | 3.58 | 5.95 | 19 |
| | | -21.2 | -6.23 | 3.42 | 6.41 | 13 |
| $2.4 < z < 3.1$ | 2.70 | -27.2 | -7.70 | 0.02 | 0.34 | 1 |
| | | -26.0 | -7.14 | 0.55 | 1.47 | 6 |
| | | -24.8 | -6.83 | 0.91 | 1.97 | 9 |
| | | -23.6 | -7.39 | 0.56 | 1.66 | 5 |
| | | -22.4 | -6.64 | 2.01 | 4.35 | 9 |
| $3.1 < z < 5.0$ | 3.25 | -27.2 | -8.13 | 0.01 | 0.24 | 1 |
| | | -26.0 | -7.04 | 0.41 | 1.78 | 3 |
| | | -24.8 | -7.63 | 0.08 | 0.54 | 2 |
| | | -23.6 | -6.99 | 0.70 | 1.44 | 10 |

Notes: ^a Φ is in units of $1 \times 10^{-7} \text{ Mpc}^{-3} \text{ mag}^{-1}$, ^b σ_{low} and σ_{upp} are in units of $1 \times 10^{-7} \text{ Mpc}^{-3} \text{ mag}^{-1}$, ^c N is the number of quasars in the corresponding M_{1450} bin.

density of RSQs is higher at $1.65 < z < 2.4$ in comparison to the other redshift intervals, i.e. the comoving space density of RSQs reaches a maximum between $1.65 < z < 2.4$. The comoving space density of RSQs is discussed with further detail in Section 5.6.3. Additionally, a good continuity between the points of the faint and bright ends is obtained in the five redshift intervals.

5.6 Results

5.6.1 Model-fitting

The observed evolution of the AGN luminosity function has been traditionally studied using luminosity (Boyle et al. 2000; Croom et al. 2009b; Richards et al. 2006), density (McGreer et al. 2013; Willott et al. 2010; Kashikawa et al. 2015), and even hybrid luminosity/density models (Ueda et al. 2003; Hasinger et al. 2005; Ross et al. 2013;

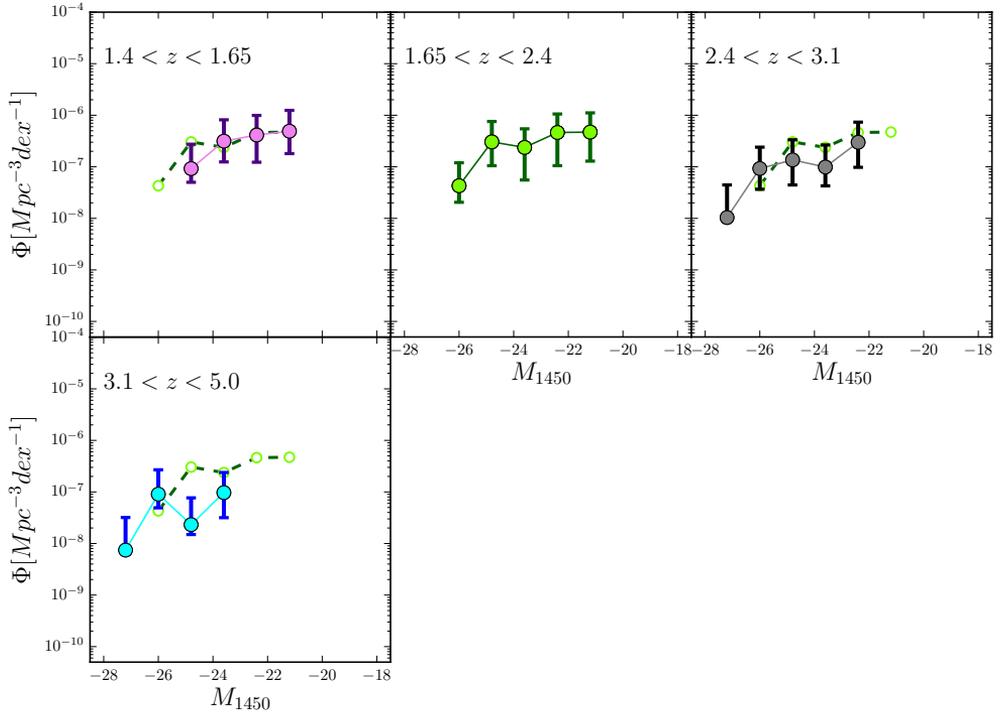


Figure 5.14: The rest-frame M_{1450} binned luminosity functions of our Boötes RSQ samples (colored circles) for five non-overlapping redshift bins between $1.4 < z < 5.0$. In each panel, we show as a reference the luminosity function at $1.65 < z < 2.4$.

Palanque-Delabrouille et al. 2016). These studies have found that the evolution of the luminosity function of quasars can be described by a pure luminosity evolution (PLE) model at $z \lesssim 2.2$ (Croom et al. 2009b), while a combined luminosity evolution and density evolution (LEDE) model can describe its evolution at $z \gtrsim 2.2$ (Croom et al. 2009b; Ross et al. 2013; Palanque-Delabrouille et al. 2016). The PLE introduces the redshift-dependence of the break magnitude using the following second-order polynomial (Croom et al. 2009b)

$$M_{1450}^*(z) = M_{1450}^*(z=0) - 2.5 (k_1 z + k_2 z^2), \quad (5.12)$$

while the LEDE model introduces the redshift-dependence in the normalization and break magnitude using the following log-linear ansatz:

$$\log(\Phi^*) = \log[\Phi^*(z=z_p)] + c_1 (z - z_p), \quad (5.13)$$

$$M_{1450}^*(z) = M_{1450}^*(z=z_p) + c_2 (z - z_p), \quad (5.14)$$

where z_p is the pivot redshift. Following previous works (e.g. Ross et al. 2013), we employ the PLE model to fit our binned luminosity function for redshift intervals $z < 2.4$, while at $z > 2.4$ we use the LEDE model with $z_p = 2.4$.

We use the χ^2 minimization to fit the luminosity function data points in each redshift bin to the corresponding models described above. Because of the relatively small area ($\sim 9.3 \text{ deg}^2$) of the Boötes field, there are only a few bright quasars in our sample. This implies that the bright-end slope β will be determined with high-uncertainty due to small number statistics. Therefore, we fix the bright-end slope β to the values reported by Ross et al. (2013) in their study of the quasar luminosity function using SDSS-DR9/BOSS data (Ahn et al. 2012). Additionally, we fix the parameters (k_1, k_2) and (c_1, c_2) in the PLE and LEDE models, respectively, to the values obtained by Ross et al. (2013). The parameter values from Ross et al. (2013) are chosen to match our redshift intervals. For the LEDE model, we use the parameter values corresponding to their S82 sample. Using the parameters from their DR9 sample produces similar results. Finally,

the best-fit parameters and their associated uncertainties are summarized in Table 5.6. The corresponding best-fit model is shown with a colored line in each subpanel of Figure 5.15. The models have a good agreement with the binned luminosity function.

5.6.2 Comparison to previous works

In Figure 5.15, we also compare our best-fit models with previous works. The SDSS-III/BOSS luminosity function (Ross et al. 2013) was estimated at $2.2 < z < 3.5$ employing a uniform sample of 22301 quasars over an area of 2236 deg^2 . Additionally, Ross et al. (2013) investigated the evolution of the QLF using a combination of SDSS (Richards et al. 2006), boss21+MMT and BOSS Stripe 82 datasets to over a redshift range of $0.3 < z < 4.75$. Ross et al. (2013) fitted their QLF data using a PLE model at $z < 2.2$, while their fittings at $z > 2.2$ were carried out employing a LEDE model. Furthermore, we compare our results with previous surveys of faint quasars ($M_{1450} < -22.0$) (Siana et al. 2008; Glikman et al. 2011; Giallongo et al. 2015; Masters et al. 2015; Akiyama et al. 2018; McGreer et al. 2018). The survey area of these studies ranges from 170 arcmin^2 (Giallongo et al. 2015) to 339.8 deg^2 (Akiyama et al. 2018). Finally, we also consider the results of SDSS and 2QZ (Croom et al. 2001, 2004) radio-selected samples (Cirasuolo et al. 2005; McGreer et al. 2009; Tuccillo et al. 2015) using FIRST (Becker et al. 1995). It is clear that the number density of our RSQs at all redshifts considered is lower in comparison with that of samples of optically bright and faint quasars, which are composed of both radio-detected and radio-undetected objects. Naturally, the number density of RSQs is higher than the density of SDSS/2QZ FIRST-selected samples (Cirasuolo et al. 2005; McGreer et al. 2009; Tuccillo et al. 2015), which are composed mainly of radio-loud quasars with fluxes $S_{1.4\text{GHz}} > 1.0 \text{ mJy}$. This flux limit corresponds to a LOFAR flux of $S_{150\text{MHz}} > 4.80 \text{ mJy}$, assuming a spectral index of $\alpha = -0.7$. As can be seen in the bottom panel of Figure 5.8, the fraction of quasars fluxes with $S_{150\text{MHz}} > 4.80 \text{ mJy}$ is just 18% of the total number of RSQs in our sample. Therefore, the lower number densities presented by SDSS/2QZ FIRST-selected samples

Table 5.6: Parametric model best-fit parameters and uncertainties. See Section 5.5.4 for more details about the models used.

| Model | Redshift range | α | β | $M_{1450}^*(z=0)$ | k_1 | k_2 | Φ^* |
|-------|------------------|------------------|--------------|---------------------|----------|----------|---|
| | | [faint-end] | [bright-end] | [mag] | | | [$\text{Mpc}^{-3} \text{mag}^{-1}$] |
| PLE | $1.4 < z < 1.65$ | -1.13 ± 0.01 | -3.55 | -20.99 ± 0.01 | 1.293 | -0.268 | $3.28 \times 10^{-7} \pm 4.03 \times 10^{-9}$ |
| PLE | $1.65 < z < 2.4$ | -1.18 ± 0.15 | -3.55 | -21.87 ± 0.43 | 1.293 | -0.268 | $2.25 \times 10^{-7} \pm 8.80 \times 10^{-8}$ |
| PLE | $1.4 < z < 2.4$ | -1.19 ± 0.16 | -3.55 | -20.99 ± 0.25 | 1.293 | -0.268 | $3.61 \times 10^{-7} \pm 1.15 \times 10^{-7}$ |
| | | | | $M_{1450}^*(z=2.4)$ | c_1 | c_2 | |
| LEDE | $2.4 < z < 3.1$ | -1.40 ± 0.16 | -3.51 | -26.98 ± 1.38 | -0.689 | -0.809 | $6.17 \times 10^{-8} \pm 6.20 \times 10^{-8}$ |
| LEDE | $3.1 < z < 5.0$ | -1.13 ± 0.65 | -3.51 | -26.41 ± 2.06 | -0.689 | -0.809 | $1.40 \times 10^{-7} \pm 2.41 \times 10^{-7}$ |
| LEDE | $2.4 < z < 5.0$ | -1.29 ± 0.18 | -3.51 | -26.12 ± 0.58 | -0.689 | -0.809 | $1.63 \times 10^{-7} \pm 1.07 \times 10^{-7}$ |

Note: Parameters without errors are kept fixed during the fitting.

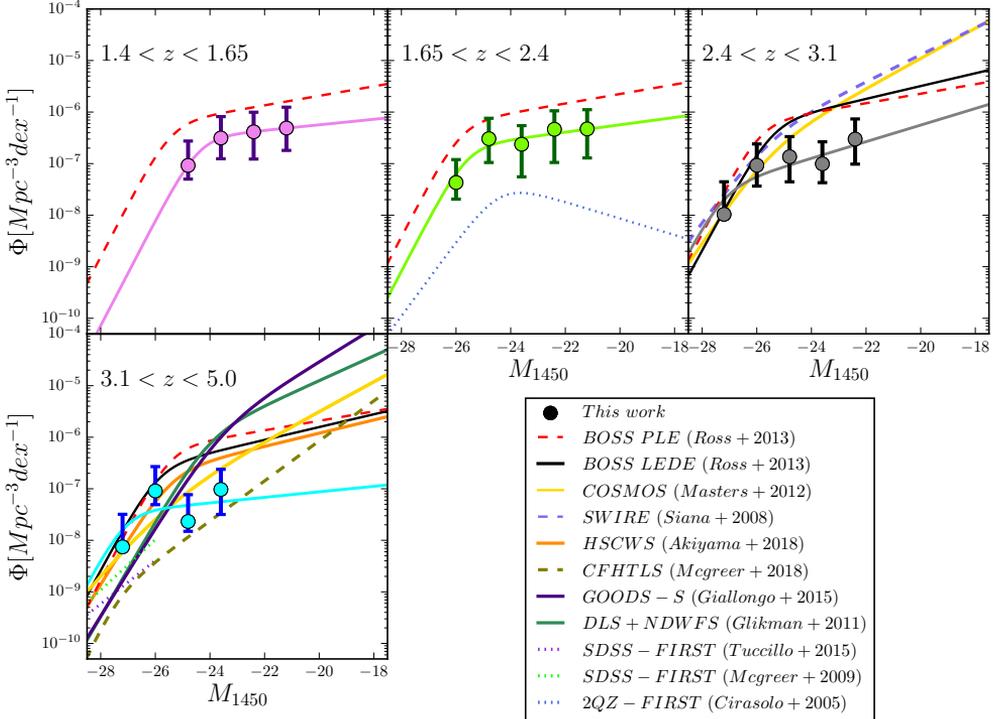


Figure 5.15: The rest-frame M_{1450} binned luminosity functions of our Boötes RSQ samples (colored circles) for four non-overlapping redshift intervals between $1.4 < z < 5.0$. The lines show the corresponding best-fit models in each redshift bin. The best-fitting parameters for each fit are presented in Table 5.6. For comparison, we show the QLFs from previous works (Cirasuolo et al. 2005; Siana et al. 2008; McGreer et al. 2009; Glikman et al. 2011; Masters et al. 2012; Ross et al. 2013; Giallongo et al. 2015; Tuccillo et al. 2015; Akiyama et al. 2018; McGreer et al. 2018) measured over similar redshift intervals. The single-power law fits by McGreer et al. (2009) and Tuccillo et al. (2015) are plotted in the range $-29 \leq M_{1450} \leq -26$, which is the original range where they were measured.

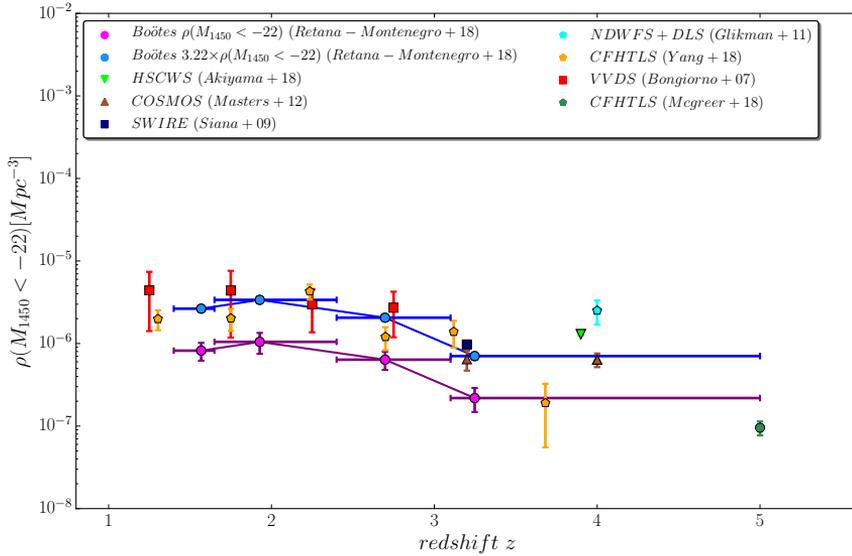


Figure 5.16: The spatial density of RSQs with $M_{1450} < -22$ as a function of redshift compared to the space density of faint quasar samples ($M_{1450} < -22$) from the literature. The spatial density of RSQs is indicated by purple circles, while estimates from the literature (Bongiorno et al. 2007; Siana et al. 2008; Glikman et al. 2011; Masters et al. 2012; Akiyama et al. 2018; Yang et al. 2018; McGreer et al. 2018) are represented by the corresponding symbols in the legend box. We also plot the spatial density of RSQs scaled by a factor of 3.22 (1/0.31) (blue circles).

are expected.

We compare our best-fit parameters with previous studies at different redshifts to constrain the evolution of the luminosity function of RSQs. In Figure 5.17, we compare the normalization constant $\log(\Phi^*)$, the break magnitude M_{1450}^* , and the faint-end slope α with previous values reported for faint quasars as a function of redshift. We also plot the PLE and LEDE models by Ross et al. (2013), as well as our PLE and LEDE models. From Figure 5.17, we find the following trends:

1. Our $\log(\Phi^*)$ values are lower in comparison with those of other samples of faint quasars (Siana et al. 2008; Glikman et al. 2011; Masters et al. 2012; Niida et al. 2016; Yang et al. 2018; Akiyama et al. 2018). For redshifts $z < 2.4$, the normalization constant within uncertainties is consistent with a PLE evolutionary trend. At $z > 2.4$, $\log(\Phi^*)$ seems to decrease following a linear-log trend reminiscent of a LEDE evolution, and similar to that of the faint quasars.
2. The break magnitude M_{1450}^* seems to get brighter with increasing redshift, a trend that is consistent with previously estimates (Siana et al. 2008; Glikman et al. 2011; Masters et al. 2015; Niida et al. 2016).
3. The faint-end slope α does become steeper with increasing redshift with a mean value of $\alpha = -1.15$ at $z < 2.4$, while at $z > 2.4$ the mean value is $\alpha = -1.26$. Our α values are consistent within error bars with those reported previously by several authors (Siana et al. 2008; Glikman et al. 2011; Niida et al. 2016; Akiyama et al. 2018; Yang et al. 2018).

5.6.3 Density evolution of RSQs

In the last 20 years, the evolution of quasar activity has been studied in the UV/optical (Fan et al. 2001; Wolf et al. 2003; Richards et al. 2006; Fontanot et al. 2007; Bongiorno et al. 2007; Croom et al. 2009b; Glikman et al. 2011; Ross et al. 2013; Masters et al. 2015; Jiang et al. 2016; McGreer et al. 2018; Yang et al. 2018), X-ray (Hasinger et al. 2005; Silverman et al. 2005; Aird et al. 2015; Georgakakis et al. 2015; Miyaji et al.

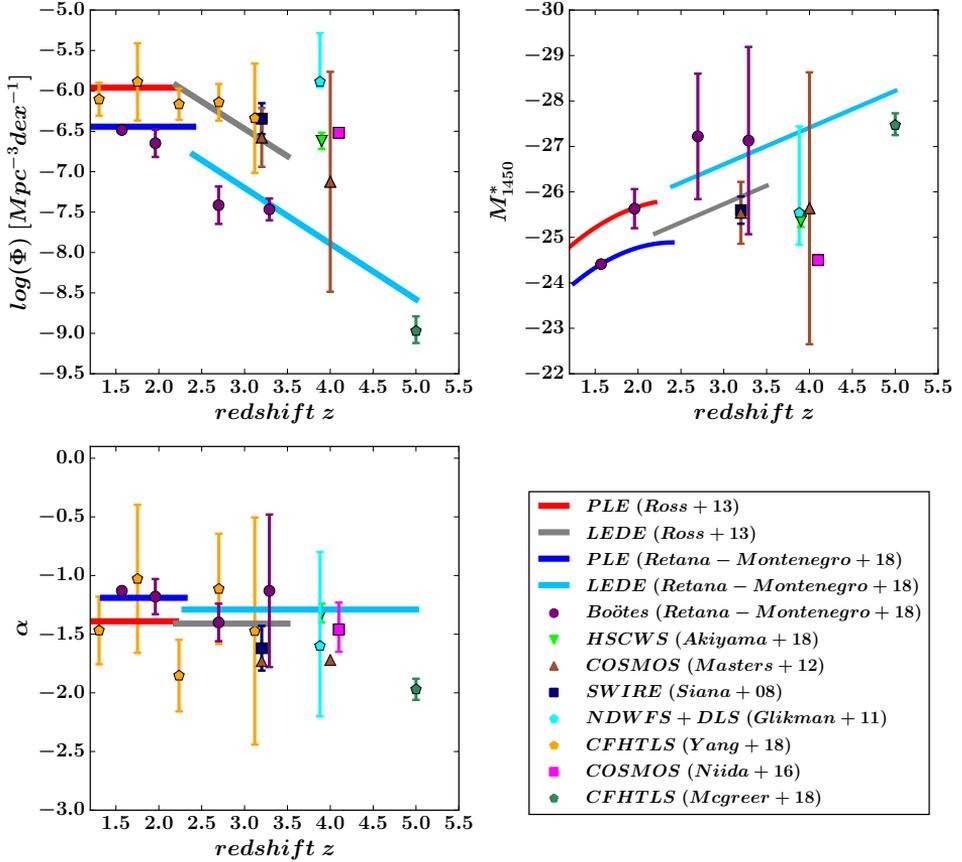


Figure 5.17: The best-fit quasar luminosity function parameters as a function of redshift. Our results are indicated by purple circles, while estimates from the literature (Siana et al. 2008; Glikman et al. 2011; Masters et al. 2012; Niida et al. 2016; Akiyama et al. 2018; Yang et al. 2018; McGreer et al. 2018) are represented by the corresponding symbols in the legend box. The red and gray lines represent the PLE and LEDE models from Ross et al. (2013), and the blue and dark cyan lines our PLE ($1.4 < z < 2.4$) and LEDE ($2.4 < z < 5.0$) models listed on Table 5.6. For clarity, we shift vertically the PLE and LEDE models from Ross et al. (2013) by a factor of +0.1 in the third panel.

2015), infrared (Brown et al. 2006; Siana et al. 2008; Assef et al. 2011), and radio (Vigotti et al. 2003; Carballo et al. 2006; Cirasuolo et al. 2006; McGreer et al. 2009; Tuccillo et al. 2015). Here, we study this evolution using the spatial density of quasars (Fan et al. 2001; McGreer et al. 2013; Tuccillo et al. 2015)

$$\rho(< M_{1450}, z) = \int_{-\infty}^{M_{1450}} \Phi(M_{1450}, z) dM, \quad (5.15)$$

where $\Phi(M_{1450}, z)$ is the luminosity function of quasars, and it is integrated over all quasars more luminous than M_{1450} . The integration is performed using the binned luminosity functions instead of using the best-fit luminosity functions, as it avoids uncertainties related to the model fitting and the extrapolation of the models. An upper limit of $M_{1450} = -22.0$ is selected for the integration, as it is the lowest luminosity limit that is common between our work and other samples of faint quasars, to compare their spatial density.

Figure 5.16 displays the space density of RSQs ($M_{1450} < -22.0$) from our sample as a function of redshift, along with other faint quasar samples from the literature. Our spatial density estimates have lower values in comparison with those of other samples. However, our spatial density values show a very similar redshift evolution to that of faint quasars, and this trend continues towards $z = 5.0$. It is clear that a peak in quasar activity occurs around $z \approx 2.0$ according to our results and the surveys carried out by Bongiorno et al. (2007) and Yang et al. (2018). This agrees with the picture provided by measurements of the luminosity function using samples of bright quasars ($M_{1450} < -24.0$) (e.g. Richards et al. 2006; Croom et al. 2009b; Ross et al. 2013). However, the quasar samples by Akiyama et al. (2018) and Glikman et al. (2011) display space densities much higher than expected in comparison with the trend suggested by our results and other samples of faint quasars (Bongiorno et al. 2007; Masters et al. 2015; Yang et al. 2018; Siana et al. 2008; McGreer et al. 2018). These higher spatial densities could be attributed to sample contamination (Akiyama et al. 2018) or cosmic variance (Glikman et al. 2011). Finally, and as expected, our values of the space density of RSQs are significantly higher in comparison to those previously estimated for RLQs. Samples of RLQs present space density values of just a few Gpc^{-3} (e.g. Vigotti et al.

2003; Tuccillo et al. 2015). This highlights the fact that deep LOFAR observations allow us to detect the radio-emission of quasars that otherwise would be classified as radio-quiet (Retana-Montenegro & Röttgering 2018).

Having computed the spatial density for the relevant samples of faint quasars and our RSQ sample, we calculate their normalized to $z \sim 2$ spatial densities as a function of redshift. In the left panel of Figure 5.18, we show the normalized spatial density of RSQs and faint quasars with $M_{1450} < -22$. It is clear that the space density of faint quasars and RSQs decreases rapidly with redshift. From its maximum at $z \sim 2$, the space density of faint quasars (RSQs) declines between $z \simeq 2$ and $z \simeq 3$ by a factor of 1.74 ± 0.84 (1.64 ± 0.41), while it reduces further from $z \simeq 2$ to $z \simeq 5.0$ by a factor of 5.23 ± 1.41 (4.80 ± 1.55). At $z \sim 1.5$, the normalized space-density ratio is 1.07 ± 0.59 (1.28 ± 0.31). Note the agreement within error bars between the evolution of the space density of RSQs and that of other faint quasar samples (Bongiorno et al. 2007; Siana et al. 2008; Masters et al. 2015; Yang et al. 2018). Considering the works by Akiyama et al. (2018) and Glikman et al. (2011), the space density of faint quasars decreases from $z \simeq 2$ to $z \simeq 5.0$ by a factor of 3.54 ± 0.87 .

Only a small fraction, less than 10%, of the quasars are classified as radio-loud (Kellermann et al. 1989; Ivezić et al. 2002; Jiang et al. 2007a). However, RLQs are often associated with massive host galaxies (Shen et al. 2009; Retana-Montenegro & Röttgering 2017, and references therein), whose radio emission is produced by large and powerful radio jets (Bridle et al. 1994; Mullin et al. 2008). However, the dependency of the radio-loud fraction (RLF) of quasars on redshift and luminosity is still a matter of debate. Some authors have found that the RLF is a strong function of luminosity (e.g. Padovani 1993; La Franca et al. 1994) and redshift (e.g. Peacock et al. 1986; Miller et al. 1990; Visnovsky et al. 1992), while others have found that it does not depend significantly on either redshift or luminosity (e.g. Stern et al. 2000; Cirasuolo et al. 2003). Kratzer & Richards (2015) found that selection effects could be biasing the conclusions about the evolution of the RLF. Using the LoTSS survey (Shimwell et al. 2019), Gurkan (2019) showed that quasars exhibit a wide continuum of radio properties, with no clear bimodality in the radio-loudness parameter.

In the context of this work, we compute the relative fraction of RSQs with respect to the spatial density of faint quasars as a function of redshift by dividing the spatial density of RSQs, $\rho_{\text{RSQs}}(z)$, by the spatial density of faint quasars (radio-detected plus radio-undetected), $\rho_{\text{QSO}}(z)$. Figure 5.18 displays the relative fraction of RSQs, $\rho_{\text{RSQs}}(z)/\rho_{\text{QSO}}(z)$, as a function of redshift. The relative fraction of RSQs considering the error bars and excluding the results of Glikman et al. (2011) and Akiyama et al. (2018) is roughly independent of redshift, with a median value of 0.31 ± 0.22 . This fraction is of 0.28 ± 0.20 considering the results of Glikman et al. (2011) and Akiyama et al. (2018). In Figure 5.18, the spatial density of RSQs is multiplied by a factor of 3.22 ($1/0.31$) to compare it with the spatial density of faint quasars. With the multiplicative factor applied, the agreement between the two spatial densities is good. Moreover, it highlights the similarity in the redshift evolution of RSQs and faint quasars up to $z \sim 5$. A fraction of ~ 0.30 of RSQs with respect to faint quasars is significantly higher than the fractions of $\sim 0.10 - 0.15$ of RLQs with respect to the whole quasar population previously estimated (e.g. Goldschmidt et al. (1999); Stern et al. (2000); Jiang et al. (2007a)). However, this is not unexpected; as previously mentioned, our deep LOFAR observations allow us to detect the radio-emission of a considerable number of quasars that otherwise would be identified as radio-quiet.

Finally, our results for the spatial density of RSQs demonstrate that the selection of quasars utilizing ML algorithms that combines optical/infrared with LOFAR observations (see Section 5.3.3.5) is very efficient and robust.

5.6.4 Contribution of RSQs to IGM Photoionization

To compute the RSQs contribution to the IGM photoionization, we consider the total quasar emissivity of hydrogen-ionizing photons, $\dot{\eta}$ [$\text{s}^{-1} \text{Mpc}^{-3}$],

$$\dot{\eta} = \epsilon_{1450} \xi_{ion}, \quad (5.16)$$

where ϵ_{1450} is the quasar emissivity at 1450 Å, and ξ_{ion} is the number of ionizing

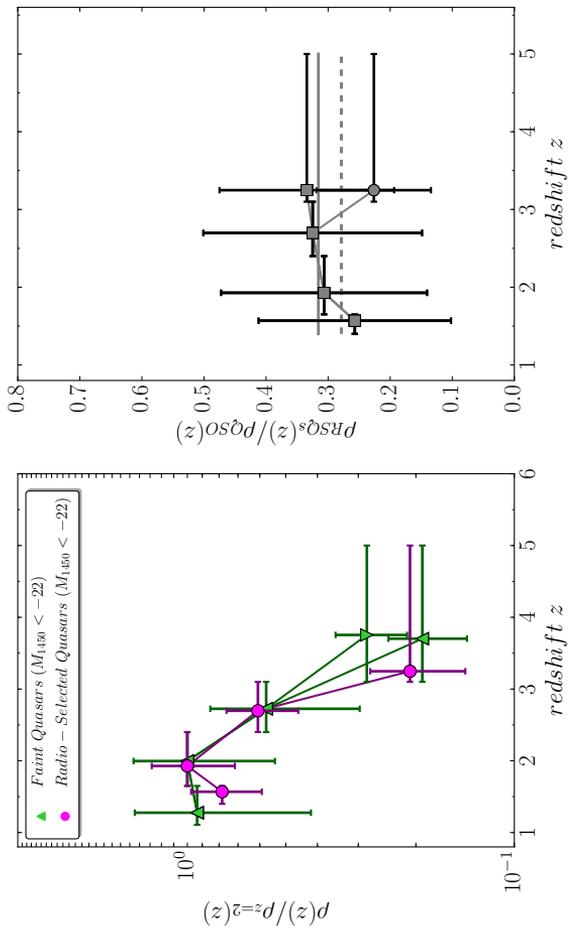


Figure 5.18: *Left panel:* Spatial densities, normalized to $z \sim 2$, as a function of redshift for optically faint quasars and RSQs. Our results are indicated by fuchsia circles, while the spatial density for faint quasars is determined from results reported in the literature (Bongiorno et al. 2007; Siana et al. 2011; Masters et al. 2012; Akiyama et al. 2018; Yang et al. 2018) and is denoted by green triangles. The upward triangle indicates the spatial density excluding the results by Glikman et al. (2011) and Akiyama et al. (2018), while the value denoted the downward triangle includes them. *Right panel:* Relative fraction of RSQs with respect to the spatial density of faint quasars as a function of redshift. The ratio is calculated between overlapping redshift bins. The gray solid line indicates the mean ratio of 0.31 excluding the works by Glikman et al. (2011) and Akiyama et al. (2018), while the dashed line denotes a mean ratio of 0.28 including these works.

photons for a source with a luminosity of $1 \text{ erg s}^{-1} \text{ Hz}^{-1}$ at 1450 \AA . The quasar emissivity (Madau et al. 1999) can be computed using the observed quasar luminosity function $\Phi(L_\nu, z)$:

$$\epsilon_\nu = \int L_\nu \Phi(L_\nu, z) dL_\nu, \quad (5.17)$$

where the integration limits are the relevant survey limiting magnitudes, and L_ν is the typical quasar SED at frequency ν . This is approximated by a broken power-law SED with slope $\alpha_\lambda = -1.7$ for $\lambda < 1100 \text{ \AA}$ (Telfer et al. 2002), and $\alpha_\lambda = -0.5$ at $1100 < \lambda < 2500 \text{ \AA}$ (Vanden Berk et al. 2001). To calculate ξ_{ion} , we integrate the quasar SED in the energy range 1-4 ryd (photons with a energy larger than 4 ryd are absorbed by Hell). The luminosity function is integrated over the luminosity range $-30 \leq M_{1450} \leq -18$. The upper limit of $M_{1450} = -18$ is chosen as it is a typical value in the literature. Additionally, a value of unity is assumed for the photon escape fraction as suggested by Grazian et al. (2018), who recently studied the Lyman Continuum escape fraction for a large sample of $z \sim 4$ faint AGNs.

Using the derived best-fit models of the luminosity function from Table 5.6, we find a total quasar emission rate of hydrogen-ionizing photons of $\dot{n} = 9.60 \times 10^{49} \text{ s}^{-1} \text{ Mpc}^{-3}$ at $z = 3.29$. We compare this value with the total photon emissivity per unit comoving volume required to ionize the universe at a given redshift (Madau et al. 1999; Fan et al. 2001)

$$\dot{N} = 10^{51.2} \left(\frac{C}{30} \right) \left(\frac{1+z}{6} \right)^3 \left(\frac{\Omega_b h^2}{0.02} \right)^2, \quad (5.18)$$

where $\Omega_b h^2 = 0.02$ is the baryon density (Spergel et al. 2007), and C is the clumping factor of the IGM. We assume that the clumping factor evolves with redshift according to the relation $C(z) = 1 + 43^{-1.71z}$ (Pawlik et al. 2009). At $z = 3.25$, the total photon emissivity required to ionize the IGM is $\dot{N} = 1.26 \times 10^{50} \text{ s}^{-1} \text{ Mpc}^{-3}$. Our quasar emission rate is approximately 76% of the rate needed to reionize the IGM at $z = 3.25$. However, taking into account the systematic uncertainties involved in the determination of the total quasar emissivity such as the escape fraction, clumping

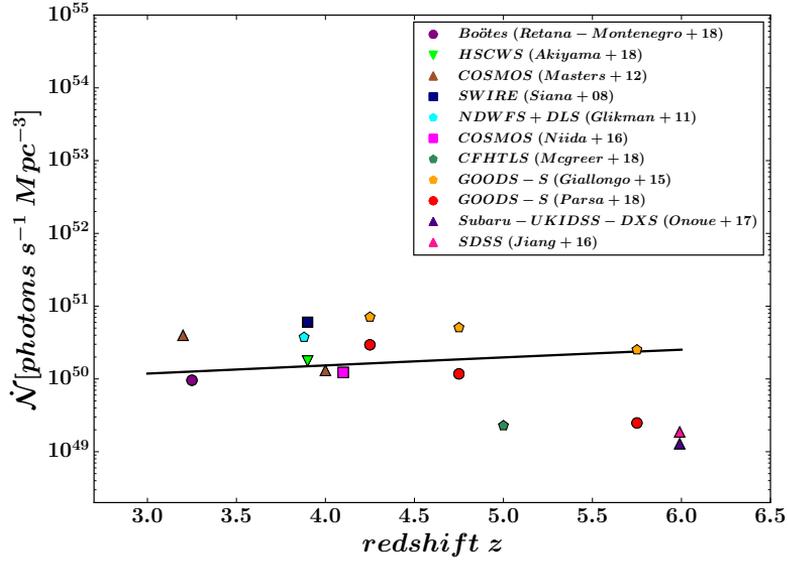


Figure 5.19: The hydrogen ionizing rate as a function of redshift. Results from this work are indicated by purple circles, while the results from the literature (Siana et al. 2008; Glikman et al. 2011; Masters et al. 2012; Niida et al. 2016; Akiyama et al. 2018; Yang et al. 2018; McGreer et al. 2018) are represented by the corresponding symbols in the legend box. The solid line represents the hydrogen ionizing rate required to ionize the IGM at any given redshift.

factor (Jiang et al. 2016), quasar SED, and the lower magnitude limit of the luminosity function, we conclude that RSQs produce only a few per cent of the photons required to ionize the IGM at $z > 3$. Additionally, as determined in Section 5.6.3, RSQs are only a fraction of the whole quasar population, therefore their spatial density is too low to be major contributors to the IGM ionization. This scenario is supported by the comparison shown in Figure 5.19 between our quasar emission rates and previous works (Siana et al. 2008; Glikman et al. 2011; Giallongo et al. 2015; Masters et al. 2015; Niida et al. 2016; Akiyama et al. 2018; Yang et al. 2018; McGreer et al. 2018). Our results agree with the majority of the previous quasar emission rates estimated in the literature and support the aforementioned conclusion.

5.7 Discussion

In this section, we discuss several aspects related to RSQs such as: the origins of their radio-emission, the environments where these objects reside, and their location in spectroscopic parameter spaces.

5.7.1 The origins of radio-emission in RSQs

An important piece of information needed to understand the origins of the radio-emission in RSQs could come from their observed radio-morphologies. Although, some of the brightest RLQs have double-lobed radio morphologies, the majority of intermediate-luminosity RLQs show core dominated radio-morphologies (de Vries et al. 2006; Lu et al. 2007; Coziol et al. 2017). Approximately, 92% of RSQs in our sample present compact radio-morphology at the resolution of the LOFAR-Boötes mosaic. A possible explanation for the origin of radio-emission in these objects lies in the interaction between outflows and the IGM, as first suggested by Stocke et al. (1992) to explain the low radio-emission in broad absorption-line quasars (BALQSOs). In this mechanism, radio-emission originates from particles accelerated on the shock fronts caused by the collision of uncollimated central outflows with the IGM of the host galaxy (Zakamska & Greene 2014; Zakamska et al. 2016). This scenario is supported by the observations of BALQSOs, as these objects have intermediate radio-luminosities and the majority present core dominated radio-morphologies (Becker et al. 2000; Liu et al. 2008; DiPompeo et al. 2011; Morabito et al. 2018), along with core-jet structures (Kunert-Bajraszewska et al. 2015) and lobes (Welling et al. 2014) in some instances. Assuming that the absorbing troughs observed in the spectra of BALQSOs are caused by uncollimated central outflows loaded into the broad emission-line region (BELR), the low-numbers of BAL systems in RLQs with double-lobed radio-morphology (DiPompeo et al. 2011; Pu 2013; Welling et al. 2014) can be attributed to the fact that in these quasars the central outflows form collimated jets which are physically separated from the BELR. In this scenario, the lack of radio-emission in quasars, traditionally classified as radio-quiet, can be explained considering that in a majority of cases the outflowing material is slowed

down by a dense interstellar clump and the formation of shock fronts is hindered. In our LOFAR-Boötes mosaic, we detect the radio-emission of many quasars that otherwise would have remained radio-undetected. The radio-emission in these quasars could have originated from the interaction between quasar outflows and the IGM. However, deeper optical and low-frequency radio surveys, in addition to LOFAR sub-arcsecond resolution observations (Varenius et al. 2015; Morabito et al. 2016), are needed to explore this mechanism in detail.

5.7.2 The environment of RSQs

A fraction of 92% of RSQs in our sample could be classified as compact steep-spectrum sources (CSS) according to their radio properties. CSS sources are usually a fraction of $\sim 10 - 30$ per cent in previous radio surveys (Peacock & Wall 1982; Fanti et al. 1990; O’Dea 1998), and they are characterized by their small projected linear sizes and median steep radio spectrum ($\alpha < -0.77$, O’Dea 1998). The brighter RSQs in our sample have a steep spectral index distribution with a median value of $\alpha \simeq -0.70$ (Retana-Montenegro & Röttgering 2018), and only 8% of RSQs in our sample present morphologies consistent with core-jet structures.

It has been suggested that CSS may be small either because they are young and still in an early stage of their evolutionary path, eventually developing into Fanaroff-Riley type-I/II (Fanaroff & Riley 1974) radio sources (Fanti et al. 1995; Alexander 2000; Snellen et al. 2000; Collier et al. 2016), or because they are embedded in a very dense environment that frustrates the propagation of the radio jets (van Breugel et al. 1984; Fanti et al. 1986, 1989; Orienti et al. 2007). The compactness of their radio-morphologies suggests that RSQs may reside in host-galaxies with a large supply of gas to fuel the early stages of quasar activity. Ultimately, these scenarios will have to be tested against high-resolution observations with submillimeter and radio interferometers that can spatially resolve the host-galaxies of RSQs and their synchrotron-dominated core/jets, respectively. These observations will, in turn, help us to shed light on the complex interplay between RSQs and their host-galaxies, and how quasar activity is triggered in these systems.

5.7.3 RSQs and their location in spectroscopic parameter spaces

The most striking features of Figure 5.18, excluding the results by Glikman et al. (2011) and Akiyama et al. (2018) from the analysis are: i) RSQs show evolutionary trends and declining factors that are similar to those presented by faint quasars ($M_{1450} < -22.0$) (see Figs. 5.16, 5.17, 5.18), and ii) the fact that RSQs may compose to up $31 \pm 22\%$ of the total faint quasar population, a fraction that within uncertainties is independent of redshift (see Fig. 5.18). Interestingly, similar decline factors in the space density of low- (Warren et al. 1994; Croom et al. 2009b; Palanque-Delabrouille et al. 2016) and high- (Schmidt et al. 1995; Kenefick et al. 1995; Palanque-Delabrouille et al. 2016) optical luminosity quasars, respectively, had been reported before. In these works, high-luminosity quasars have declining factors of $\simeq 2 - 3$ between $z \approx 2$ and $z \approx 4$, while low-luminosity quasars present steeper declining factors of $\simeq 6 - 8$ between the same redshift intervals. These factors are consistent with a downsizing evolutionary scenario. In this scenario, high-luminosity quasars evolve first at earlier epochs and reach their maximum space density at high- z , while low-luminosity quasars predominantly evolve at later epochs reaching their maximum space density at low- z (e.g. Hasinger et al. 2005; Silverman et al. 2005; Croom et al. 2009b). Similar declining factors to those of high-luminosity quasars had been reported for RLQs samples (Hook et al. 1998; Vigotti et al. 2003).

Since the early 2000's there has been substantial advances in our understanding of quasars using the broad emission line properties and their correlations. Probably, the most widely used broad emission line correlations are the eigenvector 1 (E1) and C_{IV} parameter spaces. The E1 parameter space started as the primary eigenvector in the Principal Component Analysis performed by Boroson & Green (1992), where Fe_{II} and H_{β} emission are related to line width. The generalization of this concept led to the 4D Eigenvector 1 (4DE1) parameter space, with the addition of the properties of the C_{IV} and the soft X-ray photon index (Sulentic et al. 2000a,b). The 4DE1 parameter space serves as a 4D equivalent of the 2D Hertzsprung-Russell diagrams (Hertzsprung 1909; Russell 1914). This parameter space has revealed a principal sequence of quasars char-

acterized by the Eddington ratio, and since its introduction has become an important tool for depicting the diversity of quasars and their evolutionary states (see Sulentic & Marziani 2015, and references therein). The C_{IV} parameter space (C_{IV} EW versus C_{IV} blueshift) has been used to study different quasar properties at high- z (e.g. Brotherton & Francis 1999; Sulentic et al. 2007; Richards et al. 2011; Kratzer & Richards 2015; Coatman et al. 2016). In the context of the 4DE1, E1, and C_{IV} parameter spaces, several authors (Sulentic et al. 2003, Sulentic et al. 2007, Zamfir et al. 2008, and Richards et al. 2011) have determined that RLQs and RQQs are clustered at different locations in their corresponding parameter spaces (see Fig. 14 in Richards et al. 2011 and Fig. 3 in Sulentic et al. 2003). In particular, Richards et al. (2011) and Kratzer & Richards (2015) demonstrated using the E1 and C_{IV} spaces, which may trace the relative power of radiation line-driven accretion disk winds (Richards et al. 2011), that on average RLQs present weaker radiation line-driven winds in comparison with RQQs. These authors suggest that RLQs and RQQs are two parallel evolutionary sequences, and possibly a series of spin/mergers events (Sikora et al. 2007; Sikora 2009; Schulze et al. 2017) are responsible for the triggering of radio-jets, and turning RQQs into radio-loud.

Considering that RSQs present evolutionary trends similar to those of faint quasars, and bright quasars, it is possible that faint (radio and optically) RSQs could share properties of both RLQs and RQQs. Thus, RSQs would occupy intermediate locations between RQQs and RLQs in their corresponding E1, 4DE1 and C_{IV} parameter spaces. Future spectroscopic studies of RSQs would be a major step forward towards understanding radio-loudness.

5.8 Conclusions

In this work, we train three ML algorithms: RF, SVM, and Bootstrap aggregation with optical/infrared imaging to compile a sample of quasars in the 9.3 deg^2 Boötes field. We eliminate stellar and likely galaxy contaminants from our sample by requiring a 5σ detection in deep LOFAR imaging, and applying a morphological criterium, respectively. The final sample consists of 134 spectroscopic quasars and photometric quasars in

the redshift range of $1.4 \leq z \leq 5.0$. We estimate the photometric redshifts of the photometric quasars using the NW kernel regression estimator (Nadaraya 1964; Watson 1964). When comparing the predictions of this method to the spectroscopic redshifts of 1193 Boötes spectroscopic quasars, we find that 76% of the quasars have photometric redshifts that are within $|\delta z| \leq 0.3$ of their spectroscopic redshifts. We demonstrate that in cases of lack of deep and complete mid-infrared coverage needed to perform a wedge-based mid-infrared selection of AGNs, the selection of quasars using ML algorithms trained with optical/infrared photometry in combination with LOFAR data is an effective approach for obtaining samples of quasars. We compute the fraction of quasars missed due to our selection (i.e. selection function) using a library of simulated quasar spectra. The binned luminosity function of RSQs is computed using the $1/V_{\text{max}}$ method (Schmidt 1968) in five different redshift bins between $1.4 \leq z \leq 5.0$. These luminosity functions are corrected for incompleteness due to the radio observations and selection method employed. The parametric fits to the binned luminosity function of RSQs are consistent with a PLE evolution model at $z < 2.4$, and a LEDE evolution at $z > 2.4$.

We have studied the luminosity function of RSQs down to faint luminosities of $M_{1450} = -22$. Previous studies were mostly limited to bright RLQs with luminosities $M_{1450} > -26$ (Cirasuolo et al. 2005; Carballo et al. 2006; McGreer et al. 2009; Tuccillo et al. 2015). We find evidence that suggests that the faint-end slope α is becoming steeper with increasing redshift, as found by previous measurements of the luminosity function of faint quasars (Glikman et al. 2011; Giallongo et al. 2015). Our mean values of the faint-end slope are $\alpha = -1.15$ at $z < 2.4$, while at $z > 2.4$ the mean value is $\alpha = -1.26$. We calculate the space density of RSQs over $1.4 \leq z \leq 5.0$, and find an evolutionary trend below and above the peak of their space density that is comparable to that of faint quasars. By integrating the luminosity function to $M_{1450} = -18$, we find evidence that supports a scenario where RSQs produce only a few per cent of the photons required to ionize the IGM at $z > 3$. By comparing the spatial density of RSQs with that of faint quasars at similar redshifts, we find that RSQs may compose to up $31 \pm 22\%$ of the total faint quasar population. This fraction, within uncertainties, seems to remain constant with redshift. We argue that considering the similarities in

evolutionary trends and declining factors between RSQs and faint quasars, the fainter (optically and radio) RSQs may have properties of RLQs and RQQs. Finally, we discuss several aspects of RSQs such as: the origins of their radio-emission, the environments where these objects reside, and their location in $E1$, $4DE1$ and C_{IV} parameter spaces.

Our work has demonstrated the feasibility of studying the evolution of RSQs using quasar samples compiled using the selection method proposed by Retana-Montenegro & Röttgering (2018). Future studies of the luminosity function of RSQs will benefit from the advent of the new generation of wide-field radio (LOTSS: Röttgering et al. 2011; Shimwell et al. 2017b, 2019; EMU: Norris et al. 2011), optical (LSST: Tyson 2002; LSST Science Collaboration et al. 2009; DES: Flaugher 2005), infrared (WFIRST: Spergel et al. 2013; Euclid: Laureijs et al. 2011), and spectroscopic surveys (EBOSS: Dawson et al. 2016; DESI: DESI Collaboration et al. 2016).

5.A Appendix: A sample of false color RGB ($R=B_W$, $G=R$, $B=I$) images

In this appendix, we present a sample of false color RGB ($R=B_W$, $G=R$, $B=I$) images centered on spectroscopic and photometric quasars. Each image covers $70'' \times 70''$, and the inset size is $7'' \times 7''$. The contours are $[3, 5, 7, 9, 11, 13] \times \sigma$ times the local noise level in the LOFAR (white) and FIRST (purple) images.

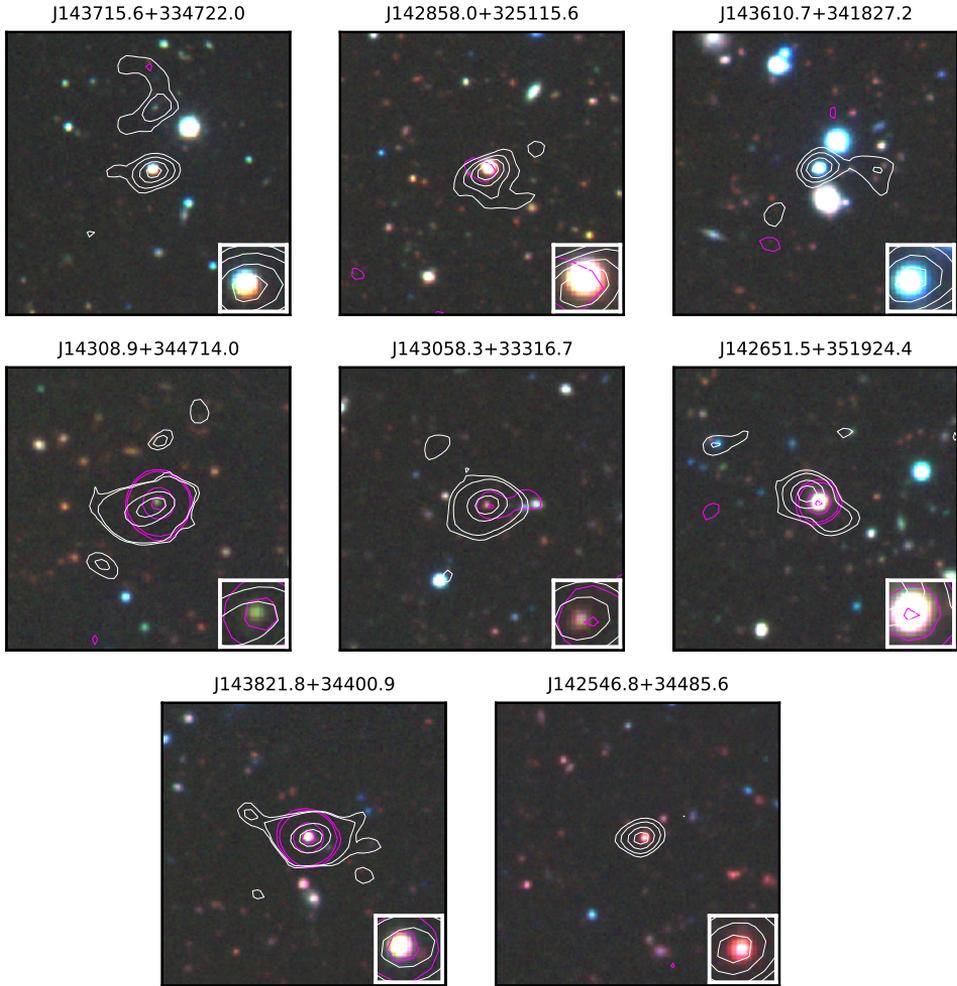


Figure 5.20: A sample of false color RGB ($R=B_W$, $G=R$, $B=I$) images centered on spectroscopic quasars. Each image covers $70'' \times 70''$, and the inset size is $7'' \times 7''$. The contours are $[3, 5, 7, 9, 11, 13] \times \sigma$ times the local noise level in the LOFAR (white) and FIRST (purple) images.

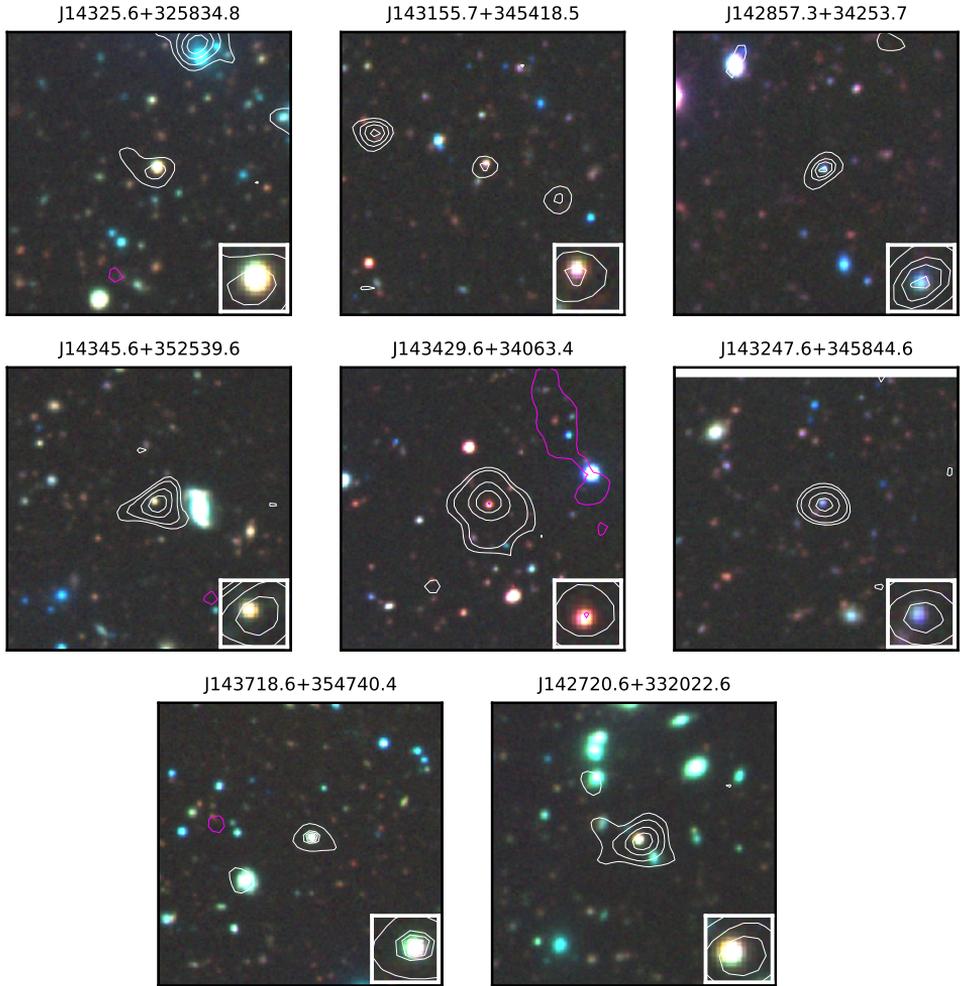


Figure 5.21: A sample of false color RGB ($R=B_W$, $G=R$, $B=I$) images centered on photometric quasars. Each image covers $70'' \times 70''$, and the inset size is $7'' \times 7''$. The contours are $[3, 5, 7, 9, 11, 13] \times \sigma$ times the local noise level in the LOFAR (white) and FIRST (purple) images.

References

- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, 235, 42
- Adelberger, K. L., Steidel, C. C., Pettini, M., et al. 2005, *ApJ*, 619, 697
- Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, *ApJS*, 203, 21
- Aird, J., Coil, A. L., Georgakakis, A., et al. 2015, *MNRAS*, 451, 1892
- Akiyama, M., He, W., Ikeda, H., et al. 2018, , 70, S34
- Alexander, P. 2000, *MNRAS*, 319, 8
- Allevato, V., Finoguenov, A., & Cappelluti, N. 2014a, *ApJ*, 797, 96
- Allevato, V., Finoguenov, A., Cappelluti, N., et al. 2011, *ApJ*, 736, 99
- Allevato, V., Finoguenov, A., Civano, F., et al. 2014b, *ApJ*, 796, 4
- Allison, R., Lindsay, S. N., Sherwin, B. D., et al. 2015, *MNRAS*, 451, 849
- Altman, N. S. 1992, *The American Statistician*, 46, 175
- Antonucci, R. 1993, *ARAA*, 31, 473
- Ashby, M. L. N., Stern, D., Brodwin, M., et al. 2009, *ApJ*, 701, 428
- Assef, R. J., Frank, S., Grier, C. J., et al. 2012, *ApJL*, 753, L2
- Assef, R. J., Kochanek, C. S., Ashby, M. L. N., et al. 2011, *ApJ*, 728, 56
- Assef, R. J., Kochanek, C. S., Brodwin, M., et al. 2010, *ApJ*, 713, 970

- Autry, R. G., Probst, R. G., Starr, B. M., et al. 2003, in Proc. SPIE, Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes, ed. M. Iye & A. F. M. Moorwood, 525–539
- Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, *Nature*, 553, 473
- Bañados, E., Venemans, B. P., Morganson, E., et al. 2015, *ApJ*, 804, 118
- Bahcall, N. A., Dong, F., Hao, L., et al. 2003, *ApJ*, 599, 814
- Baldwin, J. A. 1977, *ApJ*, 214, 679
- Barone-Nugent, R. L., Trenti, M., Wyithe, J. S. B., et al. 2014, *ApJ*, 793, 17
- Barvainis, R., Lehár, J., Birkinshaw, M., Falcke, H., & Blundell, K. M. 2005, *ApJ*, 618, 108
- Becker, R. H., White, R. L., Gregg, M. D., et al. 2000, *ApJ*, 538, 72
- Becker, R. H., White, R. L., & Helfand, D. J. 1995, *ApJ*, 450, 559
- Begelman, M. C., Blandford, R. D., & Rees, M. J. 1984, *Reviews of Modern Physics*, 56, 255
- Bershady, M. A., Charlton, J. C., & Geoffroy, J. M. 1999, *ApJ*, 518, 103
- Best, P. N., Kauffmann, G., Heckman, T. M., et al. 2005, *MNRAS*, 362, 25
- Bian, F., Fan, X., Jiang, L., et al. 2013, *ApJ*, 774, 28
- Bisnovatyi-Kogan, G. S. 2001, *Stellar physics. Vol.1: Fundamental concepts and stellar equilibrium*
- Blandford, R. D. & Znajek, R. L. 1977, *MNRAS*, 179, 433
- Blanton, M. R., Lin, H., Lupton, R. H., et al. 2003, *AJ*, 125, 2276
- Bongiorno, A., Zamorani, G., Gavignaud, I., et al. 2007, *A&A*, 472, 443
- Bonoli, S., Marulli, F., Springel, V., et al. 2009, *MNRAS*, 396, 423

- Bonzini, M., Padovani, P., Mainieri, V., et al. 2013, MNRAS, 436, 3759
- Boroson, T. A. & Green, R. F. 1992, ApJS, 80, 109
- Bovy, J., Hennawi, J. F., Hogg, D. W., et al. 2011, ApJ, 729, 141
- Boyle, B. J., Shanks, T., Croom, S. M., et al. 2000, MNRAS, 317, 1014
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503
- Breiman, L. 1996, Machine Learning, 24, 123
- Breiman, L. 2001, Machine Learning, 45, 5
- Bridle, A. H., Hough, D. H., Lonsdale, C. J., Burns, J. O., & Laing, R. A. 1994, AJ, 108, 766
- Bridle, A. H. & Perley, R. A. 1984, ARAA, 22, 319
- Bridle, A. H. & Schwab, F. R. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 180, Synthesis Imaging in Radio Astronomy II, ed. G. B. Taylor, C. L. Carilli, & R. A. Perley, 371
- Brotherton, M. S. & Francis, P. J. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 162, Quasars and Cosmology, ed. G. Ferland & J. Baldwin, 395
- Brown, M. J. I., Brand, K., Dey, A., et al. 2006, ApJ, 638, 88
- Brown, M. J. I., Webster, R. L., & Boyle, B. J. 2000, MNRAS, 317, 782
- Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
- Campbell, J., Lo, A., & MacKinlay, A. 2012, The Econometrics of Financial Markets (Princeton University Press)
- Carballo, R., González-Serrano, J. I., Montenegro-Montes, F. M., et al. 2006, MNRAS, 370, 1034
- Carrasco, D., Barrientos, L. F., Pichara, K., et al. 2015, A&A, 584, A44
- Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. 2000, ApJ, 542, 464

- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560
- Chen, T. & Guestrin, C. 2016, arXiv e-prints, arXiv:1603.02754
- Chiaberge, M. & Marconi, A. 2011, MNRAS, 416, 917
- Cirasuolo, M., Magliocchetti, M., & Celotti, A. 2005, MNRAS, 357, 1267
- Cirasuolo, M., Magliocchetti, M., Celotti, A., & Danese, L. 2003, MNRAS, 341, 993
- Cirasuolo, M., Magliocchetti, M., Gentile, G., et al. 2006, MNRAS, 371, 695
- Coatman, L., Hewett, P. C., Banerji, M., & Richards, G. T. 2016, MNRAS, 461, 647
- Cohen, A. S., Lane, W. M., Cotton, W. D., et al. 2007, AJ, 134, 1245
- Coil, A. L., Newman, J. A., Cooper, M. C., et al. 2006, ApJ, 644, 671
- Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
- Collier, J. D., Norris, R. P., Filipović, M. D., & Tothill, N. F. H. 2016, *Astronomische Nachrichten*, 337, 36
- Condon, J. J. 1992, ARAA, 30, 575
- Condon, J. J. 2007, in *Astronomical Society of the Pacific Conference Series*, Vol. 380, *Deepest Astronomical Surveys*, ed. J. Afonso, H. C. Ferguson, B. Mobasher, & R. Norris, 189
- Condon, J. J., Cotton, W. D., Fomalont, E. B., et al. 2012, ApJ, 758, 23
- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, AJ, 115, 1693
- Condon, J. J., Kellermann, K. I., Kimball, A. E., Ivezić, Ž., & Perley, R. A. 2013, ApJ, 768, 37
- Cool, R. J. 2007, ApJS, 169, 21
- Cool, R. J., Kochanek, C. S., Eisenstein, D. J., et al. 2006, AJ, 132, 823

- Cortes, C. & Vapnik, V. 1995, *Machine Learning*, 20, 273
- Cotton, W. D., Condon, J. J., Perley, R. A., et al. 2004, in *Proc. SPIE*, Vol. 5489, Ground-based Telescopes, ed. J. M. Oschmann, Jr., 180–189
- Coziol, R., Andernach, H., Torres-Papaqui, J. P., Ortega-Minakata, R. A., & Moreno del Rio, F. 2017, *MNRAS*, 466, 921
- Cristiani, S. & Vio, R. 1990, *A&A*, 227, 385
- Croom, S. M. 2011, *ApJ*, 736, 161
- Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, *MNRAS*, 356, 415
- Croom, S. M., Richards, G. T., Shanks, T., et al. 2009a, *MNRAS*, 392, 19
- Croom, S. M., Richards, G. T., Shanks, T., et al. 2009b, *MNRAS*, 399, 1755
- Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2001, *MNRAS*, 322, L29
- Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2004, *MNRAS*, 349, 1397
- Croton, D. J., Springel, V., White, S. D. M., et al. 2006, *MNRAS*, 365, 11
- Cutri, R. M. 2013, *VizieR Online Data Catalog*, 2328
- Czerny, B., Li, J., Loska, Z., & Szczerba, R. 2004, *MNRAS*, 348, L54
- da Ângela, J., Shanks, T., Croom, S. M., et al. 2008, *MNRAS*, 383, 565
- Daddi, E., Röttgering, H. J. A., Labbé, I., et al. 2003, *ApJ*, 588, 50
- Davidson, K. & Netzer, H. 1979, *Rev. Mod. Phys.*, 51, 715
- Davies, R. I., Thomas, J., Genzel, R., et al. 2006, *ApJ*, 646, 754
- Davis, M. & Peebles, P. J. E. 1983, *ApJ*, 267, 465
- Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, *AJ*, 151, 44
- de Francesco, G., Capetti, A., & Marconi, A. 2006, *A&A*, 460, 439

- de Francesco, G., Capetti, A., & Marconi, A. 2008, *A&A*, 479, 355
- de Jong, R. S., Bellido-Tirado, O., Chiappini, C., et al. 2012, in *Proc. SPIE*, Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV, 84460T
- de Vries, W. H., Becker, R. H., & White, R. L. 2006, *AJ*, 131, 666
- de Vries, W. H., Morganti, R., Röttgering, H. J. A., et al. 2002, *AJ*, 123, 1784
- de Zotti, G., Massardi, M., Negrello, M., & Wall, J. 2010, *A&A Reviews*, 18, 1
- DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, *ArXiv e-prints* [[arXiv]1611.00036]
- di Matteo, T., Springel, V., & Hernquist, L. 2005, 433, 604
- DiPompeo, M. A., Brotherton, M. S., De Breuck, C., & Laurent-Muehleisen, S. 2011, *ApJ*, 743, 71
- DiPompeo, M. A., Hickox, R. C., & Myers, A. D. 2015, *ArXiv e-prints* [[arXiv]1511.04469]
- Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *ApJ*, 748, 142
- Donoso, E., Li, C., Kauffmann, G., Best, P. N., & Heckman, T. M. 2010, *MNRAS*, 407, 1078
- Doroshenko, V. T., Sergeev, S. G., Klimanov, S. A., Pronik, V. I., & Efimov, Y. S. 2012, *MNRAS*, 426, 416
- Dunlop, J. S., McLure, R. J., Kукula, M. J., et al. 2003, *MNRAS*, 340, 1095
- Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., & Archer, S. 1959, *MEMRAS*, 68, 37
- Efstathiou, G., Bond, J. R., & White, S. D. M. 1992, *MNRAS*, 258, 1P
- Eftekharzadeh, S., Myers, A. D., White, M., et al. 2015, *MNRAS*, 453, 2779
- Einstein, A. 1916, *Annalen Phys.*, 49, 769, [,65(1916)]

- El Bouchefry, K. & Cress, C. M. 2007, *Astronomische Nachrichten*, 328, 577
- Fan, X. 1999, *AJ*, 117, 2528
- Fan, X., Narayanan, V. K., Lupton, R. H., et al. 2001, *AJ*, 122, 2833
- Fanaroff, B. L. & Riley, J. M. 1974, *MNRAS*, 167, 31P
- Fanidakis, N., Baugh, C. M., Benson, A. J., et al. 2011, *MNRAS*, 410, 53
- Fanidakis, N., Baugh, C. M., Benson, A. J., et al. 2012, *MNRAS*, 419, 2797
- Fanidakis, N., Georgakakis, A., Mountrichas, G., et al. 2013a, *MNRAS*, 435, 679
- Fanidakis, N., Macciò, A. V., Baugh, C. M., Lacey, C. G., & Frenk, C. S. 2013b, *MNRAS*, 436, 315
- Fanti, C., Fanti, R., Dallacasa, D., et al. 1995, *A&A*, 302, 317
- Fanti, C., Fanti, R., Parma, P., et al. 1989, *A&A*, 217, 44
- Fanti, C., Fanti, R., Schilizzi, R. T., Spencer, R. E., & van Breugel, W. J. M. 1986, *A&A*, 170, 10
- Fanti, R., Fanti, C., Schilizzi, R. T., et al. 1990, *A&A*, 231, 333
- Fernandes, C. A. C., Jarvis, M. J., Rawlings, S., et al. 2011, *MNRAS*, 411, 1909
- Ferrarese, L. & Merritt, D. 2000, *ApJL*, 539, L9
- Fine, S., Shanks, T., Nikoloudakis, N., & Sawangwit, U. 2011, *MNRAS*, 418, 2251
- Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. 1994, *MNRAS*, 266, 50
- Flaugher, B. 2005, *International Journal of Modern Physics A*, 20, 3121
- Flaugher, B. 2005, *International Journal of Modern Physics A*, 20, 3121
- Flesch, E. W. 2015, *PASA*, 32, e010
- Fontanot, F., Cristiani, S., Monaco, P., et al. 2007, *A&A*, 461, 39

Franzen, T. M. O., Jackson, C. A., Offringa, A. R., et al. 2016, MNRAS, 459, 3314

Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Gao, D., Zhang, Y.-X., & Zhao, Y.-H. 2008, MNRAS, 386, 1417

Gavignaud, I., Bongiorno, A., Paltani, S., et al. 2006, A&A, 457, 79

Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJL, 539, L13

Gehrels, N. 1986, ApJ, 303, 336

Georgakakis, A., Aird, J., Buchner, J., et al. 2015, MNRAS, 453, 1946

Giallongo, E., Grazian, A., Fiore, F., et al. 2015, A&A, 578, A83

Gilli, R., Daddi, E., Zamorani, G., et al. 2005, A&A, 430, 811

Gilli, R., Zamorani, G., Miyaji, T., et al. 2009, A&A, 494, 33

Glikman, E., Djorgovski, S. G., Stern, D., et al. 2011, ApJL, 728, L26

Glikman, E., Gregg, M. D., Lacy, M., et al. 2004, ApJ, 607, 60

Glikman, E., Urrutia, T., Lacy, M., et al. 2012, ApJ, 757, 51

Glikman, E., Urrutia, T., Lacy, M., et al. 2013, ApJ, 778, 127

Goldschmidt, P., Kukulka, M. J., Miller, L., & Dunlop, J. S. 1999, ApJ, 511, 612

Grazian, A., Giallongo, E., Boutsia, K., et al. 2018, A&A, 613, A44

Grier, C. J., Peterson, B. M., Pogge, R. W., et al. 2012, ApJ, 755, 60

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332

Gurkan, G. e. a. 2019, A&A

Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, ApJ, 489, 37

- Haardt, F. & Madau, P. 2012, ApJ, 746, 125
- Hales, S. E. G., Baldwin, J. E., & Warner, P. J. 1988, MNRAS, 234, 919
- Hamana, T., Yoshida, N., & Suto, Y. 2002, ApJ, 568, 455
- Hamilton, A. J. S. 1993, ApJ, 417, 19
- Härdle, W. 1990, Applied Nonparametric Regression, Econometric Society Monographs (Cambridge University Press)
- Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
- Hatziminaoglou, E., Mathez, G., & Pelló, R. 2000, A&A, 359, 9
- Heald, G. H., Pizzo, R. F., Orrú, E., et al. 2015, A&A, 582, A123
- Helfand, D. J., White, R. L., & Becker, R. H. 2015, ApJ, 801, 26
- Herrera Ruiz, N., Middelberg, E., Norris, R. P., & Maini, A. 2016, A&A, 589, L2
- Hertzprung, E. 1909, Astronomische Nachrichten, 179, 373
- Heywood, I., Jarvis, M. J., & Condon, J. J. 2013, MNRAS, 432, 2625
- Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, ApJ, 696, 891
- Hickox, R. C., Myers, A. D., Brodwin, M., et al. 2011, ApJ, 731, 117
- Hickox, R. C., Wardlow, J. L., Smail, I., et al. 2012, MNRAS, 421, 284
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. 1983, Understanding robust and exploratory data analysis
- Hodapp, K. W., Siegmund, W. A., Kaiser, N., et al. 2004, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 5489, Ground-based Telescopes, ed. J. Oschmann, Jacobus M., 667–678
- Hodge, J. A., Becker, R. H., White, R. L., Richards, G. T., & Zeimann, G. R. 2011, AJ, 142, 3

- Hook, I. M., McMahon, R. G., Shaver, P. A., & Snellen, I. A. G. 2002, *A&A*, 391, 509
- Hook, I. M., Shaver, P. A., & McMahon, R. G. 1998, in *Astronomical Society of the Pacific Conference Series*, Vol. 146, *The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift*, ed. S. D'Odorico, A. Fontana, & E. Giallongo, 17
- Hopkins, A. M., Mobasher, B., Cram, L., & Rowan-Robinson, M. 1998, *MNRAS*, 296, 839
- Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, *ApJS*, 175, 356
- Hoyle, F. 1948, *MNRAS*, 108, 372
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841
- Intema, H. T. 2014, in *Astronomical Society of India Conference Series*, Vol. 13, *Astronomical Society of India Conference Series*
- Intema, H. T., Jagannathan, P., Mooley, K. P., & Frail, D. A. 2017, *A&A*, 598, A78
- Intema, H. T., van der Tol, S., Cotton, W. D., et al. 2009, *A&A*, 501, 1185
- Ivashchenko, G., Zhdanov, V. I., & Tugay, A. V. 2010, *MNRAS*, 409, 1691
- Ivezić, Ž., Brandt, W. N., Fan, X., et al. 2014, in *IAU Symposium*, Vol. 304, *Multi-wavelength AGN Surveys and Studies*, ed. A. M. Mickaelian & D. B. Sanders, 11–17
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Ivezić, Ž., Menou, K., Knapp, G. R., et al. 2002, *AJ*, 124, 2364
- Jackson, J. C. 1972, *MNRAS*, 156, 1P
- Jannuzi, B., Weiner, B., Block, M., et al. 2010, in *Bulletin of the American Astronomical Society*, Vol. 42, *American Astronomical Society Meeting Abstracts #215*, 513
- Jannuzi, B. T. & Dey, A. 1999, in *Astronomical Society of the Pacific Conference Series*, Vol. 191, *Photometric Redshifts and the Detection of High Redshift Galaxies*, ed. R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner, 111

- Jansky, K. G. 1933, *Nature*, 132, 66
- Jarvis, M. J. & McLure, R. J. 2002, *MNRAS*, 336, L38
- Jiang, L., Fan, X., Ivezić, Ž., et al. 2007a, *ApJ*, 656, 680
- Jiang, L., Fan, X., Vestergaard, M., et al. 2007b, *AJ*, 134, 1150
- Jiang, L., McGreer, I. D., Fan, X., et al. 2016, *ApJ*, 833, 222
- Jin, X., Zhang, Y., Zhang, J., et al. 2019, *MNRAS*, 485, 4539
- Kaiser, N. 1987, *MNRAS*, 227, 1
- Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, in *Proc. SPIE, Vol. 4836, Survey and Other Telescope Technologies and Discoveries*, ed. J. A. Tyson & S. Wolff, 154–164
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in *Proc. SPIE, Vol. 7733, Ground-based and Airborne Telescopes III*, 77330E
- Kashikawa, N., Ishizaki, Y., Willott, C. J., et al. 2015, *ApJ*, 798, 28
- Kaspi, S., Netzer, H., Maoz, D., et al. 2004, in *Coevolution of Black Holes and Galaxies*, ed. L. C. Ho, 30
- Kaspi, S., Smith, P. S., Netzer, H., et al. 2000, *ApJ*, 533, 631
- Kassim, N. E., Perley, R. A., Erickson, W. C., & Dwarakanath, K. S. 1993, *AJ*, 106, 2218
- Katgert, P., Oort, M. J. A., & Windhorst, R. A. 1988, *A&A*, 195, 21
- Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, *MNRAS*, 346, 1055
- Kazemi, S., Yatawatta, S., Zaroubi, S., et al. 2011, *MNRAS*, 414, 1656
- Kellermann, K. I., Fomalont, E. B., Weistrop, D., & Wall, J. 1986, *Highlights of Astronomy*, 7, 367
- Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B., & Green, R. 1989, *AJ*, 98, 1195

- Kennefick, J. D., Djorgovski, S. G., & de Carvalho, R. R. 1995, *AJ*, 110, 2553
- Kenter, A., Murray, S. S., Forman, W. R., et al. 2005, *ApJS*, 161, 9
- Kerr, R. P. 1963, *Physical Review Letters*, 11, 237
- Kimball, A. E., Kellermann, K. I., Condon, J. J., Ivezić, Ž., & Perley, R. A. 2011, *ApJL*, 739, L29
- Kimball, A. E., Knapp, G. R., Ivezić, Ž., et al. 2009, *ApJ*, 701, 535
- Kirkpatrick, J. A., Schlegel, D. J., Ross, N. P., et al. 2011, *ApJ*, 743, 125
- Kochanek, C. S., Eisenstein, D. J., Cool, R. J., et al. 2012, *ApJS*, 200, 8
- Kratzer, R. M. & Richards, G. T. 2015, *AJ*, 149, 61
- Krumpe, M., Miyaji, T., & Coil, A. L. 2010, *ApJ*, 713, 558
- Kunert-Bajraszewska, M., Cegłowski, M., Katarzyński, K., & Roskowiński, C. 2015, *A&A*, 579, A109
- Kurk, J., Venemans, B., Röttgering, H., Miley, G., & Pentericci, L. 2004, in *Astrophysics and Space Science Library*, Vol. 301, *Astrophysics and Space Science Library*, ed. M. Plionis, 141
- La Franca, F., Gregorini, L., Cristiani, S., de Ruiter, H., & Owen, F. 1994, *AJ*, 108, 1548
- Lacy, M., Petric, A. O., Sajina, A., et al. 2007, *AJ*, 133, 186
- Lacy, M., Wilson, G., Masci, F., et al. 2005, *ApJS*, 161, 41
- Landy, S. D. & Szalay, A. S. 1993, *ApJ*, 412, 64
- Lang, D. 2014, *AJ*, 147, 108
- Laor, A. 2000, *ApJL*, 543, L111
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *arXiv e-prints* [[arXiv]1110.3193]

- Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, *A&A*, 559, A14
- Leipski, C., Falcke, H., Bennert, N., & Hüttemeister, S. 2006, *A&A*, 455, 161
- Li, Q. & Racine, J. 2011, *Nonparametric Econometrics: Theory and Practice* (Princeton University Press)
- Lin, Y.-T., Shen, Y., Strauss, M. A., Richards, G. T., & Lunnan, R. 2010, *ApJ*, 723, 1119
- Lindsay, S. N., Jarvis, M. J., & McAlpine, K. 2014a, *MNRAS*, 440, 2322
- Lindsay, S. N., Jarvis, M. J., Santos, M. G., et al. 2014b, *MNRAS*, 440, 1527
- Lira, P., Kaspi, S., Netzer, H., et al. 2018, *ApJ*, 865, 56
- Liu, Y., Jiang, D. R., Wang, T. G., & Xie, F. G. 2008, *MNRAS*, 391, 246
- Loh, Y.-S., Rich, R. M., Heinis, S., et al. 2010, *MNRAS*, 407, 55
- Lonsdale, C. J., Cappallo, R. J., Morales, M. F., et al. 2009, *IEEE Proceedings*, 97, 1497
- Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, *PASP*, 115, 897
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, *ArXiv e-prints* [[arXiv]0912.0201]
- Lu, Y., Wang, T., Zhou, H., & Wu, J. 2007, *AJ*, 133, 1615
- Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, *AJ*, 118, 1406
- Lynden-Bell, D. 1969, *Nature*, 223, 690
- MacLeod, C. L., Brooks, K., Ivezić, Ž., et al. 2011, *ApJ*, 728, 26
- MacLeod, C. L., Ivezić, Ž., Kochanek, C. S., et al. 2010, *ApJ*, 721, 1014
- Madau, P., Haardt, F., & Rees, M. J. 1999, *ApJ*, 514, 648

Maddox, N., Hewett, P. C., Péroux, C., Nestor, D. B., & Wisotzki, L. 2012, MNRAS, 424, 2876

Magliocchetti, M., Maddox, S. J., Hawkins, E., et al. 2004, MNRAS, 350, 1485

Magliocchetti, M., Maddox, S. J., Jackson, C. A., et al. 2002, MNRAS, 333, 100

Magliocchetti, M., Maddox, S. J., Lahav, O., & Wall, J. V. 1998, MNRAS, 300, 257

Magnier, E. A., Sweeney, W. E., Chambers, K. C., et al. 2016, arXiv e-prints, arXiv:1612.05244

Mahony, E. K., Morganti, R., Prandoni, I., et al. 2016, MNRAS, 463, 2997

Mandal, S. in prep.

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJL, 619, L1

Masters, D., Capak, P., Salvato, M., et al. 2012, ApJ, 755, 169

Masters, D., Capak, P., Stern, D., et al. 2015, ApJ, 813, 53

Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2016, ApJ, 828, 26

Matthews, T. A. & Sandage, A. R. 1963, ApJ, 138, 30

Mauduit, J. C., Lacy, M., Farrah, D., et al. 2012, PASP, 124, 1135

McGreer, I. D., Becker, R. H., Helfand, D. J., & White, R. L. 2006, ApJ, 652, 157

McGreer, I. D., Fan, X., Jiang, L., & Cai, Z. 2018, AJ, 155, 131

McGreer, I. D., Helfand, D. J., & White, R. L. 2009, AJ, 138, 1925

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, ApJ, 768, 105

McLure, R. J. & Dunlop, J. S. 2001, MNRAS, 321, 515

McLure, R. J. & Dunlop, J. S. 2004, MNRAS, 352, 1390

McLure, R. J. & Jarvis, M. J. 2004, MNRAS, 353, L45

- Meier, D. L. 1976a, *ApJL*, 203, L103
- Meier, D. L. 1976b, *ApJ*, 207, 343
- Meneux, B., Guzzo, L., de la Torre, S., et al. 2009, *A&A*, 505, 463
- Messias, H., Afonso, J., Salvato, M., Mobasher, B., & Hopkins, A. M. 2012, *ApJ*, 754, 120
- Metcalf, R. B. & Magliocchetti, M. 2006, *MNRAS*, 365, 101
- Miley, G. & De Breuck, C. 2008, *A&A Reviews*, 15, 67
- Miley, G. K., Overzier, R. A., Zirm, A. W., et al. 2006, *ApJL*, 650, L29
- Miller, L., Peacock, J. A., & Mead, A. R. G. 1990, *MNRAS*, 244, 207
- Miller, N. A., Bonzini, M., Fomalont, E. B., et al. 2013, *ApJS*, 205, 13
- Miyaji, T., Hasinger, G., Salvato, M., et al. 2015, *ApJ*, 804, 104
- Mohan, N. & Rafferty, D. 2015, *PyBDSM: Python Blob Detection and Source Measurement*, *Astrophysics Source Code Library*
- Møller, P. & Jakobsen, P. 1990, *A&A*, 228, 299
- Morabito, L. K., Deller, A. T., Röttgering, H., et al. 2016, *MNRAS*, 461, 2676
- Morabito, L. K., Matthews, J. H., Best, P. N., et al. 2018, *ArXiv e-prints* [[arXiv]1811.07931]
- Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, *Nature*, 474, 616
- Mountrichas, G. & Georgakakis, A. 2012, *MNRAS*, 420, 514
- Mountrichas, G., Georgakakis, A., Finoguenov, A., et al. 2013, *MNRAS*, 430, 661
- Mullin, L. M., Riley, J. M., & Hardcastle, M. J. 2008, *MNRAS*, 390, 595
- Myers, A. D., Brunner, R. J., Nichol, R. C., et al. 2007, *ApJ*, 658, 85
- Myers, A. D., Brunner, R. J., Richards, G. T., et al. 2006, *ApJ*, 638, 622

Myers, A. D., Palanque-Delabrouille, N., Prakash, A., et al. 2015, *ApJS*, 221, 27

Nadaraya, E. 1964, *Theory of Probability & Its Applications*, 9, 141

Nakoneczny, S., Bilicki, M., Solarz, A., et al. 2019, *A&A*, 624, A13

Niida, M., Nagao, T., Ikeda, H., et al. 2016, *ApJ*, 832, 208

Noordam, J. E. 2004, in *Proc. SPIE*, Vol. 5489, *Ground-based Telescopes*, ed. J. M. Oschmann, Jr., 817–825

Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, 396, 19

Norris, R. P. 2017, *Nature Astronomy*, 1, 671

Norris, R. P., Hopkins, A. M., Afonso, J., et al. 2011, *PASA*, 28, 215

Nusser, A. & Tiwari, P. 2015, *ApJ*, 812, 85

O’Dea, C. 1998, *Publications of the Astronomical Society of the Pacific*, 110, 493

O’Dea, C. P. 1998, *PASP*, 110, 493

O’Dea, C. P. & Owen, F. N. 1986, *ApJ*, 301, 841

Offringa, A. R., de Bruyn, A. G., Biehl, M., et al. 2010, *MNRAS*, 405, 155

Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, *MNRAS*, 444, 606

Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M. 2012, *A&A*, 539, A95

Oke, J. B. & Gunn, J. E. 1983, *ApJ*, 266, 713

Orienti, M., Dallacasa, D., & Stanghellini, C. 2007, *A&A*, 461, 923

Oshlack, A. Y. K. N., Webster, R. L., & Whiting, M. T. 2002, *ApJ*, 576, 81

Overzier, R. A., Röttgering, H. J. A., Rengelink, R. B., & Wilman, R. J. 2003, *A&A*, 405, 53

Padovani, P. 1993, *MNRAS*, 263, 461

Padovani, P. 2011, MNRAS, 411, 1547

Padovani, P., Bonzini, M., Kellermann, K. I., et al. 2015, MNRAS, 452, 1263

Padovani, P., Mainieri, V., Tozzi, P., et al. 2009, ApJ, 694, 235

Padovani, P., Miller, N., Kellermann, K. I., et al. 2011, ApJ, 740, 20

Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2016, A&A, 587, A41

Palanque-Delabrouille, N., Yeche, C., Myers, A. D., et al. 2011, A&A, 530, A122

Papovich, C., Cool, R., Eisenstein, D., et al. 2006, AJ, 132, 231

Papovich, C., Shipley, H. V., Mehtens, N., et al. 2016, ApJS, 224, 28

Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79

Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26

Pasquet-Itam, J. & Pasquet, J. 2018, A&A, 611, A97

Pastorini, G., Marconi, A., Capetti, A., et al. 2007, A&A, 469, 405

Pawlik, A. H., Schaye, J., & van Scherpenzeel, E. 2009, MNRAS, 394, 1812

Peacock, J. A., Miller, L., & Longair, M. S. 1986, MNRAS, 218, 265

Peacock, J. A. & Nicholson, D. 1991, MNRAS, 253, 307

Peacock, J. A. & Wall, J. V. 1982, MNRAS, 198, 843

Pearson, T. J. & Readhead, A. C. S. 1984, ARAA, 22, 97

Peebles, P. J. E. 1980, The large-scale structure of the universe

Pei, Y. C. 1992, ApJ, 395, 130

Pei, Y. C. 1995, ApJ, 438, 623

Peng, N., Zhang, Y., Zhao, Y., & Wu, X.-b. 2012, MNRAS, 425, 2599

- Peters, C. M., Richards, G. T., Myers, A. D., et al. 2015, *ApJ*, 811, 95
- Peterson, B. M. 1988, *PASP*, 100, 18
- Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, *ApJ*, 613, 682
- Porciani, C., Magliocchetti, M., & Norberg, P. 2004, *MNRAS*, 355, 1010
- Porciani, C. & Norberg, P. 2006, *MNRAS*, 371, 1824
- Prandoni, I., de Ruiter, H. R., Ricci, R., et al. 2010, *A&A*, 510, A42
- Prandoni, I., Gregorini, L., Parma, P., et al. 2000, *A&A, Supplement*, 146, 41
- Prandoni, I., Gregorini, L., Parma, P., et al. 2001, *A&A*, 365, 392
- Pu, X. 2013, *Ap & SS*, 345, 355
- Qiu, P. 2013, *Introduction to Statistical Process Control*, Chapman & Hall/CRC Texts in Statistical Science (CRC Press)
- Ramos Almeida, C., Bessiere, P. S., Tadhunter, C. N., et al. 2013, *MNRAS*, 436, 997
- Reber, G. 1940a, *Proceedings of the IRE*, 28, 68
- Reber, G. 1940b, *ApJ*, 91, 621
- Reber, G. 1944, *ApJ*, 100, 279
- Rengelink, R. B., Tang, Y., de Bruyn, A. G., et al. 1997, *A&A, Supplement*, 124
- Retana-Montenegro, E. & Röttgering, H. 2018, *Frontiers in Astronomy and Space Sciences*, 5, 5
- Retana-Montenegro, E. & Röttgering, H. J. A. 2017, *A&A*, 600, A97
- Retana-Montenegro, E., Röttgering, H. J. A., Shimwell, T. W., et al. 2018a, *A&A*, 620, A74
- Retana-Montenegro, E., Röttgering, H. J. A., Shimwell, T. W., et al. 2018b, *A&A*, 620, A74

- Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, *AJ*, 123, 2945
- Richards, G. T., Kruczek, N. E., Gallagher, S. C., et al. 2011, *AJ*, 141, 167
- Richards, G. T., Myers, A. D., Gray, A. G., et al. 2009, *ApJS*, 180, 67
- Richards, G. T., Myers, A. D., Peters, C. M., et al. 2015, *ApJS*, 219, 39
- Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, *AJ*, 131, 2766
- Richards, G. T., Weinstein, M. A., Schneider, D. P., et al. 2001, *AJ*, 122, 1151
- Rochais, T. B., DiPompeo, M. A., Myers, A. D., et al. 2014, *MNRAS*, 444, 2498
- Ross, N. P., McGreer, I. D., White, M., et al. 2013, *ApJ*, 773, 14
- Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, *ApJS*, 199, 3
- Ross, N. P., Shen, Y., Strauss, M. A., et al. 2009, *ApJ*, 697, 1634
- Röttgering, H., Afonso, J., Barthel, P., et al. 2011, *Journal of Astrophysics and Astronomy*, 32, 557
- Russell, H. N. 1914, *Popular Astronomy*, 22, 275
- Sabater, J. in prep.
- Salpeter, E. E. 1964, *ApJ*, 140, 796
- Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
- Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, *ApJ*, 742, 61
- Sandage, A. 1965, *ApJ*, 141, 1560
- Sandage, A., Véron, P., & Wyndham, J. D. 1965, *ApJ*, 142, 1307
- Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, *ApJS*, 172, 86
- Scaife, A. M. M. & Heald, G. H. 2012, *MNRAS*, 423, L30
- Schilizzi, R. T. 2005, in *EAS Publications Series*, ed. L. I. Gurvits, S. Frey, & S. Rawlings, Vol. 15, 445–463

Schindler, J.-T., Fan, X., McGreer, I. D., et al. 2017, ApJ, 851, 13

Schinnerer, E., Sargent, M. T., Bondi, M., et al. 2010, ApJS, 188, 384

Schlafly, E. F. & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schmidt, M. 1963, Nature, 197, 1040

Schmidt, M. 1968, ApJ, 151, 393

Schmidt, M. 1969, ARAA, 7, 527

Schmidt, M., Schneider, D. P., & Gunn, J. E. 1995, AJ, 110, 68

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360

Schulze, A., Done, C., Lu, Y., Zhang, F., & Inoue, Y. 2017, ApJ, 849, 4

Schwab, F. R. 1984, AJ, 89, 1076

Schwarzschild, K. 1916, Abh. Konigl. Preuss. Akad. Wissenschaften Jahre 1906,92, Berlin,1907, 1916

Scranton, R., Johnston, D., Dodelson, S., et al. 2002, ApJ, 579, 48

Shankar, F., Crocce, M., Miralda-Escudé, J., Fosalba, P., & Weinberg, D. H. 2010a, ApJ, 718, 231

Shankar, F., Weinberg, D. H., & Shen, Y. 2010b, MNRAS, 406, 1959

Shanks, T., Croom, S. M., Fine, S., Ross, N. P., & Sawangwit, U. 2011, MNRAS, 416, 650

Shaver, P. A., Wall, J. V., Kellermann, K. I., Jackson, C. A., & Hawkins, M. R. S. 1996, Nature, 384, 439

Shen, Y. 2009, ApJ, 704, 89

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169

Shen, Y. & Liu, X. 2012, ApJ, 753, 125

Shen, Y., McBride, C. K., White, M., et al. 2013, ApJ, 778, 98

Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45

Shen, Y., Strauss, M. A., Ross, N. P., et al. 2009, ApJ, 697, 1656

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017a, A&A, 598, A104

Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017b, A&A, 598, A104

Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2019, A&A, 622, A1

Shimwell, T. W., Tasse, C., Hardcastle, M. J., et al. 2018, arXiv e-prints, arXiv:1811.07926

Siana, B., Polletta, M. d. C., Smith, H. E., et al. 2008, ApJ, 675, 49

Sikora, M. 2009, Astronomische Nachrichten, 330, 291

Sikora, M., Stawarz, Ł., & Lasota, J.-P. 2007, ApJ, 658, 815

Silverman, J. D., Green, P. J., Barkhouse, W. A., et al. 2005, ApJ, 624, 630

Singal, J., Petrosian, V., Lawrence, A., & Stawarz, L. 2011, ApJ, 743, 104

Singal, J., Petrosian, V., Stawarz, Ł., & Lawrence, A. 2013, ApJ, 764, 43

Siopis, C., Gebhardt, K., Lauer, T. R., et al. 2009, ApJ, 693, 946

Skibba, R. A., Smith, M. S. M., Coil, A. L., et al. 2014, ApJ, 784, 128

Smirnov, O. M. 2011, A&A, 527, A107

Smirnov, O. M. & Tasse, C. 2015, MNRAS, 449, 2668

Smith, H. J. & Hoffleit, D. 1963, AJ, 68, 292

- Smith, J. D., Thompson, D., & Djorgovski, S. 1993, in *Astronomical Society of the Pacific Conference Series*, Vol. 43, *Sky Surveys. Protostars to Protogalaxies*, ed. B. T. Soifer, 189
- Smolčić, V., Delvecchio, I., Zamorani, G., et al. 2017a, *A&A*, 602, A2
- Smolčić, V., Novak, M., Bondi, M., et al. 2017b, *A&A*, 602, A1
- Smolčić, V., Schinnerer, E., Scodreggio, M., et al. 2008, *ApJS*, 177, 14
- Snellen, I. A. G., Schilizzi, R. T., Miley, G. K., et al. 2000, *MNRAS*, 319, 445
- Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, *MNRAS*, 391, 481
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints [[arXiv]1503.03757]
- Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, ArXiv e-prints [[arXiv]1305.5422]
- Spergel, D. N., Bean, R., Doré, O., et al. 2007, *ApJS*, 170, 377
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Starikova, S., Cool, R., Eisenstein, D., et al. 2011, *ApJ*, 741, 15
- Stern, D., Djorgovski, S. G., Perley, R. A., de Carvalho, R. R., & Wall, J. V. 2000, *AJ*, 119, 1526
- Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *ApJ*, 631, 163
- Stocke, J. T., Morris, S. L., Weymann, R. J., & Foltz, C. B. 1992, *ApJ*, 396, 487
- Sulentic, J. & Marziani, P. 2015, *Frontiers in Astronomy and Space Sciences*, 2, 6
- Sulentic, J. W., Bachev, R., Marziani, P., Negrete, C. A., & Dultzin, D. 2007, *ApJ*, 666, 757
- Sulentic, J. W., Marziani, P., & Dultzin-Hacyan, D. 2000a, *Annual Review of Astronomy and Astrophysics*, 38, 521
- Sulentic, J. W., Zamfir, S., Marziani, P., et al. 2003, *ApJL*, 597, L17

Sulentic, J. W., Zwitter, T., Marziani, P., & Dultzin-Hacyan, D. 2000b, *ApJL*, 536, L5

Sullivan, M., Hopkins, A. M., Afonso, J., et al. 2004, *ApJS*, 155, 1

Sutherland, W. & Saunders, W. 1992, *MNRAS*, 259, 413

Tasse, C. 2014, ArXiv e-prints [[arXiv]1410.8706]

Tasse, C. in prep.

Tasse, C., Hugo, B., Mirmont, M., et al. 2017, ArXiv e-prints [[arXiv]1712.02078]

Tasse, C., Hugo, B., Mirmont, M., et al. 2018, *A&A*, 611, A87

Taylor, G. B. 2007, *Highlights of Astronomy*, 14, 388

Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, *ApJ*, 565, 773

Timlin, J. D., Ross, N. P., Richards, G. T., et al. 2016, *ApJS*, 225, 1

Timlin, J. D., Ross, N. P., Richards, G. T., et al. 2018, *ApJ*, 859, 20

Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, *PASA*, 30, e007

Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, *ApJ*, 750, 99

Trump, J. R., Impey, C. D., Elvis, M., et al. 2009, *ApJ*, 696, 1195

Tuccillo, D., González-Serrano, J. I., & Benn, C. R. 2015, *MNRAS*, 449, 2818

Tyson, J. A. 2002, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4836, *Survey and Other Telescope Technologies and Discoveries*, ed. J. A. Tyson & S. Wolff, 10–20

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, *ApJ*, 598, 886

Ulvestad, J. S., Antonucci, R. R. J., & Barvainis, R. 2005, *ApJ*, 621, 123

Urry, C. M. & Padovani, P. 1995, *PASP*, 107, 803

van Breugel, W., Miley, G., & Heckman, T. 1984, *AJ*, 89, 5

- van Breugel, W., Miley, G., Heckman, T., Butcher, H., & Bridle, A. 1985, *ApJ*, 290, 496
- van den Bosch, F. C. 2002, *MNRAS*, 331, 98
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, 556, A2
- van Velzen, S. & Falcke, H. 2013, *A&A*, 557, L7
- van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016, *ApJS*, 223, 2
- van Weeren, R. J., Williams, W. L., Tasse, C., et al. 2014, *ApJ*, 793, 82
- Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, *AJ*, 122, 549
- Varenius, E., Conway, J. E., Martí-Vidal, I., et al. 2015, *A&A*, 574, A114
- Vasconcellos, E. C., de Carvalho, R. R., Gal, R. R., et al. 2011, *AJ*, 141, 189
- Vernstrom, T., Scott, D., Wall, J. V., et al. 2016, *MNRAS*, 462, 2934
- Vigotti, M., Carballo, R., Benn, C. R., et al. 2003, *ApJ*, 591, 43
- Visnovsky, K. L., Impey, C. D., Foltz, C. B., et al. 1992, *ApJ*, 391, 560
- Volonteri, M. & Rees, M. J. 2005, *ApJ*, 633, 624
- Wake, D. A., Croom, S. M., Sadler, E. M., & Johnston, H. M. 2008, *MNRAS*, 391, 1674
- Walsh, J. L., Barth, A. J., Ho, L. C., & Sarzi, M. 2013, *ApJ*, 770, 86
- Wang, D., Zhang, Y. X., Liu, C., & Zhao, Y. H. 2007, *MNRAS*, 382, 1601
- Warren, S. J., Hewett, P. C., & Osmer, P. S. 1994, *ApJ*, 421, 412
- Watson, G. S. 1964, *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 26, 359
- Wayth, R. B., Lenc, E., Bell, M. E., et al. 2015, *PASA*, 32, e025
- Weinstein, M. A., Richards, G. T., Schneider, D. P., et al. 2004, *ApJS*, 155, 243

- Welling, C. A., Miller, B. P., Brandt, W. N., Capellupo, D. M., & Gibson, R. R. 2014, MNRAS, 440, 2474
- White, G. J., Hatsukade, B., Pearson, C., et al. 2012, MNRAS, 427, 1830
- White, R. L., Helfand, D. J., Becker, R. H., Glikman, E., & de Vries, W. 2007, ApJ, 654, 99
- Williams, W. L., Intema, H. T., & Röttgering, H. J. A. 2013, A&A, 549, A55
- Williams, W. L., van Weeren, R. J., Röttgering, H. J. A., et al. 2016, MNRAS, 460, 2385
- Willott, C. J., Delorme, P., Omont, A., et al. 2007, AJ, 134, 2435
- Willott, C. J., Delorme, P., Reylé, C., et al. 2010, AJ, 139, 906
- Willott, C. J., McLure, R. J., & Jarvis, M. J. 2003, ApJL, 587, L15
- Wilman, R. J., Miller, L., Jarvis, M. J., et al. 2008, MNRAS, 388, 1335
- Wilson, A. S. & Colbert, E. J. M. 1995, ApJ, 438, 62
- Windhorst, R., Mathis, D., & Neuschaefer, L. 1990, in *Astronomical Society of the Pacific Conference Series, Vol. 10, Evolution of the Universe of Galaxies*, ed. R. G. Kron, 389–403
- Windhorst, R. A., Fomalont, E. B., Partridge, R. B., & Lowenthal, J. D. 1993, ApJ, 405, 498
- Windhorst, R. A., Miley, G. K., Owen, F. N., Kron, R. G., & Koo, D. C. 1985, ApJ, 289, 494
- Wolf, C., Wisotzki, L., Borch, A., et al. 2003, A&A, 408, 499
- Woo, J.-H. & Urry, C. M. 2002, ApJL, 581, L5
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
- Wu, H. & Zhang, J. 2006 (Wiley Press)

Wylezalek, D., Galametz, A., Stern, D., et al. 2013, ApJ, 769, 79

Yang, J., Wu, X.-B., Liu, D., et al. 2018, AJ, 155, 110

Yang, Q., Wu, X.-B., Fan, X., et al. 2017, AJ, 154, 269

Yao, S., Wu, X.-B., Ai, Y. L., et al. 2019, ApJS, 240, 6

Yatawatta, S., de Bruyn, A. G., Brentjens, M. A., et al. 2013, A&A, 550, A136

Yèche, C., Petitjean, P., Rich, J., et al. 2010, A&A, 523, A14

Yi, W.-M., Wang, F., Wu, X.-B., et al. 2014, ArXiv e-prints [[arXiv]1410.2689]

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zakamska, N. L. & Greene, J. E. 2014, MNRAS, 442, 784

Zakamska, N. L., Lampayan, K., Petric, A., et al. 2016, MNRAS, 455, 4191

Zamfir, S., Sulentic, J. W., & Marziani, P. 2008, MNRAS, 387, 856

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1

Zeimann, G. R., White, R. L., Becker, R. H., et al. 2011, ApJ, 736, 57

Samenvatting

Rond de tijd van de ontdekking van de eerste quasar, 3C273, stond het algemeen geaccepteerde paradigma van een statisch heelal onder druk vanwege de theorie dat zware zwarte gaten aanwezig zijn in het centrum van de meerderheid van sterrenstelsels, en vanwege de waarneming dat sterrenstelsels zich van ons af bewegen. Sindsdien heeft de continue verbetering van sterrenkundige waarnemingen ons in staat gesteld om de belangrijke rol van zwarte gaten in het ontstaan en de evolutie van sterrenstelsels beter te begrijpen.

Nu, meer dan 50 jaar na de ontdekking van 3C273, zijn er meer dan een half miljoen quasars gevonden die alsmaar in massa toenemen, verspreid over een enorme tijdspanne van het heden (het lokale heelal) tot aan een tijd waar het (verre) heelal minder dan zeventien miljoen jaar oud was. De grote lichtkracht van quasars stelt ons in staat deze objecten in groot detail te bestuderen, met name de karakteristieken van de zwarte gaten die de drijvende kracht zijn achter quasars.

Een kernthema van dit proefschrift is het bestuderen van quasars vanuit een radio-sterrenkundig perspectief. Radiostraling van quasars geeft onder andere informatie over hoe quasars ontstaan en evolueren, over het verband tussen de quasar en z'n directe omgeving, en over eigenschappen van het sterrenstelsel waarin de quasar zich bevindt. Bijvoorbeeld, de meest krachtige radio-quasars (radio-loud quasars; RLQs) worden waarschijnlijk aangedreven door de zwaarste zwarte gaten en bevinden zich in de meest massieve structuren in het heelal. Daarintegen worden zwakke radio-quasars (radio-quiet quasars; RQQs) waarschijnlijk aangedreven door minder zware zwarte gaten.

Quasars

Omdat quasars compacte bronnen zijn met een hoge lichtkracht, werd de eerst ontdekte quasar verkeerd geclassificeerd als zijnde een ster in ons eigen sterrenstelsel

(de Melkweg) met een ongewoon spectrum. Deze vergissing leeft nog steeds voort in de naam quasar, een afkorting van quasi-stellar radio source. De kosmische aard van quasars werd als eerste herkend door de Nederlands-Amerikaanse sterrenkundige Maarten Schmidt. Schmidt identificeerde de brede emissielijnen in het optische quasar-spectrum als emissielijnen van waterstof die meer dan 15% verschoven waren naar de rode kant van het spectrum. Quasars hebben verscheidene karakteristieke eigenschappen die ze onderscheidt van andere actieve kernen van sterrenstelsels (active galactic nuclei; AGN). De aanwezigheid van sterke, brede emissielijnen is een kenmerkende eigenschap van quasars, en wordt toegeschreven aan de aanwezigheid van wolken van hoog geëxciteerd gas van hoge dichtheid en zeer hoge temperatuur, die met hoge snelheid nabij het centrale zwarte gat bewegen. Onder de aanname dat de beweging van de gaswolken wordt beschreven door klassieke Newtoniaanse mechanica kan de massa van het zwarte gat worden afgeleid. Hieruit volgt een typische massa van zwarte gaten in quasars van meer dan een paar miljoen zonsmassas.

Variatie in lichtkracht is een andere belangrijke eigenschap van quasars die al vroeg werd bestudeerd. Zo varieert de lichtkracht van de vroegst-bekende quasar 3C273 op tijdschalen van dagen tot jaren. Deze variaties impliceren dat de emissie afkomstig is van een compacte bron, niet groter dan een paar lichtjaar. De kleur van quasars is significant anders dan die van de meeste sterren en sterrenstelsels vanwege de brede emissielijnen (vooral de L_{y_α} emissielijn). Deze onderscheidende eigenschap heeft in de afgelopen twee decennia geleid tot een aantal grote zoektochten (surveys) naar quasars op grote afstand (hoge roodverschuiving). Hierdoor zijn er nu meer dan een half miljoen quasars bekend tot op een roodverschuiving van 7.5. Radio-waarnemingen zijn een andere belangrijke manier om quasars beter te begrijpen. De meerderheid van quasars zijn zwakke radiobronnen (RQQs) en blijven meestal ongedetecteerd in grote radio surveys, maar ongeveer 10–15% van de quasars is honderden tot duizenden malen krachtiger (RLQs). Radiostraling van bronnen in deze laatste categorie is afkomstig van gecollimeerde jets nabij het zwarte gat. Waarom quasars radio-loud worden is nog steeds een openstaande vraag.

Radio-sterrenkunde op zeer lage frequenties

Er is nog steeds veel onduidelijk over oorsprong van radiostraling in quasars. In RLQs wordt radiostraling gegenereerd in de vorm van synchrotron-straling. Voor RQQs bestaat het vermoeden dat ook stervormingsprocessen in het sterrenstelsel waar de quasar gevestigd is verantwoordelijk zijn voor een deel van de radiostraling. Het radiospectrum kan goed worden beschreven met een simpele functie $S_\nu \propto \nu^{-\alpha}$, waarbij S_ν de waargenomen radio-lichtkracht (flux density) is van de bron, ν is de waarneemfrequentie, en α de zogenaamde spectraalindex. Voor quasars heeft α een positieve waarde, waardoor de radio-lichtkracht toeneemt bij lagere waarneemfrequenties.

De Low-Frequency Array (LOFAR) is een nieuwe radio-telescoop, gebaseerd op revolutionaire phased-array technologie, die waarneemt op zeer lage-frequenties (10-240 MHz). Een phased-array radio-telescoop bestaat uit stations van kleine antennes (dipolen) die niet bewegen. Voor ieder station worden de ontvangen signalen van de antennes digitaal gecombineerd tot een virtuele bestuurbare schotel-antenne. Dit elimineert het gebruik van dure, mechanisch bestuurbare schotel-antennes zoals bij de Amerikaanse Very Large Array (VLA) en de Indiase Giant Metrewave Radio Telescope (GMRT). LOFAR is opgebouwd twee soorten antennes: de Low Band Antenna (LBA), werkend tussen 10 en 80 MHz, en de High Band Antenna (HBA), werkend tussen 120 en 240 MHz. LOFAR bestaat momenteel uit 38 Nederlandse stations, en 13 internationale stations in Duitsland, Frankrijk, Verenigd Koninkrijk, Ierland, Zweden en Polen. Het Nederlandse deel van LOFAR alleen produceert al waarnemingen van ongekeerde resolutie en gevoeligheid op lage radiofrequenties. En door het gebruik van de internationale stations is het mogelijk om de beeld-resolutie met nog een factor tien te verhogen. De volledig digitale besturing van LOFAR maakt het mogelijk om tegelijkertijd in verschillende richtingen en op verschillende frequenties waar te nemen, waarbij alleen de totale bandbreedte van de telescoop de beperkende factor is. Deze zogenaamde multi-beam capaciteit geeft een groot instantaan blikveld, die samen met de hoge gevoeligheid en resolutie van LOFAR een enorm krachtige survey telescoop maken.

Het verleggen van de grenzen door diepere waarnemingen en hogere resoluties op zeer lage-frequenties leidt ook tot nieuwe uitdagingen. Één van de grootste uitdagingen

is verstoring van radiosignalen door de Aardse ionosfeer, waardoor radiobronnen aan de hemel zich continue lijken te verplaatsen. Dit kan een grote (negatieve) invloed hebben op de kwaliteit van de radio-afbeeldingen. Het onderdrukken van ionoferische effecten vereist de ontwikkeling en toepassing van nieuwe kalibratie-algoritmes.

Dit proefschrift

Quasars representeren de actieve fase van zware zwarte gaten, en bevinden zich onder de meest heldere, krachtigste, en meest energetische objecten in het heelal. Het doel van dit proefschrift is om radio-waarnemingen op lage en hoge frequenties te gebruiken om de volgende vragen te beantwoorden:

- Zijn radio-sterke en radio-zwakke quasars verschillende objecten?
- Zijn diepe, laag-frequente radio-waarnemingen effectief te gebruiken om quasars op hoge roodverschuiving te vinden?
- Hoe evolueert de populatie van zwakke radio-quasars als een functie van roodverschuiving?
- Is er een samenhang tussen de directe omgeving van quasars en de oorsprong van hun radiostraling?

In dit proefschrift wordt gebruik gemaakt van radio-afbeeldingen op lage en hoge frequenties, van bestaande catalogi van optische spectra van quasars, en van ondersteunende survey data uit optische en infrarood-waarnemingen. In het tweede hoofdstuk worden bestaande radio en spectroscopische catalogi van quasars gebruikt om gemeenschappelijke eigenschappen van RQQs en RLQs te bestuderen. In het derde hoofdstuk wordt beschreven hoe een richtingsafhankelijk kalibratie-algoritme is gebruikt om een diepe LOFAR afbeelding op 150 MHz te maken van het Boötes veld, (een onderdeel van de NOAO Deep Wide-field Survey; NDWFS) en hoe daaruit een catalogus van radiobronnen is geconstrueerd. De daarop volgende hoofdstukken combineren deze catalogus van radiobronnen met data uit de optische en infrarood surveys om quasars op

hoge roodverschuiving te selecteren, en om de evolutie van deze geselecteerde quasars te bestuderen als een functie van roodverschuiving.

Hoofdstuk 2 begint met het samenvoegen van twee surveys: de Sloan Digital Sky Survey (SDSS) en de Faint Images of the Radio Sky at Twenty centimeters (FIRST) survey. Dit resulteert in een dataset van 45441 radio-quiet quasars en 3493 radio-loud quasars met roodverschuivingen van 0.3 tot maximaal 2.3. Hieruit leiden we af dat RLQs zich in halos van donkere materie bevinden die tenminste een orde van grootte zwaarder zijn dan halos met RQQs. Verder vinden we dat de mate waarin quasars zich groeperen afhangt van de massa van de bijbehorende zwarte gaten, waarbij quasars met zwaardere zwarte gaten sterker groeperen dan quasars met minder zware zwarte gaten. Deze resultaten suggereren dat er inderdaad een intrinsiek verschil is tussen RLQs en RQQs.

Hoofdstuk 3 presenteert de resultaten van een diepe radio survey van het NDWFS-Boötes veld, waargenomen met LOFAR tussen 120 en 168 MHz. Hiervoor is in totaal is 55 uur aan LOFAR-waarnemingen gekalibreerd met het richtings-afhankelijk algoritme *facet calibration*. De resulterende afbeelding heeft een resolutie van ongeveer $\sim 5.0''$ en een gevoeligheid van $55 \mu\text{Jy}/\text{beam}$ in het centrum van het veld. De hieruit afgeleide catalogus bevat 10091 radiobronnen (gedetecteerd met een significantie van minstens 5σ) in een gebied van 20 deg^2 . We bevestigen dat de aantallen zwakke radiobronnen (beneden 1 mJy) als een functie van waargenomen radio-lichtkracht (flux density) overeenkomen met de resultaten uit eerdere surveys op hoge en lage radio-frequenties. De reden voor de waargenomen afvlakking in de zwakke radiobron-aantallen is een toenemende bijdrage van stervormende sterrenstelsels en zwakke AGN. Verder bestuderen we de bijdrage aan de onzekerheid van radiobron-aantallen door kosmische variantie, waaruit volgt dat de spreiding door kosmische variantie groter is dan de Poisson-onzekerheid op het aantal bronnen. Dit kan mogelijk de verschillen verklaren tussen eerder gerapporteerde diepe radiobron-aantallen beneden 1 mJy .

Hoofdstuk 4 bouwt voort op de resultaten van **Hoofdstuk 3**. We presenteren een nieuwe methode voor het selecteren van quasars op hoge roodverschuiving, gebruikmakend van onze diepe LOFAR waarnemingen van het NDWFS-Boötes veld. Deze methode

combineert selectie middels optische en infrarode kleuren met $> 5\sigma$ LOFAR detecties op 150 MHz, waarbij wordt voorkomen dat sterren ten onrechte worden geïdentificeerd als quasars. De unieke gevoeligheid van onze LOFAR waarnemingen stelt ons in staat om radiostraling te detecteren van quasars die anders als radio-quiet zouden worden geclassificeerd. Verder onderzoeken we het effect van de spectraalindex op de selectie van quasars door de LOFAR waarnemingen te combineren met radio-afbeeldingen op 1400 MHz van de Westerbork Synthesis Radio Telescope (WSRT). We vinden dat de door ons geselecteerde quasars relatief stijle radiospectra hebben met een gemiddelde spectraalindex van $\alpha \simeq -0.73$ tussen 150 en 1400 MHz.

In **Hoofdstuk 5** gebruiken we machinaal leren technieken om een lijst van quasars samen te stellen uit het NDWFS-Boötes veld. Hiermee bestuderen we de ruimtelijke dichtheid van quasars als functie van intrinsieke lichtkracht (de zogenaamde *luminosity function*) voor radio-geselecteerde quasars (RSQs) met roodverschuivingen tussen 1.4 en 5.0. Onze resultaten suggereren dat heldere RSQs een vergelijkbare evolutionaire trend vertonen met lichtzwakke quasars. Door de ruimtelijke dichtheid van RSQs te vergelijken met lichtzwakke quasars op dezelfde roodverschuiving volgt dat RSQs tot 39% van de *totale* populatie quasars vormen. Binnen de onzekerheden lijkt deze fractie niet afhankelijk te zijn van roodverschuiving. Wij beargumenteren dat de meest lichtzwakke quasars een overgangsfase zouden kunnen zijn tussen radio-quiet en radio-quiet quasars.

English Summary

At the time of the discovery of the first quasar, 3C273, the idea that black holes (BHs) exist at the center of the majority of galaxies, and the fact that galaxies are moving away from us, challenged the paradigm of a non-evolving universe that was accepted by contemporary cosmologists and astronomers. Since then, the continuous advances in astronomical instrumentation have allowed us to understand the profound role of BHs in the formation and evolution of galaxies.

More than 50 years since the discovery of 3C273, we now know of more than half a million quasars that are actively accreting matter, both in the local universe and out to an epoch when the universe was less than seven hundred millions of years old. Their intrinsic high luminosity allows us to study in great detail many of the observational signatures of the BHs and their active engines.

This thesis focuses on studying the properties of quasars from a radio-astronomy perspective. Quasars exhibit radio properties that provide important insights about their triggering mechanisms, feedback processes, environments, host galaxies, and evolution. The brightest radio quasars, often referred to as radio-loud quasars (RLQs), are possibly powered by the most massive BHs and are located in the most massive structures in the universe; however the majority of quasars are undetected at radio-wavelengths (radio-quiet quasars, RQQs), and are thought to be powered by less massive BHs.

Quasars

Initially, the first quasar discovered was mistaken for a star with an unusual spectrum, within our own galaxy due to its star-like morphology and high-luminosity. This confusion is still reflected in the name “quasar”, which is an acronym for quasi stellar radio source. The extragalactic nature of quasars was first recognized by Dutch-American

astronomer Maarten Schmidt, who identified the broad emission lines in the spectra as hydrogen emission lines, which were shifted towards the red end of the spectrum by over 15%, confirming them as extragalactic objects. Quasars have a variety of characteristics that makes them different from other active galactic nuclei. The presence of strong, broad emission lines is a defining characteristic of quasars, and they are attributed to clouds of hot, dense, highly excited gas moving rapidly, close to the central BH. The mass of the central BH can be estimated, assuming that classical Newtonian mechanics describes the motion of gas around it, and is typically found to be several million solar masses. (or you could express this as $> 10^6 M_{\odot}$). Temporal variability in luminosity is another important characteristic of quasars and one of the first to be studied. For example, the first quasar discovered, 3C273, exhibits variability on periods ranging from days to years. This short-term variability suggests that the emission comes from a compact source with a size no larger than a few light years. The broad emission lines of quasars (especially the Ly $_{\alpha}$ line) make the colors of quasars very different from those of galaxies and most stars. This has motivated a number of major surveys dedicated to search for high- z quasars in the last two decades. As a result, there are more than 592000 quasars currently known, up to $z = 7.5$. The radio continuum observations are also important tools that help to improve our understanding of quasars. Although, the majority of quasars have weak radio emission (RQQs) and usually remain radio-undetected in wide-field shallow radio surveys, around 10 – 15% of the quasars are hundreds to thousands of times brighter in radio maps (RLQs). RLQs often present radio morphologies characterised by bright collimated jets emanating from their nucleus. The reasons why a quasar becomes radio-loud are still a subject of debate and investigation.

Very Low-Frequency Radio Astronomy

The origin of the radio emission in quasars is still not completely understood. In the brightest radio quasars, the radio emission is generated by the conversion of potential energy into electromagnetic energy released as synchrotron radiation and manifesting itself as large-scale structures (radio jets and lobes), while in the fainter radio quasars

star-forming processes taking place in their host galaxies are also thought to be responsible for a fraction of the energy output at radio frequencies. The radio spectrum can be described using a power-law of the form $S_\nu \propto \nu^{-\alpha}$, where S_ν is the source flux, α is the spectral index, and ν is the observing frequency. The spectral index is typically $\alpha \approx -0.70$, therefore, the radio brightness of quasars increases at low-frequencies.

The Low Frequency Array (LOFAR) is a new radio telescope operating at very low-frequencies (10-240 MHz) based on revolutionary phased-array technology. A phased-array radio-telescope is composed of stations that contain a certain number of dipoles at fixed orientation. The signals of the dipoles are combined electronically to simulate a digital dish. This eliminates the need for expensive mechanically operated antennas used in radio-telescopes based on a steerable antenna design such as the Very Large Array (VLA) and Giant Metrewave Radio Telescope (GRMT). LOFAR has two different types of dipole antennas: Low Band Antenna (LBA) and High Band Antenna (HBA), optimized to operate at 10-80 MHz and 120-240 MHz, respectively. The majority of LOFAR stations are located in The Netherlands, with at least 13 stations located in Germany, France, United Kingdom, Ireland, Sweden, and Poland. Even LOFAR operating only with the Dutch stations provides unprecedented sensitivity and angular resolution at low-frequencies, while the addition of international stations increases the angular resolution by a factor of 10. Because the LOFAR beams are digital, it is possible to create different combinations of pointing directions and observing frequencies, limited only by the total bandwidth of the radio-telescope. Thus, the multi-beam capabilities, large instantaneous field-of-view, excellent sensitivity, and high-angular resolution make LOFAR a powerful sky-survey machine.

Pushing towards higher sensitivities and angular resolutions at very low-frequencies brings new challenges. The most difficult of these challenges is the Earth's ionosphere, which can have a large impact on the observations by causing the apparent spatial displacement of stationary radio sources in the images. Therefore, new calibration algorithms that can efficiently account for the ionospheric distortions are required.

This Thesis

In this thesis, the main tools used are low-frequency and high-frequency radio imaging, spectroscopic quasar catalogs, and ancillary optical and infrared data. The second chapter of the thesis uses existing radio and spectroscopic catalogs of quasars to study the clustering properties of RQQs and RLQs. The third chapter uses a directional-dependent calibration algorithm known as facet-calibration to obtain a deep LOFAR 150 MHz mosaic of the NOAO Deep Wide-field Survey (NDWFS) Boötes field. The remaining chapters combine the catalog obtained from the deep LOFAR mosaic presented with ancillary optical/infrared data to study the selection of high- z quasars using LOFAR observations, and to investigate the evolution of radio-selected quasars with redshift.

Chapter 2 investigates the clustering properties of 45441 radio-quiet and 3493 radio-loud quasars in the range $0.3 < z < 2.3$. Our results support a scenario where RLQs inhabit dark matter haloes that are at least one order of magnitude more massive compared with those hosting RQQs. This suggests that there is a real dichotomy between RLQs and RQQs, at least considering the masses of their host dark matter haloes.

Chapter 3 presents a deep LOFAR observation of the NDWFS-Boötes field at 120-168 MHz. Differential source counts present a flattening below sub-mJy flux densities, which agrees with previous results from high- and low- frequency surveys. This flattening has been argued to be due to an increasing contribution of star-forming galaxies and faint active galactic nuclei.

In Chapter 4, we propose a method to select high- z quasars using the LOFAR observations presented in **Chapter 3**. The method combines optical/infrared color selection with 5σ LOFAR detections at 150 MHz to guarantee that stellar contamination in our sample is negligible. Furthermore, the depth of our LOFAR observations allow us to detect the radio-emission of quasars that otherwise would be classified as radio-quiet.

In Chapter 5, we use machine learning methods to compile a sample of quasars in the NDWFS-Boötes field to investigate the cosmic evolution of radio-selected quasars (RSQs) between $1.4 \leq z \leq 5.0$. Our results suggest that RSQs show evolutionary trends

that are similar to those exhibited by faint quasars ($M_{1450} < -22.0$). By comparing the spatial density of RSQs with that of faint quasars at similar redshifts, we find that RSQs may compose to up $31 \pm 22\%$ of the *total* (radio-detected and radio-undetected) faint quasar population.

Publications

Refereed first and second author publications:

- **Retana-Montenegro, E.** & Röttgering, H. J. A., *The luminosity function of LOFAR radio-selected quasars at $1.4 \leq z \leq 5.0$ in the NDWFS-Boötes field*, in prep., 2019.
- **Retana-Montenegro, E.** et al., *Deep 150MHz imaging survey of the Boötes field: Unveiling the faint low-frequency sky*, *A&A*, 620, A74, 2018.
- **Retana-Montenegro, E.** & Röttgering, H. J. A., *On the selection of high- z quasars using LOFAR observations*, *Frontiers in Astronomy and Space Sciences*, Vol. 5, id 5, 2018.
- *Andreani, P.*, **Retana-Montenegro, E.**, et al., *Extreme conditions in the molecular gas of lensed star-forming galaxies at $z \sim 3$* , *A&A*, 615, A142, 2018.
- **Retana-Montenegro, E.** & Röttgering, H. J. A., *Probing the radio loud/quiet AGN dichotomy with quasar clustering*, *A&A*, 600, A97, 2017
- **Retana-Montenegro, E.** & Frutos-Alfaro, F. , *Lensing properties of the Einasto profile in terms of the Meijer G function*, *Revista de Matemática Teoría y Aplicaciones*, 21, 2, 2014.
- Frutos-Alfaro, F., **Retana-Montenegro, E.**, Cordero-García, I., Ulloa-Esquivel, O., *Metric of a Slow Rotating Body with Quadrupole Moment from the Erez-Rosen Metric*, *International Journal of Astronomy and Astrophysics*, 3, 431-437, 2013.
- **Retana-Montenegro, E.**, Van Hese, E., Gentile, G., Baes, M. & Frutos-Alfaro, F. , *Analytical properties of Einasto dark matter haloes*, *A&A*, 540, A70, 2012.

- **Retana-Montenegro, E.**, Frutos-Alfaro, F. & Baes, M., *Analytical shear and flexion of Einasto dark matter haloes*, A&A, 546, A32, 2012.

Refereed co-author publications:

- Shimwell, T. W., et al. (including **E. Retana-Montenegro**), *The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release*, A&A, 460, 2385-2412, 2016.
- Williams, W. L., et al. (including **E. Retana-Montenegro**), *LOFAR 150-MHz observations of the Boötes field: catalogue and source counts*, MNRAS, 459, 277-290, 2016
- Shimwell, T. W., et al. (including **E. Retana-Montenegro**), *A plethora of diffuse steep spectrum radio sources in Abell 2034 revealed by LOFAR*, MNRAS, 459, 277-290, 2016
- Mahony, E. K., et al. (including **E. Retana-Montenegro**), *The Lockman Hole project: LOFAR observations and spectral index properties of low-frequency radio sources*, MNRAS, 463, 2997-3020, 2016

Curriculum Vitae

I was born in San José, Costa Rica. Since middle school, I have been interested in learning how nature works, which differed from the then-contemporary interests of many kids. I used to read advance biology, physics and astronomy books from all the places I could borrow them. After I completed my high-school studies, I enrolled at the University of Costa Rica (UCR) to pursue studies in Physics. During my undergraduate studies, I worked **on** several projects within the Astronomy and Astrophysics group at the UCR, including the modeling of gravitational lensing systems under the supervision of Dr. Francisco Frutos-Alfaro. During this time, I also worked as a teaching assistant for general physics laboratories, and Electromagnetism I and II. After my undergraduate studies, I started my Master studies once again working under the supervision of Dr. Frutos-Alfaro, studying the analytical properties of Einasto dark matter haloes.

In 2013, I began my doctoral studies at Leiden University, working under the supervision of Prof. Huub Röttgering. My PhD research focused on using radio observations to study the properties and evolution of quasars. I was involved in commissioning work for the Low Frequency Array (LOFAR), and in applying new calibration algorithms to process LOFAR data. I had short working visits to the European Southern Observatory in Garching, Germany, along with the opportunity to participate in observations runs at the Isaac Newton Telescope in La Palma, Spain and the Giant Metrewave Radio Telescope located in Pune, India. I presented the work of this thesis at various conferences and workshops in the Netherlands (Leiden and Amsterdam), Germany (Munich), Italy (Padova), and South Africa (Cape Town).

Acknowledgements

This would not **have** been possible without the continuous support of my parents, Edwin and Elieth, you always have been **there** to support me in all moments of my life, and encouraged me to achieve my life goals. I am also grateful to my two brothers, Rodrigo and Oscar, for your endless support through these years.

To my colleagues in the LOFAR group: Huib, Reinout, Aayush, Tim, Emma, Wendy, Soumyajit (Jit), Kim, Gabriela (Gaby), Pedro, Alex, Josh, and Frits. Huib thanks for always being open to share your knowledge in radio-astronomy with others. To my officemate Frits, thanks for translating the summary of my thesis into dutch.

I would like to thank my spanish-speaking friends for our conversations about daily life and your scientific results: Heather, Pedro, Cristina, Henry, Gabriela, Paola, Noel, and Santiago.

I would also like to thank the support staff members of Leiden Observatory: Evelijn, Marjan, Alexandra, Caroline, Monica, Liesbeth, and Anita. They were always willing to help me with all my questions, translated letters written in dutch for me, or helped me with any administrative paperwork. I am grateful for the help of the people at the Service Centre International Staff, particularly Ivonne Kluijt, who always addressed all my questions with great care.

Also, I would like to thank my supervisor Huub Röttgering for the opportunity to study at Leiden Observatory. I appreciate his guidance during my PhD studies. I would like to express my gratitude to Xander Tielens for his support in the final part of my PhD. The computer group at Leiden Observatory did an excellent job in maintaining the latest versions of the LOFAR software on our computer clusters. They always responded quickly to all my hardware and software questions.

I am grateful for the guidance of Prof. Francisco Frutos, Prof. Herberth Morales, Prof. Roberto Magaña, and Prof. Rodrigo Carboni during my bachelor and master studies

at the University of Costa Rica. I also would like to express my gratitude to Paola Andreani from the European Southern Observatory for her mentoring.

To the friends that I made during my PhD: Francois, Clement, and Darko. Thanks for all the interesting and enjoyable conversations.

To my friends, Hugo and Wendy, thanks for joining me in the defense of my thesis as my paranymphs.