

Universidad de Cantabria

FACULTAD DE CIENCIAS

DEPARTAMENTO DE FÍSICA MODERNA



TESIS PRESENTADA PARA OPTAR AL GRADO DE DOCTOR EN FÍSICA

**Radio cuásares a alto
desplazamiento al rojo en
cartografiados de gran área**

Autor:

Diego Tuccillo

Director:

Prof. José Ignacio
González Serrano

Programa Oficial de Postgrado en Ciencias, Tecnología y Computación

2015

Universidad de Cantabria

FACULTAD DE CIENCIAS

DEPARTAMENTO DE FÍSICA MODERNA



A THESIS SUBMITTED IN CONFORMITY WITH THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

High redshift radio-loud quasars in large-area surveys

Author:
Diego Tuccillo

Supervisor:
Prof. José Ignacio
González Serrano

Programa Oficial de Postgrado en Ciencias, Tecnología y Computación

2015

Declaración de Autoría

José Ignacio González Serrano, Doctor en Ciencias Físicas y Catedrático de Universidad de la Universidad de Cantabria,

CERTIFICA que la presente memoria

Radio cuásares a alto desplazamiento al rojo en cartografiados de gran área

ha sido realizada por Diego Tuccillo bajo mi dirección en el Instituto de Física de Cantabria, para optar al título de Doctor por la Universidad de Cantabria. Considero que esta memoria contiene aportaciones científicas relevantes, suficientes para constituir la Tesis Doctoral del interesado.

Santander, a 8 de octubre de 2015

Fdo. José Ignacio González Serrano

Conosciamo le malattie del corpo, con qualche difficoltà le malattie dell'anima, quasi per nulla quelle della mente. Eppure, anche le idee della mente si ammalano, talvolta si irrigidiscono, talvolta si assopiscono, talvolta, come le stelle, si spengono. E siccome la nostra vita è regolata dalle idee, di loro dobbiamo avere cura, non tanto per accrescere il nostro sapere, quanto piuttosto per metterlo in ordine¹.

Umberto Galimberti

¹We know many diseases of the body, with some difficulty the diseases of the soul, almost none of the mind. Yet even the mind's ideas get sick, sometimes become rigid, sometimes become dormant, sometimes as the stars, they turn off their light. And since our life is governed by ideas, we must take care of them, not so much to increase our knowledge, but rather to put it in order

Aknowledgements

Getting a PhD is not just about growing your knowledge or acquiring new skills and expertise, it is also a human experience that changes you and hopefully improves you as a person. A journey that would have been way less exciting and fulfilling, with much more downsides, without some people that I would like to thank.

Mi primer agradecimiento va para mi supervisor, Ignacio. Por la gran autonomía que me ha dado siempre, su confianza en mis iniciativas y por todo lo que he aprendido gracias a él. Por los inolvidables recuerdos de las observaciones en Roque de los Muchachos buscando cuásares y rezando para que las nubes se fuesen, por mantener el sentido del humor cuando yo había perdido el mío y, sobre todo, por haber sido un amigo antes que un supervisor.

During these years I had the opportunity to work with and, in some cases, visit some great people I would like to thank. In particular, I am thankful to Chris Benn for the many insightful scientific suggestions during our collaboration and also for the more informal e-mails and comments exchanged.

During the first years of University many "aspiring scientists" have an ideal image of the scientist as someone brilliant, curious, open-minded and always passionate about science no matter the work-charge. Since scientists are people after all, reality naturally runs on a different path. But it is very inspiring to collaborate, or even just meet one of those rare cases in which the reality is close to that ideal juvenile image we had. I want to thank Richard McMahon for having been such an inspiration, for all the science I have learned from you and for all the advice you have given me.

I thank Mike Brotherton, one of the most anticonformist and brilliant persons that I have met, for his great hospitality in Laramie and for his precious support during our collaboration. I am also grateful to all the experienced researchers I have worked with, for their valuable guidance: Karl-Heinz Mack, Mike DiPompeo, Manda Banerji, Alex Kraus.

Un agradecimiento especial a esos colegas que se han convertido también en muy buenos amigos. En particular, gracias a Gabriele Bruni por todas nuestras charlas acerca de la vida y de la ciencia, por los partidos de fútbol y las "rabas

VIII

de Nobel”. Mi agradecimiento a Eduardo González por haberme mostrado que también en Cambridge hay algún buen restaurante y por nuestras charlas, que abarcan desde las astronomical surveys hasta la muerte de Ned Stark.

Mi agradecimiento a todos los compañeros del IFCA que durante estos años me han mostrado su simpatía y amabilidad. Mi agradecimiento a todos los miembros de las "parent populations" de los grupos de AGN y cosmología. Gracias por las barbacoas, las comidas navideñas y las charlas casuales en los pasillos. En particular, a Xavier Barcons y Francisco Carrera, por su disponibilidad para darme opiniones científicas, y por su constante gentileza. Un agradecimiento también a mis compañeros de despacho que en estos años han contribuido a que mi entorno de trabajo fuese más agradable y alegre: Nuria, Serena, Anuar, Nacho y Judit. Gracias también a mi compañera de estancia en Cambridge, Biuse, por la simpatía y las muchas informaciones fiscales que me ha proporcionado. Entre compañeros del IFCA y de Cambridge sitúo mi agradecimiento sincero a Marina Migliaccio por su amistad y nuestras buenas charlas en el IoA a horarios en que los ingleses estaban ya digiriendo la cena.

Ha merecido la pena venir a vivir a Cantabria ya sólo por conocer a algunas personas, que han sido compañeros constantes en esta aventura y han llegado a estar entre mis mejores amigos. Gracias a Carlos Blanco por las charlas entre entrenamientos y chuletas de 1 kg. Nos queda pendiente escribir ese libro y hacer juntos un peregrinaje a Londres. Agradezco a mi amiga Ana Quirce (Annarella) todas nuestras conversaciones surrealistas y haber estado siempre dispuesta a salir de su "batcueva-laboratorio" y ofrecerme su amistad y apoyo. Gracias al ladrón de mi escritorio original, Javier Brochero, por todas las risas a carcajadas incomprensibles para los demás, y que espero que volvamos a vivir.

Mi agradecimiento a Rolando y Eloísa, que me han acogido en España y en su propia casa, y que han sido para mí, durante estos años, mi familia española. Gracias a Ismael y Elisa por su sincera acogida y...por el descubrimiento de los percebes.

Grazie alla mia compagna di vita, Sofía. In te ho trovato la compagna solida e leale, coerente e sensibile, che tutti vorrebbero avere la fortuna di incontrare. Grazie per il tuo grande aiuto morale, materiale e per essere il mio "centro di gravità permanente". Grazie, amore, per aver reso così speciali questi ultimi due anni.

Grazie ai miei amici "de toda la vida", che più che amici per me sono come dei fratelli aggiunti: Marco T., Francesco F.R. e Vincenzo C. Con voi lo spazio ed il tempo sono molto più relativi di quanto una certa teoria prevede. Tra gli altri amici a distanza ringrazio in particolare Lello T. per le consulenze e le chiacchierate telefoniche su tutto e niente; prima o poi riusciremo anche a fare quell'articolo insieme!

Ultimo ma non ultimo va il ringraziamento alla mia bella famiglia. Grazie ai miei genitori per tutto quello che mi hanno insegnato, per l'esempio costante di vita, per il sostegno incondizionato. A mio padre per la tua grande curiosità per capire le cose, e per saper affrontare tutte le cose con un sorriso. A mia madre per la grande forza nell'affrontare le difficoltà, e la sensibilità per capire le persone. Grazie a entrambi per essere le migliori persone che ho conosciuto nella vita. A voi va il ringraziamento costante per tutti i traguardi che ho raggiunto, questo che ho raggiunto ora e per tutti quelli che raggiungerò nel futuro. Grazie a Paola e Dario per comportarsi ed essere per me dei fratelli maggiori anche oramai che "maggiori" sono anche io. Grazie per i consigli, la solidarietà e per l'appoggio anche a distanza di un 2000 Km.

Contents

Contents	XI
Summary	1
Resumen	5
1 Introduction	9
1.1 Quasars: historical background	9
1.2 Active Galactic Nuclei properties	10
1.3 Black Hole paradigm	12
1.4 AGN taxonomy and unified scheme	14
1.5 Radio Loud Quasars	18
1.5.1 Morphology	18
1.5.2 Radio loudness	19
1.5.3 Radio-loud Fraction and Dichotomy	20
1.5.4 Phenomenology and origin of RL QSOs	22
1.6 High redshift quasars in the epoch of Big Surveys	23
2 Selection of RL QSOs at $3.6 \leq z \leq 4.4$	29
2.1 Introduction	29
2.2 Data	31
2.2.1 FIRST	31
2.2.2 SDSS	32
2.2.3 FIRST-SDSS overlap area	35
2.2.4 DR7 Quasar catalogue	35
2.3 Pre - selection criteria	36
2.4 Data mining and Learning Machines	37
2.4.1 Artificial Neural Networks	38
2.4.2 Our Neural Network	41
2.5 QSOs candidates	42
2.5.1 Spectroscopic classification and pre-processing	42

2.5.2	Training and testing of the ANN	43
2.5.3	High-redshift QSO candidates	44
2.6	Conclusions	46
3	Data analysis	47
3.1	Introduction	47
3.2	Spectroscopic follow-up	47
3.3	Check of completeness and efficiency	48
3.3.1	Check of completeness from SDSS DR9	48
3.3.2	Assessment of the sample of high- z candidates in C08	53
3.3.3	Spectroscopic completeness of SDSS for high- z QSOs	55
3.4	Conclusions	55
4	RL QSOs luminosity function at $z \sim 4$	59
4.1	Introduction	59
4.2	Quasar Luminosity Function	60
4.3	Data	62
4.3.1	K -correction	62
4.3.2	Sample	65
4.4	Completeness	70
4.4.1	Optical image quality	70
4.4.2	Radio incompleteness	71
4.4.3	Incompleteness of selection by the neural-network	72
4.4.4	Net completeness	72
4.5	Binned Luminosity function: Method	73
4.6	The QSO Luminosity Function	75
4.7	Discussion	79
4.7.1	The space density of RL QSOs at $3.6 \leq z \leq 4.4$	79
4.7.2	Total space density of QSOs at $3.6 \leq z \leq 4.4$	82
4.7.3	The bright-end slope of the luminosity function for RL QSOs	83
4.8	Conclusions	88
5	High-z BALs QSOs	91
5.1	Introduction	91
5.2	BAL Quasars: definition and theoretical models	92
5.3	Radio studies of BALs QSOs	94
5.4	Selection of RL BAL at high- z	96
5.4.1	BALs and comparison samples	98
5.5	Radio observations and data reduction	99
5.5.1	JVLA	101

5.5.2	Effelsberg-100m telescope	102
5.6	Observational results	102
5.6.1	Morphology	105
5.7	SEDs shape and spectral index	105
5.7.1	Synchrotron peak frequencies	105
5.7.2	Spectral indices	106
5.8	Characterization optical/uv	109
5.8.1	Broadband optical colors	110
5.8.2	RL and RQ BALs	110
5.8.3	Discussion	116
5.9	Fraction	116
5.10	Conclusions	118
6	Selection of Candidates high-z RL Red QSOs	121
6.1	Introduction	121
6.2	Astrometry of the surveys	122
6.2.1	FIRST astrometry	123
6.2.2	SDSS Astrometry	124
6.2.3	UKIDSS Survey and astrometry	126
6.3	Preselection	129
6.3.1	Cross-match FIRST-UKIDSS	129
6.3.2	Cross-match [FIRST-UKIDSS] vs SDSS	132
6.4	Sample undetected in SDSS	132
6.4.1	False negatives	132
6.4.2	False positives	137
6.5	Eye Ball check of the pre-selection	143
6.6	Exploratory Data Analysis	145
6.7	Likelihood ratio	146
6.7.1	Estimation of the functions	153
6.8	Reliability of the UKIDSS morphological classification	156
6.8.1	Efficiency of the morphological classification	157
6.9	Catalogue and Quasar selection	161
6.10	Conclusions	162
7	Selection of RL QSOs at $z > 4.7$	165
7.1	Introduction	165
7.2	Preselection of the sample	166
7.3	Candidates selection	167
7.3.1	Candidate list and preliminary observational results	170
7.3.2	WHT observations	170
7.3.3	GTC observations	171

7.4	Conclusions	171
8	Conclusions and future prospective	175
	References	179

Summary

Around 8 – 13% of the known population of quasars (QSOs) are powerful radio emitter and therefore classified as Radio-Loud (RL), while the remaining part is classified as Radio-Quiet (RQ). The reason of the radio emission is explained in terms of the presence of a radio-jet, responsible for the synchrotron emission and relativistic particles in intense magnetic fields. Nevertheless, the reason why only a minority of the quasars show strong radio-emission and which is the physical connection between these two major classes of objects, still lack a convincing explanation.

Quasars are bright sources but they are relatively rare, especially at high redshift, since their comoving density is a strong function of redshift that peaks at $z \sim 2 - 3$ and declines exponentially at higher redshifts. The radio-loud quasar population at $z > 3.5$ is an even more elusive population, since the fraction of radio-loud quasars on the entire population of quasars seems to decrease strongly with redshift.

The population of radio-loud quasars at high- z is therefore quite small and does not allow exhaustive statistical studies and comparisons between radio-loud and radio-quiet populations in the early universe, where it might be possible to gather clues on the connection between radio and optical activity in QSOs. Accurate estimation of the RL quasar population at high redshift is also fundamental for studies on the impact of QSO feedback on the formation and evolution of galaxies and large structures in the young universe.

The quest to increase the number of known radio-loud quasars at high redshift require the use of large area surveys, advanced techniques of data mining to deal with the sheer volume of data of modern surveys, and the refinement of the candidate-selection techniques to select quasar-candidate of potential interest from broad-band photometric surveys.

This thesis is focused on the selection and study of high-redshift RL QSOs at $z \geq 3.6$. We combined modern data mining techniques and machine-learning algorithms for an efficient and complete exploitation of multi-wavelength data from large-area surveys in the optical (SDSS), in the near- and mid-infrared (UKIDSS, WISE) and in the radio (FIRST). We completed and tested our modern techniques

of analysis and selection through follow-up optical spectroscopy and radio observations of the samples of RL quasars of our interest. We analyzed the results of our observations and drew conclusions both on the techniques used for the candidate-selection and on the physics of the studied objects.

In Chapter 1 we give an overview on quasars, from their discovery to their main properties. We focus our attention on radio-loud quasars, on their theoretical and observational characterization, and on the debate on the origin of the existence of the radio-loud and radio-quiet population.

In Chapter 2 we present the results of our original selection of RL QSOs with $3.6 \leq z \leq 4.4$. We selected the initial sample by a cross-match between FIRST and SDSS DR7 photometric catalogue, then we used a feed-forward neural network to obtain a sample of 15 RL QSOs candidates out of a sample of 2,916 radio-optical sources lacking spectroscopic identification.

In Chapter 3 we present the results of the spectroscopic follow-up of the 15 candidates, observed at the NOT telescope. Once the spectroscopic types and redshifts of the observed objects were identified, we were able to compare the theoretical efficiency and completeness of our neural network selection (based on the training sample) with the actual results on objects without previous spectra. In this way we confirmed the predicted results and we determined that our selection methodology is 97% complete and 60% efficient.

In Chapter 4, we use the data selected in the previous chapters to calculate the optical luminosity function for quasars. The sample of radio-loud quasars with $3.6 \leq z \leq 4.4$ and $M_{1450} < -27$ known previously to our research accounted for only 72 known QSOs, but this small sample has been increased to 87 objects with our neural network selection method. The completeness of our selection strategy and the identification of new sources offer the possibility to carry out the most accurate determination of the optical luminosity function for these objects known at the time. In this chapter we present the data that we later used for our determination of the luminosity function, we discuss also the K-correction, we estimate the completeness of our sample and we illustrate our new methodology for the calculation of the luminosity function.

- The research presented in Chapter 2, 3 and 4 has been published in Monthly Notices of the Royal Astronomical Society with the title: *Neural-network selection of high-redshift radio-loud quasars, and the luminosity function at $z \sim 4$* (Tuccillo et al. 2015).

In the work presented in Chapter 5, we selected a sample of 22 high- z RL Broad Absorption Line (BAL) QSOs within $3.6 \leq z \leq 4.8$, extending the known number of objects of this class. Afterwards we observed the brightest objects in several radio frequencies (at the EVLA radio-array and at the 100-m Effelsberg

telescope), together with a comparison sample of non-BAL RL QSOs matched in redshift and radio-optical luminosity. With these data we tested the two main scenarios that have been proposed to explain the existence of the BAL phenomenon. In particular we tested the *orientation scenario* by calculating the orientation of the radio-jet with respect to the line of sight. The shapes of the radio spectra allowed us also to evaluate statistically the radiative age of the sources (counting the number of compact steep-spectrum and gigahertz peaked-spectrum sources), and thus test the *evolutionary scenario*. We also used our sample of RL high- z BAL QSOs to compare their broad-band optical colors with the remaining population of quasars in the same range of redshift. The motivation of this study was to understand whether the radio characteristics found in BAL quasars at lower redshift are representative of the radio-loud BAL QSO population at higher redshift or if there are indications of evolutionary trends in BAL properties.

- The research presented in Chapter 5 is presented in a manuscript that has been submitted to MNRAS: Tuccillo, D., Bruni, G., Di Pompeo, M., Brotherton, M., González-Serrano, J. I., Kraus, A, "*A multi-wavelength continuum characterization of high-redshift broad absorption line quasars*"

In Chapter 6 we describe how we built a catalogue of sources detected in the radio (FIRST) and in the infrared (UKIDSS) but undetected in the optical (SDSS). The majority of the work consists in data mining techniques developed to "clean" the sample from false or unreliable detections. The final part of the chapter consists in the selection of a sample of very red QSO candidates.

- The research presented in Chapter 5 is presented in an advanced draft that will be soon submitted to MNRAS: Tuccillo, D., McMahon, R., González-Serrano, J. I., Solares-González, E. "*A catalogue of highly reddened radio-sources from FIRST-UKIDSS*"

In Chapter 7 we present our ongoing research on the selection of RL quasar candidates $z > 4.4$ selected using data in the radio (FIRST), in the optical (SDSS DR10) and in the infrared (UKIDSS Large Area Survey DR10, WISE, VHS). We present a list of 22 candidates with $4.7 \leq z \leq 5.1$ that we selected as well as the preliminary results and conclusions obtained from their partial observation.

Finally in Chapter 8 we present the conclusions and the further perspective that the research presented in this thesis opens.

Resumen

Los cuásares (QSOs) son fuentes extragalácticas compactas, muy brillantes, originadas por material de acreción que cae en un agujero negro supermasivo (SMBH) con una masa $> 10^6 M_{\odot}$, y forman parte de la familia de los núcleos activos galácticos (AGN). Los cuásares tienen una distribución de energía espectral amplia y se encuentran entre los objetos más brillantes del cielo en todas las longitudes de onda en que han sido observados. Alrededor del $\sim 8 - 13\%$ de la población conocida de cuásares son potentes emisores en radio y, por tanto, se clasifican como radio-intensos (*radio-loud*, RL), mientras que a la parte restante se le denomina radio-silenciosos (*radio-quiet*, RQ). La emisión en radio se explica por la presencia de un *radio-jet*, responsable de la emisión sincrotrón y que está originado por partículas relativistas en presencia de campos magnéticos intensos. Sin embargo, el motivo por el que sólo una minoría de los cuásares muestra una fuerte emisión en radio y cuál es la conexión física entre estas dos clases de objetos no cuenta aún con una explicación convincente. La densidad comóvil de los cuásares luminosos es una fuerte función del desplazamiento al rojo (z) que alcanza un máximo a $z \sim 2 - 3$ y desciende exponencialmente hacia mayores (y menores) desplazamientos al rojo. La densidad comóvil a $z \sim 6$ es ~ 40 veces menor que a $z \sim 3$, a medida que nos aproximamos a la época de formación de agujeros negros supermasivos en el universo. La población de cuásares radio-intensos a $z > 3.5$ es una población aún más elusiva, no sólo porque se trata de una submuestra de la población de cuásares, sino porque además la fracción de cuásares radio-intensos parece disminuir considerablemente a medida que aumenta el desplazamiento al rojo. Por tanto, la población de cuásares radio-intensos a alto z es bastante reducida y no permite llevar a cabo estudios estadísticos exhaustivos ni comparaciones entre poblaciones radio-intensas y radio-silenciosas en el universo temprano, donde podría ser posible recopilar pruebas sobre la conexión entre actividad radio y óptica en los cuásares. Del mismo modo no es posible realizar una evaluación certera de la fracción de radio cuásares, la cual está íntimamente conectada con el debate acerca de la dicotomía entre AGNs RL y RQ, es decir, si las dos poblaciones pertenecen a dos poblaciones distintas o si forman parte de un secuencia continua única. Además, realizar una estimación

rigurosa de la población de cuásares RL a alto desplazamiento al rojo permitiría comparaciones con modelos de formación de QSOs y de la formación y evolución de las galaxias y de grandes estructuras en el universo temprano.

Las observaciones para aumentar el número de cuásares radio-intensos y reducir los sesgos asociados a su selección requieren el uso de cartografiados de gran área de nueva generación, especialmente en el infrarrojo cercano y lejano, ya que la combinación de campo de visión y sensibilidad permite buscar cuásares débiles a mayor desplazamiento al rojo. Esto requiere del perfeccionamiento de técnicas de minería de datos, para procesar y combinar tal cantidad de datos multi-longitud de onda, y así aprovechar completamente su potencial. Por último, es necesaria la mejora en la determinación de los sesgos de selección, que deben ser corregidos minuciosamente para poder inferir las propiedades estadísticas de la población de cuásares.

Esta tesis doctoral está centrada en la selección y estudio de los RL cuásares a alto desplazamiento al rojo $z \geq 3.6$. Se combinaron las técnicas modernas de minería de datos y algoritmos de aprendizaje automático para explotar los datos multi-longitud de onda procedentes de cartografiados de gran área en el óptico (the Sloan Digital Sky Survey or SDSS), en el cercano y medio infrarrojo (the UKIRT Infrared Deep Sky Survey or UKIDSS; the Wide-Field Infrared Survey Explorer or WISE) y en el radio (the Faint Images of the Radio Sky at Twenty-Centimeters or FIRST). Completamos y evaluamos la eficiencia de nuestras técnicas de análisis y selección mediante espectroscopía óptica, y observamos en radio una submuestra de cuásares de especial interés. Analizamos los resultados de nuestras observaciones y obtuvimos conclusiones tanto en lo referente a las técnicas empleadas para la selección de candidatos como sobre la física de los objetos estudiados.

En el capítulo 1 exponemos una visión de conjunto de los cuásares, desde su descubrimiento a sus propiedades principales. Centramos nuestra atención en los cuásares radio-intensos, en su caracterización teórica y observacional. Finalmente aportamos un resumen acerca del debate sobre el origen de la existencia de las dos poblaciones radio-intensa y radio-silenciosa.

En el capítulo 2 presentamos los resultados de nuestra selección original de RL QSOs con $3.6 \leq z \leq 4.4$, basada en un algoritmo de red neuronal. Se seleccionó la muestra inicial mediante un *cross-match* entre fuentes de radio de FIRST (del catálogo publicado el 11 de abril de 2013), y objetos de morfología estrellar del catálogo fotométrico SDSS DR7. En este capítulo analizamos los cartografiados empleados, los criterios de preselección aplicados a la muestra inicial, y proporcionamos una visión general del aprendizaje automático y de las redes neuronales artificiales (ANN). Por último, tratamos en particular el algoritmo de red neuronal *feed-forward* que se empleó en este trabajo y presentamos la lista de 15 candidatos RL QSOs en el rango de desplazamiento al rojo $3.6 \leq z \leq 4.4$, seleccionados de entre una muestra de 2.916 fuentes radio-ópticas sin identificación

espectroscópica.

En el tercer capítulo presentamos el seguimiento espectroscópico de los 15 candidatos seleccionados previamente. Una vez identificados los tipos espectroscópicos y los desplazamientos al rojo de los objetos observados, comparamos la eficiencia y completitud teóricas de nuestra selección de red neuronal (basada en la muestra de prueba) con los resultados reales para los objetos sin espectros previos. De este modo confirmamos los resultados previstos y demostramos que nuestra red neuronal puede seleccionar QSOs de alto desplazamiento al rojo de cartografiados radio-ópticos, con un 97% de completitud y un 60% de eficiencia. Comparamos nuestra metodología con el algoritmo de selección de SDSS y concluimos que, a pesar de que las eficiencias de las dos metodologías son comparables, nuestro algoritmo de red neuronal detecta el $\sim 97\%$ de los QSOs con alto z , mientras que SDSS detecta sólo el $\sim 85\%$ de ellos.

En el capítulo 4 llevamos a cabo la determinación más precisa de la función de luminosidad óptica para estos objetos conocida hasta el momento. De hecho, la muestra de cuásares radio-intensos con $3.6 \leq z \leq 4.4$ y $M_{1450} < -27$ está compuesta de 72 QSOs conocidos, sin embargo esta muestra reducida ha sido aumentada a 87 objetos con nuestros métodos de selección con redes neuronales artificiales. Presentamos los datos que empleamos para nuestra determinación de la función de luminosidad y discutimos la corrección K. Asimismo, hacemos una estimación de la completitud de nuestra muestra e ilustramos nuestra nueva metodología para el cálculo de la función de luminosidad. Finalmente, calculamos la función de luminosidad y discutimos los resultados de nuestra determinación mediante su comparación con otros resultados y modelos en la literatura. Determinamos la pendiente de la función de luminosidad en dos rangos de desplazamiento al rojo, hallando valores de la pendiente consistentes con otras determinaciones realizadas para las poblaciones radio-intensa y radio-silenciosa. La consistencia de nuestros resultados sugiere una evolución similar para ambas poblaciones en este rango de desplazamiento al rojo. Nuestros resultados pueden también sugerir un aplanamiento de la pendiente *bright-end* de $z \sim 2$ a $z \sim 4$, exclusivamente para la población radio-intensa. Si esto se confirmase, implicaría una evolución de la densidad de los agujeros negros supermasivos asociada a QSOs radio-intensos, de modo que éstos serían más abundantes en $z \sim 4$. Sin embargo, para clarificar la evolución de la población RL en relación a la población total de QSOs son necesarios más datos observacionales, especialmente en desplazamientos al rojo por encima de 4.

En el quinto capítulo, exponemos la selección de una muestra de 22 RL *Broad Absorption Line* (BAL) QSOs con alto z entre $3.6 \leq z \leq 4.8$, ampliando el número conocido de objetos de esta clase. Posteriormente, observamos los objetos más luminosos de esta muestra en varias frecuencias en radio (en el EVLA radio-array y en el telescopio 100m de Effelsberg), junto con una muestra de com-

paración de RL QSOs no-BAL en el mismo intervalo de desplazamiento al rojo. Con estos datos, pusimos a prueba los dos escenarios principales que han sido propuestos para explicar la existencia de estos objetos, en particular testamos el "escenario de orientación", calculando la orientación del *jet* respecto a la línea de observación. Las formas de los espectros radio nos permitieron también evaluar estadísticamente la edad radiativa de las fuentes (contando el número de fuentes CSS (*Compact Steep Spectrum*) o GPS (*Gigahertz Peaked-Spectrum*) y probando, por tanto, el "escenario de evolución". Empleamos también nuestra muestra de RL high z BAL QSOs para comparar sus colores ópticos de banda ancha con la población restante de cuásares en el mismo rango de desplazamiento al rojo. La motivación de esta investigación ha sido la de determinar si las características halladas en los cuásares BAL con menor desplazamiento al rojo son representativas de la población de RL BAL de alto desplazamiento al rojo o si existe alguna indicación de tendencias evolutivas en las propiedades de los BAL.

En el sexto capítulo presentamos un catálogo de fuentes detectadas en el radio (FIRST) y en el infrarrojo (UKDISS) pero no detectadas en el óptico (SDSS), y una lista de candidatos a cuásares enrojecidos. La mayor parte del trabajo trata los métodos desarrollados para 'limpiar' la muestra de falsas fuentes y para asociar a cada objeto final la probabilidad de ser un objeto radio-infrarrojo real con luminosidad óptica por debajo del flujo límite de detección de SDSS. La segunda parte consiste en la selección de la muestra de candidatos QSOs muy enrojecido en el óptico. Además de en el catálogo en sí, el interés de esta investigación reside en la recopilación de técnicas de minería de datos utilizadas y en la ilustración de los problemas que pueden surgir en este tipo de análisis.

En el capítulo 7 exponemos la investigación que estamos desarrollando actualmente acerca de la selección de candidatos RL cuásar con $z > 4.7$ empleando los datos en radio (FIRST), en el óptico (SDSS DR10) y en infrarrojo (UKDISS Large area Survey DR10, WISE, VHS). Presentaremos una lista de candidatos $4.7 \leq z \leq 5.1$, así como los resultados y conclusiones preliminares obtenidos a partir de su observación parcial.

Finalmente, en el capítulo 8 presentamos las conclusiones y la perspectiva de futuro que se abre a partir de la investigación realizada en esta tesis doctoral.

Chapter 1

Introduction

*Studia prima la scienza, e poi seguita la pratica, nata da essa scienza.
Quelli che s'innamoran di pratica senza scienza son come'l nocchier
ch'entra in navilio senza timone o bussola, che mai ha certezza dove si
vada¹*

Leonardo da Vinci

1.1 Quasars: historical background

The discovery of the first quasar finds its origin and foundation on the determination of accurate position of radio-sources, in an era in which their typical positional accuracy was of the order of a few arcmin. In 1962 Cyril Hazard successfully tried out a new method to determine the position of a radio source in the sky, obtaining the coordinates of the radio-source 3C-273 (a source with two components separated by about 20 arcseconds) with an accuracy of ~ 1 arcsec (Hazard et al. 1963). This positional accuracy was sufficient for astronomers to identify the optical counterpart of the source, and in 1963 Marteen Schmidt took its optical spectrum (Schmidt 1963). The optical counterpart of 3C 273 turned out to be a very bright star-like source ($B = 13.1$ mag) whose spectra was unusual because it had four broad emission lines. But the bigger peculiarity of its spectra was the redshift implied by the wavelength of those emission lines, that if of cosmological origin, implied a redshift of $z \approx 0.158$.

In the early 1960s the measured redshifts of galaxies usually ranged up to $z \sim 0.2$, and the more distant galaxy known was 3C 295, discovered by Minkowsky in 1960 (Minkowski 1960) and having a redshift of 0.46. Therefore the big news

¹Study the theory before of the practice . He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast

about 3C273 was not the high redshift by itself, but the intrinsic brightness implied for this star-like source placed outside of our galaxy. Being the bolometric luminosity of 3C273 $\sim 10^{47}$ erg/s, it was a source a thousand times brighter than whatever was known so far, having the order of the luminosity of 1000 bright galaxies. The hypothesis that the quasar redshift was not cosmological was object of intense debate in the first decade of research on quasars. But the alternative-hypothesis on the cosmological origin of quasar's redshift were generally ruled out by an increasing volume of observational evidences. The role of theoretical inputs given by quasars was considered so important that many international symposium were convened in the years subsequent the 1963, bringing together general relativists, theoretical astrophysicist and observers to discuss the implications of the discovery of these remarkable objects. The field of research opened with the discovery of 3C-273, has been so profitable, extended and full of exciting overturning, that the name itself chosen to indicate this class of objects it is nowadays somehow contradictory. In fact, the word "quasar" was chosen to indicate "quasi-stellar radio source" but now we know that the large majority of the known quasars are not active in the radio and that only about the 50% of this type of objects are quasi-stellar (Stern et al. 2012). More than 50 years after their discovery, quasars are still one of the more vibrant scientific field of research. Many of the questions that dominated the first decades of quasar research have been definitively resolved, while others have risen and are being object of great debate and research, promising to have profound implications on our understanding of the universe and its origins.

1.2 Active Galactic Nuclei properties

Short after the discovery of the first quasar, many other objects with similar characteristic were discovered and studied. The great brightness was not the only peculiarity of these objects and in the early years of quasar searches, the following broad criteria specified by Maarten Schmidt were established to distinguish a quasar from a star or a galaxy:

- i Star-like object identified with a radio source
- ii Large ultraviolet flux of radiation
- iii Large redshift
- iv Variable light
- v Broad emission lines in the spectra with absorption lines in some cases

Later searches showed that quasars are part of a larger family called Active Galactic Nuclei (AGNs) characterized by many, but not all the properties indicated by Schmidt. In fact, nowadays we know that the radio property (i) is not generic of all AGNs and only $\sim 10\%$ of them may be considered radio-loud. On the other hand AGNs are considered the most powerful not-explosive sources in the universe, and they all show large intrinsic brightness. (ii) The latter property make them visible at large distances (iii) and for many years, in the observational "hunt" for distant objects, quasars have been the "most distant object" observed. The spectra of quasars have allowed pioneer studies on the physical environment of the universe at large redshift. Nowadays the most distant object observed is not a quasar ² anymore, but they still play a key-rule in observational studies on the early universe.

Another early surprise about quasars involved the discovery that their emission was time variable in intensity and on small timescales (week/months) in basically all the wavebands (iv), tending to show larger variations at shorter wavelength. This property has been confirmed as a defining characteristic of AGNs, implying very compact (in the order of light months, i.e. \sim a few 10^3 Schwarzschild radii) emitting volumes (Greenstein & Schmidt 1964). The time variability of the emission was one of the first properties of quasars to be explored in detail (e.g Smith & Hoeffleit 1963) and has turned out to be a powerful tool to study AGNs physical properties. Studies sampling the variability of certain broad emission lines in response to continuum variations, the so called "reverberation mapping", have been widely used to infer the size of the broad line region and to estimate the black hole masses (see for example Peterson & Horne 2004; Bentz et al. 2009). Even the most elusive form of photometric variability, i.e. "microvariability" provides valuable information on the accretion process. Although the exact reasons of the variability in AGN are still an open discussion, there is a general agreement that it could be originated in the accretion mechanisms and that the mass accretion rate onto the black hole is likely the fundamental parameter underlying thus property.

Another early noticed observational characteristic of AGN is the presence of broad emission lines (v) with widths upwards of 1000 km/s. The strongest observed lines are the Balmer-series lines ($H\alpha$ $\lambda 6563$, $H\beta$ $\lambda 4861$ and $H\gamma$ $\lambda 4340$), hydrogen $Ly\alpha$ $\lambda 1216$, and prominent lines of abundant ions ($Mg II$ $\lambda 2798$, $C III$ $\lambda 1909$ and $C IV$ $\lambda 1549$). This property has been revealed to be especially relevant for the selection of quasars through broad-band photometry. In fact, the presence of broad emission lines (in particular $Ly \alpha$ emission) in a given band enhances its total flux rendering the color of the quasar generally distinguishable from other

²At the time, the most distant objects is a γ ray burst (Tanvir et al. 2009), having redshift $z = 8.2$, and in general there are several galaxies at redshift ~ 8 . The most distant observed quasar has $z \sim 7.1$ and was discovered in 2011 UKIDSS (Mortlock et al. 2011)

kinds of objects.

Another feature common to all AGNs that is worth pointing out is that they are a strongly evolving population. AGNs are bright in the close universe, but they were even more powerful in the past, with a peak at $z \approx 2$. Studies on the demographic evolution of AGNs, obtained with the determination of their Luminosity Function, play a fundamental role to understand the nature of Black Holes and on the impact of AGNs on the evolution of galaxies and on the early universe. In Chapter 4 of this research-thesis we will discuss the current "state of the art" on the AGN luminosity function and we will present our results on its determination.

1.3 Black Hole paradigm

The basic physics of quasars was recognized quite soon after Schmidt's interpretation of the redshift of 3C 273, by Zel'dovich & Novikov (1964), Salpeter (1964) and Lynden-Bell (1969). Their model predicted that the quasar was powered by accretion material falling onto a supermassive black hole (SMBH) with a mass $> 10^6 M_\odot$. It was a powerful model because it explained at once: the extreme luminosities of quasars based on the release of gravitational energy through accretion phenomena; the broadening of the emission lines in terms of doppler effect from gas clouds moving with high velocities; the small size of the emitting regions and, related to it, the short variability time scales of AGN.

A detailed treatment of the physics of black holes requires general relativity, however it is useful to remind here some of their fundamental properties. The basic idea defining a black hole, is that it is an object so compact that even an escape velocity equal to the velocity of light it is not sufficient to gain its gravitational force³. A black hole is then defined by its "event horizon" R_S , i.e. the radius within which matter and light fall inward the black hole but can never re-emerge. The expression for R_S (also called *Schwarzschild radius*) can be easily derived using classical mechanics posing the escape velocity equals the speed of light, thus:

$$R_S = \frac{2GM_{BH}}{c^2} \quad (1.1)$$

³The hypothesis of the existence of such kind of objects was discussed first by John Mitchell in 1784. Independently, in 1796, Pierre-Simon Laplace developed the same concept in what he called the "dark stars". The basic idea was then abandoned since the light was considered as made of massless particles. Black holes re-appeared in general relativity as an hypothetical solution of the gravitational field generated by a spherical distribution of mass, in the Schwarzschild derivation (1916). But only after the discovery of AGN the black-hole hypothesis started to be considered more than a pure speculation.

where G is the gravitational constant, c is the speed of light and M_{BH} is the mass of the black hole. The highest luminosity that a source of mass M can have and still be stable against radiation pressure is known as the "Eddington luminosity":

$$L_{edd} \leq \frac{4\pi GMm_p c}{\sigma_T} \approx 1.3 \times 10^{38} \frac{M}{M_\odot} \quad (1.2)$$

where σ_T is the Thomson scattering cross-section and m_p is the proton mass. The rate at which the potential energy of infalling material is converted onto radiation is given by:

$$L_{acc} \approx \frac{dU}{dt} = \frac{GM\dot{m}}{R} = \eta c^2 \dot{m} \quad (1.3)$$

in which the accretion luminosity L_{acc} is driven by gravitational potential, and it is assumed that the energy extracted by the in-falling material of mass m is some fraction of its rest energy $E = \eta mc^2$, with η being the efficiency factor. A rough estimation of η can be obtained considering that most of the optical/UV continuum radiation is originated at $\sim 5R_S$. Therefore changing in the expression above R for $5R_S$, we have that $\eta \sim 0.1$, the calculation suggest that η is an order of a magnitude more efficient than fusion of hydrogen to helium⁴, for which $\eta = 0.007$

The "Eddington power" \dot{M}_{edd} is defined as the mass accretion rate needed to sustain the Eddington luminosity and, known η , it is easily calculable from:

$$\dot{M}_{edd} = \frac{L_{edd}}{\eta c^2} \quad (1.4)$$

From this expression it is interesting to notice that for a luminous AGN having $L = 10^{46} \text{ ergs}^{-1}$ it is "only" needed an accretion rate of $\sim 2M_\odot \text{ yr}^{-1}$.

Although the black hole paradigm is long-dated, actual observational evidence for black holes in AGN took much longer to emerge and the existence of black holes has been questioned until less than a couple of decades ago⁵. Nowadays the situation has changed and the last 2 decades of research have accumulated several proving studies on black holes existence, included the one in our own galaxy (see for example Ghez et al. 2005). In general there are compelling evidences that back up the existence of SMBH and not only at the center of the AGNs, but also at the center of almost all the galaxies that have a strong spheroidal component (Peterson 2008).

⁴A more accurate calculation of the efficient rate give $\eta \sim 0.06$ for a non rotating black hole (Schwarzschild metric) and $\eta \sim 0.42$ for a rotating black hole (Kerr metric)

⁵In 1992 Blandford (Blandford 1992) wrote that: "it remains true that, even in the lax standards of astronomy, there is no proof that black holes exist in an AGN or indeed everywhere else".

1.4 AGN taxonomy and unified scheme

In the years following the discovery of the first quasar, increased considerably the number of objects discovered or recognized as having properties similar to those considered as characteristic of quasars. The term "active galactic nucleus (AGN)" was used for the first time in 1971 by Ambartsumian (Ambartsumian 1971) and soon became popular to indicate the whole family of objects whose energetic output was mainly attributable to gravitational accretion onto a supermassive black hole. The AGNs are divided into a considerable number of classes and subclasses based on different defining observational characteristic. As pointed out by Tadhunter (2008), the taxonomy of AGNs is complicated from the fact that they emit powerfully over the full electromagnetic spectrum accessible and AGNs were discovered and classified separately at different wavelengths. The first 25 years of research on AGN culminated successfully in an attempt to simplify the classification of these objects, proving effectively that most of the differences among AGN classes were not caused by real physical differences and that the distinct appearance of AGNs can be explained by a common underlying model. The so called "AGN unification scheme" model, occurred in the mid '80s and over the last decades have withstood most major statistical tests. However, it fails to explain all the differences observed in AGNs and in particular cannot explain the connection between radio-loud and radio-quiet phenomena.

The root of unification is represented by the Antonucci & Miller (1985) article on polarization of the Seyfert galaxy NGC 1068. Seyfert galaxies were studied for the first time by Carl Seyfert (Seyfert 1943) two decades before the discovery of quasars, as unusual nearby spiral galaxies having a bright central core and a bluer optical spectrum with strong, broad emission lines. Seyfert galaxies were recognized as part of the AGN family and they were eventually classified in Type 1 and Type 2 depending on the presence (type 1) or absence (type 2) of broad permitted lines, in addition to narrow forbidden lines of [OIII], [OII], [OI], [NeIII], [NeV], [NII] that are a feature of both classes (Khachikian & Weedman 1974). In the model proposed by Antonucci & Miller (1985), the key element to explain the observed differences between the two types of Seyfert was constituted by an opaque dusty structure, described as a torus, surrounding the black hole accretion region. The observational differences between the two types of Seyfert galaxies were explained in terms of orientation of the obscuring torus relative to the line of sight: so if the torus is seen face-on, our view of the central regions is unobscured and we detect the broad lines; if our view is closer to edge-on, the central regions are not seen directly and no broad lines are detected.

The fundamental idea of an orientation-based unified scheme was successful. Extending it to all types of AGN, it affirms that they belong to the same parent population and they have similar intrinsic properties. Differences between the

observed properties of the classes of AGNs are due to anisotropy and orientation effects. According to the unified scheme the main components of a typical AGN are summarized in Fig. 1.1 and they are:

- A *supermassive black hole* having a mass $M = 10^6 - 10^9 M_\odot$, the "central engine" of the AGN.
- An *accretion disk*, surrounding the central black hole resulting from the residual viscosity and angular momentum of the accreting gas. A simplified model (Shakura & Sunyaev 1973) for the disk predict that it is geometrically thin and optically thick, emitting as a black body and therefore having luminosity $\propto 2\pi R^2 \sigma T^4$. The temperature of the disk (up to $T \sim 10^6 \text{K}$) depends on the distance R from the black hole (outer layers are cooler than inner ones) and on the black hole mass as: $T(R) = \left(\frac{GM_{BH}\dot{m}}{4\pi\sigma R^3}\right)^{1/4}$. Thermal emission from the disk accounts for most of the UV/optical continuum radiation of the AGN.
- The outer part of the accretion disk is the inner edge of the "*Broad-Line Region*" (*BLR*), probably made by gaseous clouds close to the engine (up to 100 light days) having a density $n_e \sim 10^{10} - 10^{11} \text{cm}^{-3}$, too high to emit forbidden lines, and a temperature $T \sim 10^4 \text{K}$. Photoionization is the dominant physical process in the BLR gas. The emission lines (like the hydrogen Balmer series) have a large width, typically a few 10^3km/s . The principal broadening mechanism is Doppler motion of individual clouds, being the thermal broadening $\Delta\nu \sim \left(\frac{kT}{m_p}\right)^{0.5} \sim 10 \text{km/s}$ for a $T \sim 10^4 \text{K}$ gas.
- A dusty cool *torus* surrounding the central regions at a distance $R \sim (0.3 - 3 \text{pc})$. Its temperature must be lower than the dust sublimation temperature, which can vary for different dust composition (e.g. $T_{subl} \sim 900 \text{K}$ for silicates, $T_{subl} \sim 1800 \text{K}$ for graphite). The presence of the torus is necessary to explain why type 2 AGN are obscured, but its structure is still not well known: several models were developed proposing a homogeneous or clumpy distribution with different density and temperature profiles, or even an outflowing wind driven by a magnetic field. The torus emits thermally in the infrared.
- All the previous structures are embedded in a much larger region, up to $(100 - 1000)$ parsecs, of low-density gas ($n_e \sim 10^3 - 10^4 \text{cm}^{-3}$ at a temperature $T \sim 10^3 \text{K}$): the "*Narrow-Line Region*" (*NLR*). This region produces the narrow forbidden lines having typically a width of a few 10^2km/s . Forbidden lines, like [OIII] at 5007\AA and 4959\AA , are emitted because the low

density of the gas decreases the probability that collisional de-excitation occurs before the radiative de-excitation which, for forbidden lines, has very long time scales.

- A gaseous rarefied *hot corona* located above the disk that has a diffuse or clumpy structure depending on the models and a temperature $T \sim 10^9\text{K}$. The hot corona is necessary to explain the high-energy X-ray emission through Inverse Compton (IC) interaction with the low energy thermal photons emitted by the disk, but the heating mechanism is not known yet.
- Relativistic jets, perpendicular to the torus plane, which emit mostly through synchrotron mechanism from the radio band to the γ -rays and that can reach the length of $R \sim 1\text{Mpc}$. The mechanism originating them is not completely understood, but it seems to be related to the black hole rotation-driven magnetic field (Pudritz et al. 2012). The interaction between jets and environmental medium causes the formation of the hot spots and the lobes. Only radio-loud AGN show the presence of prominent jets.

Given the described, general model of AGN, the different features of the several classes of these objects can then be explained in terms of structure and orientation, as can be seen in Fig. 1.1. The difference between radio loud and radio-quiet AGN is explained in terms of presence or absence of the radio-jet. The radio emission becomes more and more important as the line of sight approaches the jet axis. This is easily explained by the relativistic beaming affecting the synchrotron emission. Thus radiogalaxies, i.e. radio-loud AGN for which the line of sight forms a wide angle with the jet axis, are weaker radio sources than blazars, for which that angle is very small or null. The orientation does not affect the mid-IR luminosity (mainly emitted by the dusty torus itself) which is indeed equal for the two classes.

The orientation-based unified schemes are highly successful at explaining some aspects of AGN classification, in particular, the relationship between broad and narrow line AGN. However there is a lot that it cannot explain like the already mentioned relation between radio-loud and radio-quiet quasars and other aspects like the differences found in the luminosity functions of different AGN types (e.g. in the blazar population BL LAC and FSRQ), or the fact that variability does not show the same patterns independently of the source type. It appears evident, nowadays, that although the Unification scheme has yielded key information to understand the structure of AGNs, it also represents a generalization, an attempt to simplify a complex situation.

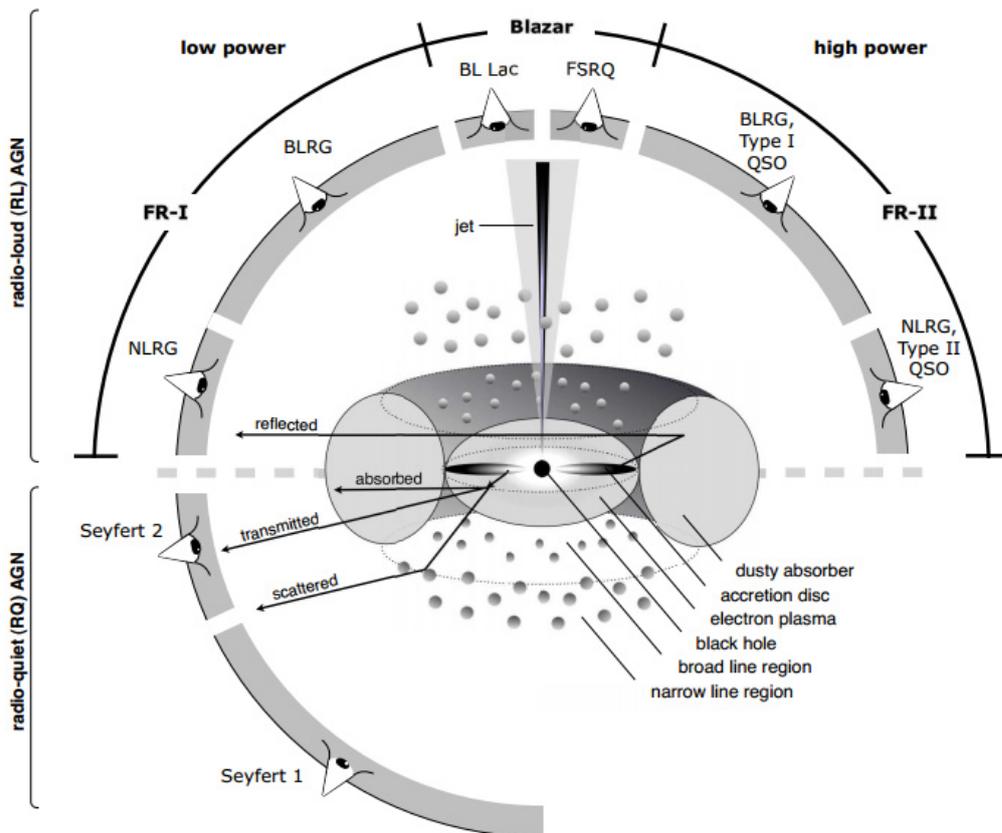


Figure 1.1: Schematic representation of the unification scheme for the AGN phenomena. The viewing angle determines the type of object observed. The difference between radio-loud and radio-quiet AGNs is given by the presence or not of the radio-jet.

1.5 Radio Loud Quasars

Although the first quasar discovered was radio-loud (RL), it is now known that only a minority of their whole population is a powerful radio emitter. In particular only the $\sim 8 - 13\%$ of the known quasars are classified as RL, while the remaining part is classified as Radio-Quiet (RQ) (Ivezić et al. 2002; Jiang et al. 2007; Baloković et al. 2012); after more than 50 years of research, the reason of the existence of these two major classes still lacks a convincing explanation. In the AGN unification scheme (see previous section) the difference between radio-loud and radio-quiet AGN is explained in terms of presence or absence of the radio-Jet, responsible for the synchrotron emission. But it is not known why some QSOs produce a Jet and others do not, neither we know why it is only a minority of QSOs that rise a jet.

1.5.1 Morphology

AGN radio emission is non-thermal and it is originated by synchrotron mechanism, resulting in a power law spectrum (e.g. Risaliti & Elvis 2004). Relativistic electrons having a Lorentz factor γ and a mass m_e are accelerated by a magnetic field of intensity B and emit at a characteristic frequency $\nu_S \propto BE^2$, where $E = \gamma m_e c^2$.

Morphologically, radio-loud sources consist broadly of two components, a steep-spectrum extended source, which generally has a double-lobed structure, and a highly beamed flat-spectrum core, i.e with $\alpha < -0.5$, where α is the index of the power law spectrum $F_\nu \propto \nu^\alpha$. The lobes of the extended component are more or less symmetrically located on either side of the optical quasar or center of the galaxy; their linear extent can be as large as megaparsecs. The position of the optical quasar is coincident with that of the compact radio source. The core is often seen, at high angular resolution, to originate a linear extended jet structure (Bridle & Perley 1984). The appearance of the jet suggest that they transport energy and particles from the compact source to the extended regions. Unlike the extended components, the jet structures are almost always one sided, but co-aligned with the axis of the extended radio-lobes. The usual explanation for the one-sidedness of the jets is that their radiation is relativistically beamed with fairly large Lorentz factors, which strongly favor the detection of the approaching jet over the receding one (Garrington et al. 1988; Laing 1988).

The relative strength of the extended, compact and jet components varies with frequency since the different components have different spectral shapes. The relative strengths also show considerable variation from source to source, with fainter, core dominated FR-I and brighter, lobe-dominated FR-II radio-sources (in agreement with the scheme developed by Fanaroff & Riley (1974). In Fig. 1.2 we show

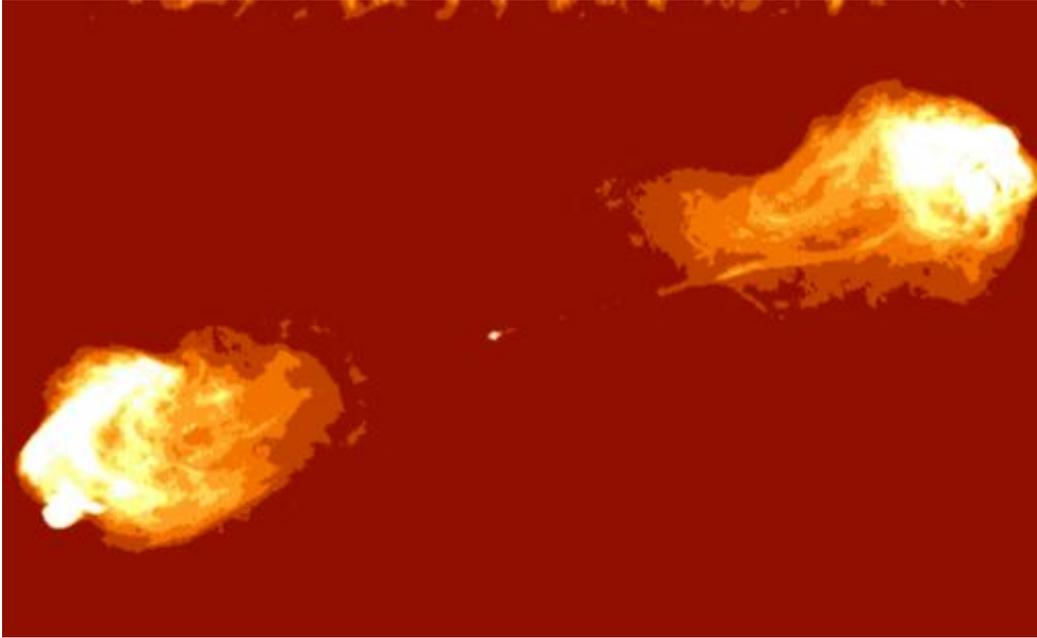


Figure 1.2: *This image is a radio map (at wavelength of 22 cm) of the powerful radio galaxy Cygnus A, produced from the observations at the Very Large Array in all four of its antenna configurations. The compact nucleus is the small point in the center of the image. The nucleus is emitting jets of material in opposite directions. These jets impact material surrounding the galaxy, giving rise to the giant "lobes" of radio emission seen in this image. Cygnus A is a FRII radio galaxy and among the first such galaxies to be detected, it has played a key role in the field of radio-loud active galaxy research (e.g., Burbidge et al. 1963).*

the radio map of Cygnus A, a well known example of bright FRII radio-galaxy.

The radio-loud quasars are also sub-divided on the basis of whether they are steep radio spectrum dominated (steep spectrum radio loud quasars: SSRLQ), flat radio spectrum dominated (FSRLQ), core-dominated or lobe dominated. Nevertheless, following the unified scheme that we saw in the pervious section, part of the observed differences among radio loud are explainable in terms of orientation.

1.5.2 Radio loudness

Since the boundary between RL and RQ QSOs is phenomenological and a physical explanation of their distinction is still elusive, it is not surprising that there is no unique definition of radio-loud quasar. Two are the most common criteria employed in the literature to assign a quasar the status of "radio loud". The first one is based on the radio luminosity P emitted by the source (e.g. Gregg et al.

1996), and defines an object to be radio-loud if $\log P_{1.4, \text{GHz}} (\text{W/Hz}) > 25.5$). The second one is the so called " R -parameter" and it is geared by the rest-frame ratio of the flux density at 6 cm (5 GHz) to the flux density at 2500 Å (e.g., Stocke et al. 1992). Generally, objects are considered to be RL if $R > 10$, and RQ if $R < 10$ (Kellermann et al. 1989). It is still an open question whether one of these two criterion is more relevant or more justified than the other.

While some authors argue in favor of the radio luminosity as the real fundamental parameter to discriminate (Miller et al. 1990) the two classes of AGN, others favor the R -parameter because, relating the radio and the optical luminosity, it offers a scaling of non thermal and thermal mechanism at work within the quasar. The definition of the R -parameter, by comparing the radiation mainly originated from the disk (optical) with the one mainly originated from the jet (radio), also has the advantage of making clear that the difference between radio loud and radio quiet phenomena is most likely originated from a real underlying physical difference in the quasar structure. On the other hand the R -parameter is basically meaningless when a significant part of the the optical radiation is extinct (for example for many Type 2 QSOs). Also, since R tends to be luminous dependent ($R \sim L^{-0.5}$, Laor 2003) some authors have pointed out that using a fixed boundary of $R = 10$ may not be appropriate for lower luminous sources (Ho & Peng 2001) and artificially create biased trends in the demography of the RL population. Some further attempt to improve the definition of radio loudness include the incorporation of the radio morphology in the definition (Sulentic et al. 2003), but in general there is a general agreement to consider that the adopted definition to discriminate RL from RQ QSOs should be always determined by the specific outline of the problem meant to be addressed.

1.5.3 Radio-loud Fraction and Dichotomy

The exact Radio Loud Fraction (often indicated with the acronym RLF), i.e. the fraction of quasars that is radio loud out of the whole population of quasars, is still unknown and it has been subject of considerable debate over the past years. It is useful to remind that radio-quiet is not the same as radio-silent and that at a certain flux level probability all AGN emit radio waves, therefore the exact RLF also depends on the adopted definition of radio-loudness. However, the main theme of debate on the RLF has been focused on its behavior regarding redshift and luminosity. In fact, some studies have found no evidence for significant change of RLF with either redshift or luminosity (e.g., Goldschmidt et al. 1999; Stern et al. 2000; Cirasuolo et al. 2003; Vigotti et al. 2003), while others have found that the RLF changes with both parameters (e.g., Miller et al. 1990; Visnovsky et al. 1992; Schneider et al. 1992; Jiang et al. 2007; Baloković et al. 2012). In particular Jiang et al. (2007), using a sample of 30,000 optically selected quasars from the

SDSS quasar catalogue obtained from Data Release 3 Schneider et al. (2005), it has been found that the RLF of quasars decreases strongly with increasing redshift and decreasing luminosity:

$$RLF/(1 - RLF) = b_o + b_z \log(1 + z) + b_M(M_{2500} + 26) \quad (1.5)$$

Where M_{2500} is the absolute magnitude at rest frame 2500Å, and the parameters b_z and b_M depend on the definition of radio loudness. In this thesis we will test the Jiang et al. (2007) determination, based on our determination of luminosity function.

The question of the RLF is intimately connected with the debate on the dichotomy between RL and RQ AGN, i.e. if the two populations form part of two distinct populations or they are part of a unique continuous sequence. There is a vast literature on this question that seems by itself bimodal. But, although the existence of the dichotomy has been questioned (e.g., Cirasuolo et al. 2003), in the last years there is mounting evidence in its favor. In fact, on the observational side there is a lack of evidence of continuous radio-properties in the AGN population. For instance, Seyfert galaxies are classical examples of RQ AGNs, but they are also weak radio emitters. High resolution observations of the radio cores of Seyfert galaxies (Lal et al. 2011) do not seem to detect any relativistic beaming, which would be a clear indication for a jet; therefore indicating that the radio emission in Seyfert is driven by different mechanisms from the ones responsible of the radio emission in RL quasars.

Two recent and complete studies on this dichotomy also argue in favor of the existence of two distinct populations. The first, by Baloković et al. (2012), uses a Monte Carlo simulation and the SDSS DR7 Quasar catalogue matched to the FIRST survey, showing that the radio-loudness is inconsistent with a single distribution of the radio-loudness and the distribution of quasars likely consists of at least two components. The second work, by Kratzer (2014), analyze critically the literature against the dichotomy, and conclude that also the results presented in literature as indications against the dichotomy, can be actually re-interpreted as in agreement with a bimodal distribution.

However, despite the indications in favor of this dichotomy, as pointed out by several authors, other effects like AGN evolution and biases in their selection may cloud the analysis (Laor 2003). Raising the number statistics and reducing the selection biases linked to flux-limited samples would probably abate the main source of discrepancies (Ivezić et al. 2002). In fact, when White et al. (2007) show the existence of a shallowing minimum between the RL and RQ parts of the radio loudness distribution, it is stressed that optical selection effects probably dominate this distribution.

In this context, studies on the RLF for high redshift RL QSO are particularly

interesting because differences, if any, in the early evolutionary behavior of radio and optical activity would provide valuable clues in favor or against the existence of two distinct populations. Unfortunately, as we will discuss in this thesis, at high- z the biases and the statistical uncertainty are enhanced.

1.5.4 Phenomenology and origin of RL QSOs

Among the models considered to explain the production of the radio-jet, one of the most commonly quoted parameter is the spin of the black hole (Sikora et al. 2007). In general relativity, Kerr-metric describes the space-time geometry in the vicinity of a rotating black hole, and through the so-called Penrose-process it is possible to extract large amounts of energy from the rotating black hole in the expanse of its rotational energy. This process is suspected to play a role in the launching of AGN jet.

In Wilson & Colbert (1995) the authors have considered a model where the black hole spin rates are determined not by the normal accretion process but by mergers of large black holes following the mergers of their parent galaxies. In their model the rate of the spin⁶ that determines if the radio energy (which appear in form of jets) is mechanically extracted from the black hole and thus whether or not a source is RL. Another model proposed by Garofalo et al. (2010) considers the relative spin on the central black hole with respect to the accretion disk to be the crucial factor here. In their scenario AGN would start with a black hole which has a retrograde spin with respect to the accretion disk, leading to the strong interaction with the disk and thus strong jets. As the black hole is spun up in the direction of the accretion disk, the interaction of the rotating black hole with their magnetospheres becomes less efficient and the jet weakens.

Unfortunately, it is extremely difficult to accurately determine the spin of a black hole, and the few methods known for its calculation, like for example the one based on the shape of the fluorescent Fe K_{α} , are critically connected to the model used for the description of the black hole dynamic.

Nevertheless, in the quest to understand the reasons behind the difference in radio emission for RL and RQ quasars, many have uncovered valuable phenomenological properties that aid in understanding them. Among them, many studies have established confidentially that the majority of RQ quasars generally show smaller values of M_{BH} and larger values of L_{bol}/L_{Edd} (Boroson 2002; Dunlop et al. 2003). Optical analysis of nearby AGN (Capetti & Balmaverde 2005) shows that RLs are hosted by elliptical galaxies while radio quiet agn are mainly (but not

⁶In particular it is the orbital motions of charged particle around the black hole to determine whether or not electromagnetic effects will be powerful enough to generate and collimate a relativistic jets

exclusively) found in spiral galaxies (Hamilton 2010). The latter circumstance has been argued (Wilson & Colbert 1995) due to the fact that black holes in the ellipticals are rapidly spun up due to mergers. In general all these results provide support in favor on the hypothesis that RL and RQ quasars are tied to different evolutionary paths of their hosts through mergers, and interactions. However, there are numerous observational inconsistencies, and in general the problem seems entangled with unknown aspects of the quasars evolution. The ultimate origin of the radio emission, and its related problem of the possible existence of two fundamentally different quasars population in structure and kinematics, are still unanimously considered as unresolved issues.

1.6 High redshift quasars in the epoch of Big Surveys

Over the years after their discover, quasars have been identified using many different methods based on their peculiar emission properties in the various wavebands. Including: radio-selection, optical variability, optical colors, X-ray selection, lack of proper motion, infrared-selection. Most of these techniques can be applied at all redshifts, however, each of them suffers of its own selection bias and the resulting survey can be inefficient or completely miss certain type of quasars.

As it has been already discussed, quasars were first discovered via the optical identification of radio sources (Hazard et al. 1963; Schmidt 1963), a technique that was both effective and efficient because normal stars and galaxies are much weaker sources of radio emission. It was soon realized (Sandage et al. 1965) that the bulk of the quasar population was radio quiet and could be identified through the ultraviolet radiation excess (UV) that quasars demonstrated compared to normal stars, i.e. this excess is traduced , using Johnson UBV photometry, in showing U-V color remarkably small, bluer than stars. Nevertheless, as shown in Fig. 1.3 the most distant quasars were still all initially identified via radio-emission until the end of the 80'. After that, the advent of high speed computers , digital rapid plate-scanning machines and CCDs such as COSMOS and APM, allowed large areas surveys and enabled the extension to higher redshift of color-based techniques for discovering quasars. The machines and multi-color techniques made it possible to use more sophisticated combinations of colors to separate quasars from stars and to provide quantitative estimates of the selection efficiency as a function of redshift and apparent magnitude, which were crucial for determining the luminosity function.

In general, the exponential rise in available computer power and, as a related consequence, the enormous quantities of observed data, primary in digital form,

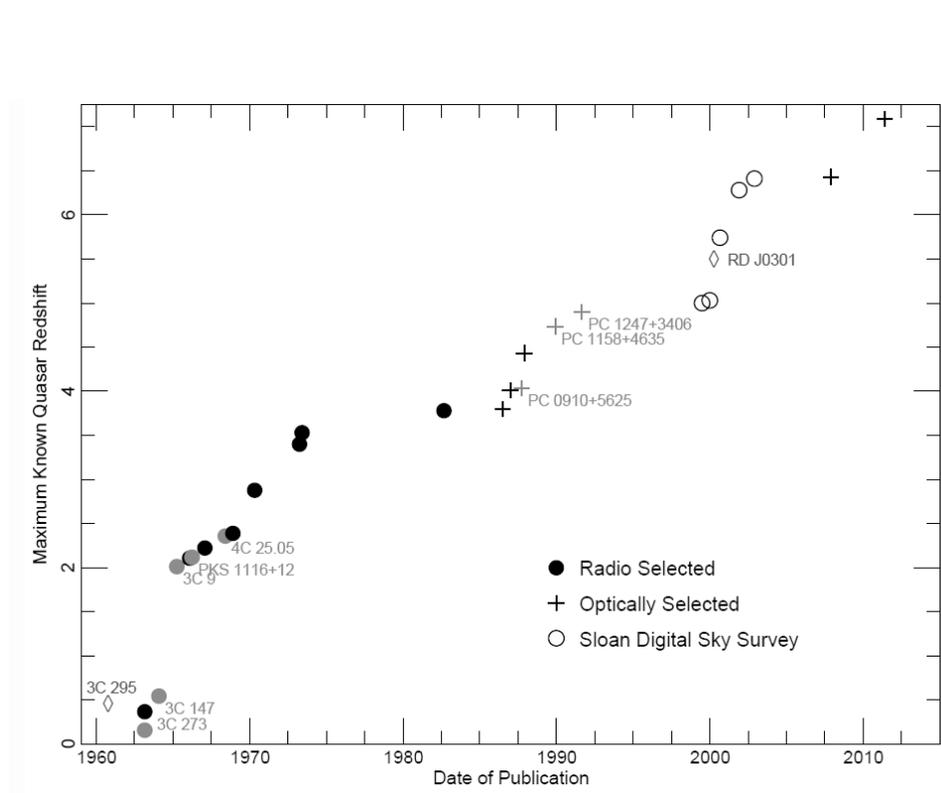


Figure 1.3: Plot of quasars that held the record for "highest redshift" as function to the year of their discovery. Different symbols represent the different methods used for their initial identification.

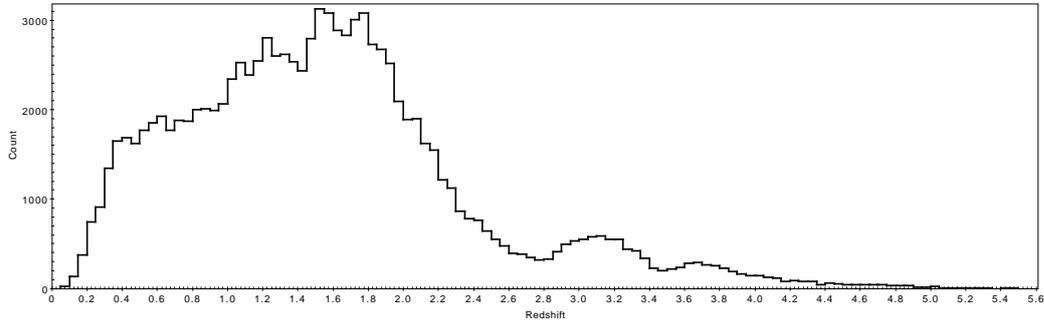


Figure 1.4: Redshift distribution of the $\sim 105,000$ QSOs of the SDSS DR7 quasar catalogue (Schneider et al. 2010).

constitute probably the major change in astronomy in the last decades. The rise in data made necessary the birth of a new field of research known as Data mining, that has been described as the "4th paradigm of science"⁷ Ball & Brunner (2010). Data mining approaches have been effectively used to deal with the sheer volume of data of modern surveys and in the last 20 years surveys of the most luminous quasars are dominated by large projects using wide field optical imaging and spectroscopy. The best examples of them are the 2dF (Two Degree Field) quasar redshift survey (2QZ, Croom et al. 1998) and the quasar survey in the SDSS (Schneider et al. 2010) which discovered respectively $\sim 25,000$ and $\sim 105,000$ quasars. In Fig. 1.4 we show the distribution in redshift of the $\sim 105,000$ quasars of the DR7 QSOs catalogue

Despite their great success, optical surveys can suffer of severe biases. In particular they completely miss obscured quasars and type 2, and they tend to bias against quasars that are modestly reddened due to foreground dust or having weak emission lines. Surveys based on optical colors are also not complete in certain redshift ranges where quasars colors are similar to those of stars or compact galaxies. For example SDSS is notoriously not efficient to find quasars at $z \sim 2.5 - 3$, where the quasar locus cross the stellar in the color space (this effect is visible in Fig. 1.4 as the artificial drop in the population of quasars around that redshift range).

At higher redshift, optical surveys become more and more inefficient in finding quasars, since as the Ly- α forest and the Ly- α emission move toward near infrared waveband. The limitation of optical surveys is assessed at $z \sim 6.4$ where the Ly- α peak around $\lambda \sim 9000\text{\AA}$ and the sources become too faint in the reddest optical band of the survey (in SDSS, the z-band) due to absorption by the intervening Ly- α forest. To overcome this limit and reduce the selection - biases,

⁷The first two paradigm are the well known pair of theory and observation, the third is another relatively recent addition, computer simulation (Ball & Brunner 2010 ; Borne 2009).

data from wide-field near and infrared (NIR, IR) surveys are needed. A number of ambitious NIR-surveys are carried out or in the planning stage, including the UK Infrared Telescope Deep Sky Survey (UKIDSS) (Lawrence et al. 2007) and the Visible and Infrared Survey Telescope for Astronomy (VISTA). Over longer term, ground based LSST survey and ultimately, space based projects such as Euclid, JANUS and WFIRST. Using the UKIDSS Large Area Survey (LAS), Mortlock et al. (2011) discovered the first quasar beyond $z > 6.5$, ULAS J1120 + 0641 at $z \sim 7.1$, and the future surveys will be sensitive to discover a large number of quasars at $7 < z < 10$

The interest in finding quasars at high redshift is mostly connected with the fact that the universe at $z > 7$ contains many landmark events, such as the formation of the first stars and galaxies, the growth of the first massive black holes and the re-ionization of the neutral hydrogen in the intergalactic medium (IGM). The one $z \sim 7$ quasar discovered, already provides constraints on the neutral fraction of hydrogen (Bolton et al. 2011; Mortlock et al. 2011), on the metallicity of the IGM (Simcoe et al. 2012), and on the formation of dust and stars in the host galaxy (Venemans et al. 2012). The population of higher redshift quasars known nowadays sit at the threshold of the earliest generation of quasars, and any discovery of quasars at $z > 7$ or lack of, will provide powerful constraint to the formation of the first supermassive black holes in the universe.

However one of the key issue is whether numerous luminous quasars such as those discovered by SDSS and other surveys at $z \sim 4$ continue to exist at $z > 7$. The comoving density of luminous quasars is determined as a strong function of redshift that peak at $z \sim 2 - 3$ and declines exponentially toward higher (and lower) redshifts. As shown in Fig. 1.5, at $M_{1450} < 27.6$, the comoving density at $z \sim 6$ is ~ 40 times smaller than at $z \sim 3$, as we close in to the epoch of the formation of the first supermassive black hole in the universe.

The correct determination of the quasars density per comoving volume and its evolution with increasing redshift, requires not only the identification of a greater number of quasars, but also the proper valuation of all the selection biases. These valuations are even more crucial for the Radio-Loud population that, representing only the 10% of the whole population of quasars, is affected by much larger errors and uncertainty.

In conclusion, the study of quasars at high redshift is a key element for the understanding of the early universe, but the quest to assess the properties of this population goes through some obligatory steps. These are the implementation and use of the new generation of large-area surveys especially in the near and far infrared, since the combination of field of view and sensitivity will enable to search fainter quasars at higher redshift; the refining of the data mining techniques, in order to deal and combine such a large volume of multi-wavelength data, exploiting their full potentiality. Finally, the improvement of the statistic determination

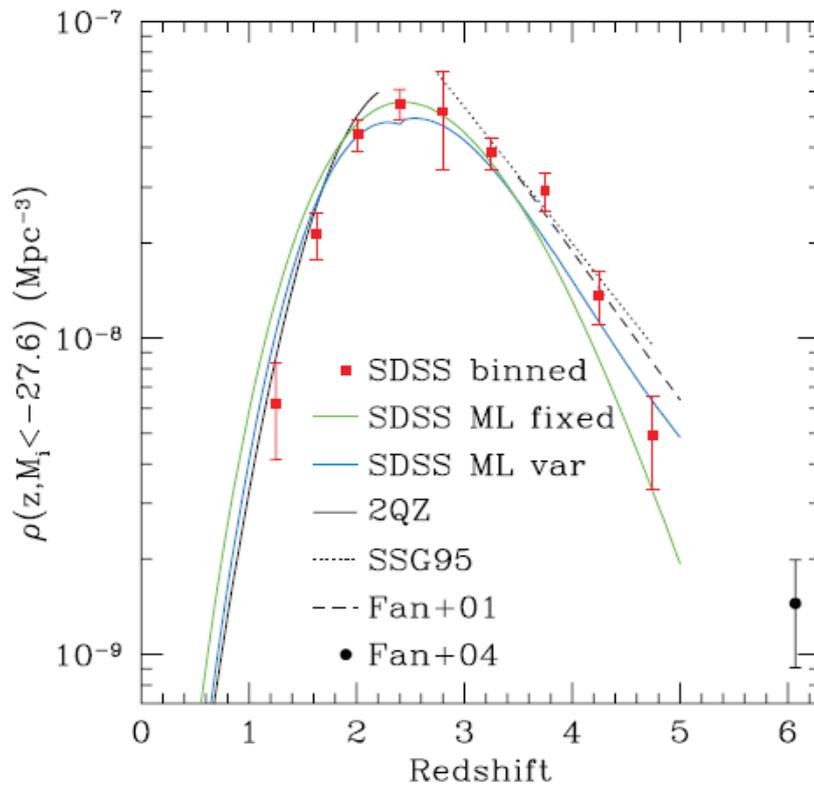


Figure 1.5: Space density of quasars with $M_{1450} < -26.7$, from Richards et al. (2006).

of the selection biases, that have to be carefully corrected in order to derive the statistical properties of the quasar population. The research work presented in this thesis is focused on all these issues related to the selection and identification of high-redshift Radio Loud quasars.

Chapter 2

Selection of RL QSOs at

$$3.6 \leq z \leq 4.4$$

2.1 Introduction

The first catalogue of quasars was published in 1971 (de Veny et al. 1971) and contained 202 objects. Since then, many catalogues have been published, gathering together an increasing number of quasars either from heterogeneous samples (Hewitt & Burbidge 1993; Veron-Cetty & Veron 2010 and previous releases) or from large surveys as LBS (the Large Bright Quasar Survey, Morris et al. 1991; Hewitt et al. 1995), the 2dF (Two degree Field) QSO Redshift Survey (2QZ, Boyle et al. 2000; Croom et al. 2001) and the various releases of the SDSS quasar catalogues (Schneider et al. 2010, see section 1.6).

Large catalogues of quasars covering a broad range of redshift allow a number of fundamental studies like the evolution of the quasar luminosity function, the spatial clustering of quasars as a function of redshift, the quasar correlation function (Croom et al. 2001) and the characterization of the general quasar spectral properties (Croom et al. 2002; Corbett et al. 2003). They have also been used to search for rare/unusual objects (Londish et al. 2002), to carry out lensing studies (Miller et al. 2003) and to place limits on cosmological parameters (Outram et al. 2001; Hoyle et al. 2002).

Many of the studies on the properties of quasars are effective only if the catalogues are constituted on the basis of well-defined uniform selection criteria. When the catalogues are built upon the use of surveys, the selection involves the use of methodologies to target quasar candidates for follow-up spectroscopy and the choice of this methodology assesses the completeness of the resulting spectroscopic survey. As discussed in section 6 of chapter 1, there are many selection methods to identify quasar candidates without spectroscopy, but the most effective

are the ones based on color-color diagrams in the optical, since quasars stand out as a distinct population from stars, making them easy to identify. The selection of quasars from multicolor imaging data was pioneered by Sandage et al. (1965) and has continued through the years (see, e.g., Koo & Kron 1982; Schmidt & Green 1983; Warren et al. 1991; Hewett et al. 1995; Hall et al. 1996; Croom et al. 2001).

The most numerous quasar catalogue is the one based upon the SDSS Seventh Data Release (DR7) and includes a remarkable 105,783 spectroscopically-confirmed QSOs (Schneider et al. 2010). The approach used in the SDSS to select quasars from the imaging data, it is based of their non-stellar colours in *ugriz* bands and by matching unresolved sources to the FIRST radio survey (Faint Images of the Radio Sky at Twenty-Centimeters, Becker et al. 1995). Therefore although it is similar to previously used color-color criteria in the optical, it is enormously improved by the use of large data and the use of an effective decision tree algorithm (Richards et al. 2002). According to Richards et al. (2002), the SDSS target-algorithm is sensitive to QSOs at redshifts $z \leq 5.8$. Completeness (fraction of QSOs selected as such) and efficiency (number of actual QSOs amongst the candidates, divided by the total number of candidates) of the selection are a complex function of apparent magnitude i and redshift. Although QSOs of type 2 and certain QSOs of type 1 are missed, the overall estimated completeness is high (Vanden Berk et al. 2005), above 90% for $16.0 \leq i \leq 19.0$ (Richards et al. 2006). At higher redshift, both completeness and efficiency drop, with an overall completeness of $\sim 80\%$ for $3 \leq z \leq 5.3$ and efficiency $\sim 55\%$ for QSOs with $z > 3$.

In spite of this remarkable success, the development and refinement of new techniques for targeting quasar candidates are still of great interest. In fact, using different selection strategies it is possible to discover quasars that are missed by the SDSS selection, to reduce and better quantify the selection biases. In this context, the selection of Radio-Loud quasars is particularly interesting to check the completeness of samples selected on the basis of optical data only, since the use of radio-data reduces the biases due to reddening and/or dust obscuration. Also, increasing their population (representing $\sim 10\%$ of the whole population of quasars) and studying the evolution of their demography with more complete and unbiased samples, it is possible to gain some insight into the relationship between Radio-Loud and Radio-Quiet phenomena.

In this chapter we present the results of our original selection of RL QSOs with $3.6 \leq z \leq 4.4$ based on a Neural Network algorithm. We selected the initial sample by a cross match between FIRST radio sources (from the 2003 April 11 release of the catalogue), with star-like objects in the SDSS DR7 photometric catalogue. After a section dedicated to the surveys used in this work (section 2.2), we will discuss (section 2.3) the pre-selection criteria applied to the initial sample. In section 2.4 we will give a general overview on learning machine and artificial

neural networks (ANN), we will discuss in particular the algorithm of the ANN used in this work. Finally we will discuss the further steps needed to prepare the data for the application of the ANN, the training and the test of ANN and finally we will present the list of candidates quasars (section 2.5) extracted by our ANN for follow up spectroscopy.

2.2 Data

For the selection of RL QSO candidates in the redshift range $3.6 \leq z \leq 4.4$ we used the data of the FIRST radio survey (section 2.2.1) and the SDSS DR7 (section 2.2.2) optical survey. In order to calculate the quasar luminosity function we need to know the overlap area of the two surveys used, that will be discussed in section 2.2.3. Our selection is based on the use of spectroscopically confirmed QSOs, and we extensively used the SDSS QSOs catalogue (described in section 2.2.4) as reference.

2.2.1 FIRST

The Faint Images of the Radio Sky at Twenty-centimeters (FIRST) survey use the NRAO Very Large Array (VLA) in its B-configuration at 1.4 GHz and was originally designed to produce the radio equivalent of the optical Palomar Observatory Sky Survey. Pilot observations for the FIRST survey began in 1993 and the final observations were completed in the Spring of 2011. Over the period of 18 years in which the data were collected, the original plan was improved and the survey area was chosen to make it ideal for comparison with the Sloan Digital Sky Survey (SDSS), so that the final footprint of $10,575 \text{ deg}^2$ is largely coincident with SDSS.

The survey was designed to provide high-resolution maps of the radio sky and it produces images with $1''.8$ pixels, a resolution of $5''$, a typical rms of 0.15 mJy and a flux-density limit of 1 mJy. The positional accuracy of the sources at the survey flux limit is $\sim 1''$, implying high reliable cross matches with other surveys and a low rate of chance coincidences. The FIRST catalogue includes a sidelobe warning flag that indicates a likelihood that the source entry is actually a sidelobe of a bright nearby source rather than a real object (White et al. 1997).

The basic attributes of the FIRST project are described in detail in Becker et al. (1995) and in White et al. (1997), that are the primary references when making use of FIRST results. On the other hand, the recent article of Helfand et al. (2015), is an interesting overview of the survey history, discussing hardware and software changes that affect the catalogue reliability and completeness, and the major science results of the survey.

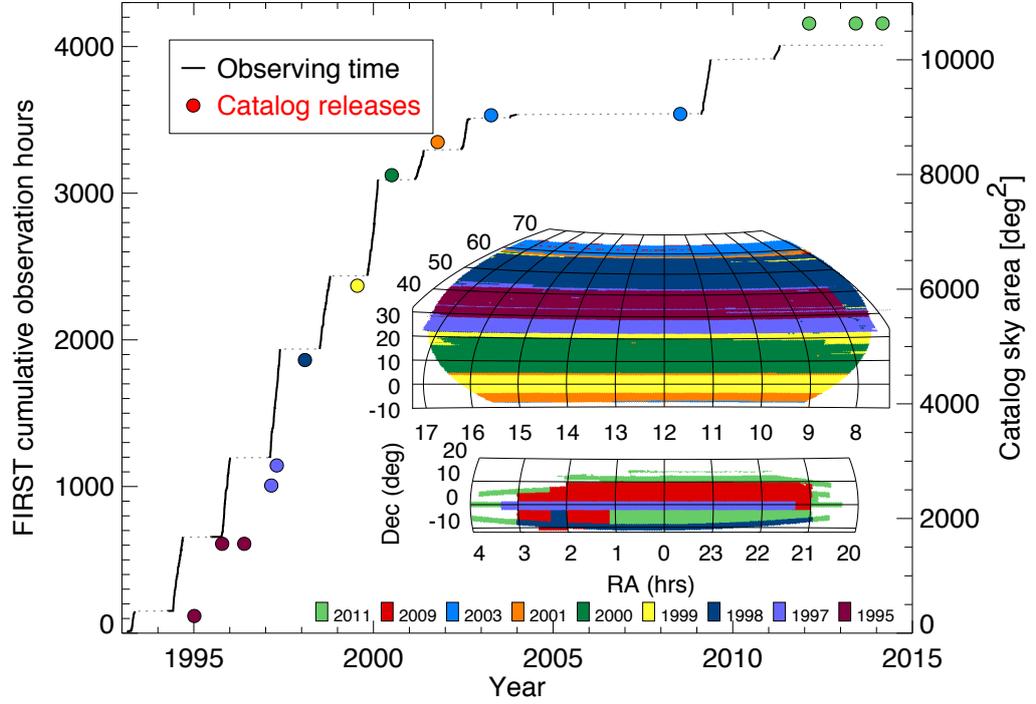


Figure 2.1: *The twenty-year history of the FIRST survey showing the sky coverage (nearly vertical dark lines) as a function of time. The right axis shows the sky area in square degrees and the left axis shows the cumulative observing time. The insert shows the sky coverage color-coded by the catalogue release dates, which are also indicated by circles colored to match the coverage map. The final project, representing an investment of 4009 hours of VLA time, generated a survey covering 10,575 deg² of sky. From Helfand et al. (2015)*

In our work we used the 2003 April 11 version of the FIRST catalogue containing 811,117 sources covering a total area of 9,033 deg² (8,422 deg² in the Northern Galactic Cap and 611 deg² in the Southern Galactic Cap). In Fig. 2.1 it is reported the sky coverage of the FIRST survey over the twenty-year of its history.

2.2.2 SDSS

The Sloan Digital Sky Survey (SDSS) is one of the most ambitious and influential surveys ever attempted. It maps one quarter of the sky, presents imaging information for some 460 million stars and galaxies, spectra for approximately 2 million objects using a dedicated wide-field 2.5 m telescope at Apache Point Observatory in Southern New Mexico (Gunn et al. 2006). There have been a number of data

releases ranging from the Early Data Release (EDR; Stoughton et al. 2002), to the twelve Data Release (DR12; Alam et al. 2015). Each successive release has included a larger area of sky and catalogued a larger number of objects, in both photometry and spectroscopy.

The imaging is carried out in five broad optical bands u, g, r, i, z , see Fig. 2.2), covering the wavelength range from 3000 to 10000 Å (Fukugita et al. 1996; Gunn et al. 1998). The SDSS camera works by scanning across a continuous strip of the sky, resulting in six image strips from the six CCDs in the focal plane. Once the entire sky is imaged in a collection of overlapping 12 strips, the complex set of SDSS photometric pipelines (Lupton et al. 2001) processes the images and extracts the relevant data. The vast amount of data produced by the pipelines is stored in a convenient form in the Catalogue Archive Server (CAS), with all the imaging results provided in SQL database format. As well as many tables with other useful information, the PhotoObjAll table stores most of the useful photometric outputs such as object type (star or galaxy), position, magnitudes, and sizes. Importantly, there is also a variety of flags indicating where the pipeline has encountered a problem and the photometry is not to be trusted. There are a series of magnitudes stored for each object in the catalogue. Model magnitudes are fitted with either a pure de Vaucouleurs profile, a pure exponential profile or a Point Spread Function model (PSF) to each object and the best fit is retained depending on whether the object is an elliptical galaxy, a spiral galaxy or a star. Galactic extinction corrections are also provided as a function of the object position (Schlegel et al. 1998).

From the resulting catalogues of objects, quasar candidates (Richards et al. 2002) are selected for spectroscopic follow-up. Spectroscopy is performed using a pair of double spectrographs with coverage from 3800 to 9200 Å and a resolution $\lambda/\Delta\lambda$ of roughly 2000. The SDSS quasar survey spectroscopically targets quasars with $i < 19.1$ at low redshift ($z < 3$) and $i < 20.2$ at high redshift ($z \geq 3$). The low-redshift selection is performed in $ugri$ color space, and the high-redshift selection is performed in $griz$ color space. In addition to the optical selection, a SDSS object is also considered to be a primary quasar candidate if it is an optical point source located within 2.0'' of a FIRST radio source (and $i < 19.1$).

In this work we used the SDSS Data Release 7 (Abazajian et al. 2009), that covers a total imaging area of 11,663 deg² (7,646 deg² in the Northern Galactic Cap). A total of 357 million distinct objects are included in the imaging catalogue, of which approximately 1.6 million are also included in the spectroscopic catalogue. The survey reaches magnitude limits (95% detection repeatability for point sources) in photometric bands u, g, r, i and z of 22.0, 22.2, 22.2, 21.3 and 20.5 respectively. Absolute astrometric errors are $< 0''.1$. In this work, we consider only the images flagged by SDSS as 'Primary'. These are unique detections, i.e. they do not include duplicate detections from the overlap between survey

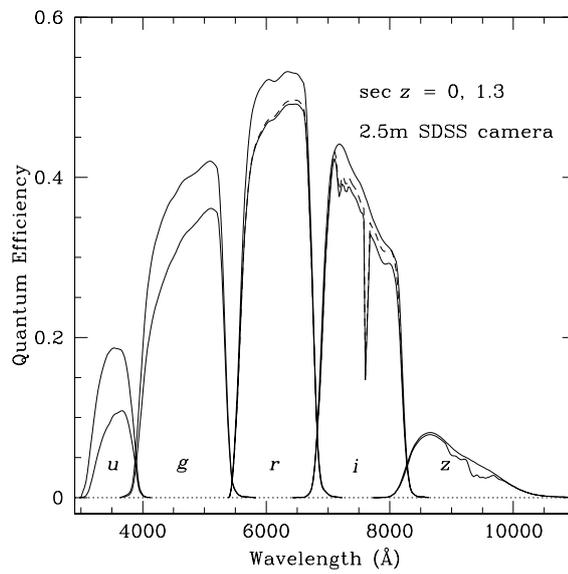


Figure 2.2: *SDSS filter responses, in u, g, r, i, and z. The upper curve in each case is the filter response including the quantum efficiency of the CCD and the reflectivity of the primary and secondary, ignoring the atmosphere, and the lower curve assumes an air mass of 1.3. Scattering within the thin chips affects the r and i bands; this has no effect on extended objects, and the corresponding response curve is given by the dashed line in these two cases. From Stoughton et al. (2002)*

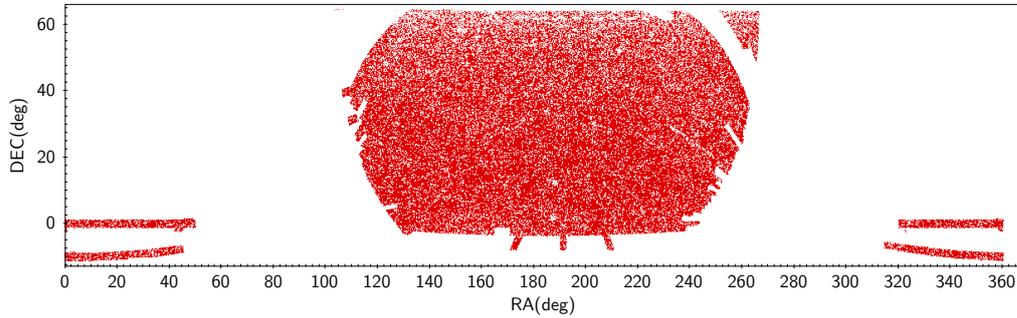


Figure 2.3: Sources detected in both *FIRST* (April 2003 version) and *DR7 SDSS* surveys, as used to calculate the overlap area of the two surveys (section 2.2.3)

stripes. Each such object is associated with a run and a field which is the primary source of imaging data at this position. The remaining objects can be "secondary" or "family" (see paragraph 4.7 in Stoughton et al. 2002 for a definition of these terms).

2.2.3 FIRST-SDSS overlap area

To determine the overlap area of the two surveys we first determined the area of the *FIRST* survey, which has an irregular boundary, by constructing a Delaunay triangulation using the source coordinates. We used code developed by Bernal (1988) which provides the coordinates of the vertices of the unique set of triangles over the *FIRST* area. We then computed the area covered by *FIRST* by adding up the areas of the individual triangles. The resulting area is 9,032.27 square degrees. Finally, for each *FIRST* source, we queried the *SDSS* database to see if the position of the sources was included in the survey. The result of the query was that 89.38% of *FIRST* sources fall in the *SDSS-DR7* imaging area, implying an overlap of 8,073.04 square degrees. In Fig. 2.3 we show the overlap area of *FIRST* 2003 and *SDSS DR7*.

2.2.4 DR7 Quasar catalogue

The fifth edition of the Sloan Digital Sky Survey (*SDSS*) Quasar catalogue, which is based upon the *SDSS* Seventh Data Release, contains 105,783 spectroscopically confirmed quasars. It represents the conclusion of the *SDSS-I* and *SDSS-II* quasar survey and consists of quasars that contain at least one broad emission line ($\text{FWHM} > 1000 \text{ km s}^{-1}$) or unambiguously broad absorption line (*BAL*) quasars and it includes highly reliable determinations of the QSOs redshifts. The catalogue is fully described in Schneider et al. (2010), and includes quasars selected from

the SDSS photometric catalogue either for their colors or for their positional coincidence with radio sources in FIRST catalogue (within $2''$) or ROSAT X-ray sources (within $10\text{-}20''$) and absolute optical magnitude $M_i \leq -22.0$ (in a cosmology with $H_0 = 70\text{kms}^{-1}\text{Mpc}^{-1}$, $\Omega_M = 0.3$, and $\Omega_\lambda = 0.7$). The sample also includes some supplementary quasars that meet the above criteria but were selected initially as galaxy targets. Note that the magnitude limits for various candidates are different. The catalogue was created by inspecting all spectra that were either targeted as quasar candidates, or classified as a quasar by the spectroscopic pipelines. The catalogue covers an area of $\sim 9380\text{ deg}^2$. The quasar redshifts range from 0.065 to 5.46, with a median value of 1.49. The number of the quasars (radio-loud + radio-quiet) included in the catalogue having redshift > 3.6 is 3067, i.e. $\sim 3\%$.

2.3 Pre - selection criteria

We used the data from FIRST and SDSS to create a starting sample of radio-optical sources among which we aimed to select the list of our quasar candidates.

In order to do that, we matched each FIRST source, not flagged as possible sidelobe or nearby bright source ($\sim 3.6\%$ of the sources in the catalogue have this warning flag), with the closest optical object in the 'PhotoPrimary' view of the SDSS DR7 catalogue within a $1''.5$ radius. This radius is the same as used by Carballo et al. (2008) (here and after C08), and is a compromise between completeness and efficiency. The adopted value is lower than the $2''.0$ radius used by SDSS in their algorithm for QSO selection. However, the excellent astrometry of FIRST and SDSS means that the peak in the distribution of optical/radio offsets occurs at about $0''.2$ (Schneider et al. 2010, Fig. 6), supporting our adoption of a $1''.5$ radius. From this match we obtained a starting sample of 222,517 sources. In this sample there is no selection by radio flux density or radio morphology other than the requirement that the radio source have at least a weak core component. The FIRST catalogue itself introduces several minor selection effects: the FIRST sensitivity limit is somewhat non-uniform over the sky, with small variations due to the observing strategy and large variations due to decreasing sensitivity in the vicinity of bright sources. However, the fraction of the survey area affected by sensitivity variations is small, less than 15% (Becker et al. 1995). Another effect is that the FIRST survey limit of ≥ 1.0 mJy refers to the peak flux density of sources rather than to the integrated flux density; consequently, extended sources with total fluxes greater than 1 mJy may not appear in the catalogue because their peaks fall below the detection threshold.

From these 222,517 matches, we first selected the 13,956 star-like objects with $15.0 \leq r \leq 20.2$, where r refers to SDSS PSF r magnitude, corrected for Galactic

Table 2.1: *Steps in the pre-selection process*

Pre-selection criterion	Selected sources
Starting FIRST-SDSS sample ($\leq 1''.5$)	222,517
$15 \leq r \leq 20.2$	74,853
Point-like	13,956
Without mag error > 0.2 in all bands	13,934
Exclude sources with 'fatal' error flags	13,287
Exclude sources with 'CHILD' flag	9,139

extinction according to Schlegel et al. (1998).

We then filtered the sample on the basis of several SDSS quality-control parameters used by others (e.g. Richards et al. 2002) when selecting QSO targets for spectroscopy. Specifically, we rejected all objects with magnitude errors > 0.2 mag in all five bands, and any for which the SDSS 'fatal' error flags 'BRIGHT', 'SATURATED', 'EDGE' or 'BLENDED' were set, indicating unreliable photometry. This left 13,287 objects.

Finally, in contrast with Richards et al. (2002), we rejected all objects with the 'CHILD' flag set (another 4,148 sources), indicating objects obtained by deblending an image flagged 'BLENDED'. This criterion ensures that only one optical object is associated with each radio source, and we adopt it for consistency with C08. In this way we avoid introducing differences in the pre-selection that may change the final efficiency of the neural-network algorithm. This is the main source of incompleteness in our sample, as will be discussed in Chapter 3.

This pre-selection process, summarized in Table 2.1, left us with 9,139 star-like objects coinciding with FIRST radio sources.

2.4 Data mining and Learning Machines

The selection of quasar candidates from an astronomical survey is a well-known and important problem in astronomy, and one of the best suited for a data mining approach.

Data mining not only renders dealing with a sheer volume of data simpler and more automatic, but it also improves the efficiency and completeness of the candidates targeting. In fact, many of these selections are based on machine learning algorithms that have the ability to combine, in complex models, the full set of known characteristic of the astronomical objects, allowing a more refined selection of quasar-candidates in multi-dimensional space in comparison to the old techniques based on simple color-color diagrams.

White (2000), in a review of the methods of object-classification for astronomy

based on multi-parameter data set, proves that *decision tree* methods improve in average the reliability of the quasar selection to 85% compared to only 60% for simpler color-color criteria.

The techniques that have been used in literature to target quasars or to classify astronomical objects are many: decision trees (Ball et al. 2006), principal component analysis (Yip et al. 2004), support vector machines (Zhang & Zhao 2003), kernel density estimators (Richards et al. 2009). Unfortunately, there is no a simple method to select the optimal algorithm to use, because the most appropriate algorithm can depend not only on the dataset, but also on the application for which it will be employed. In this section we will describe the Artificial Neural Network algorithm that we developed and used to extract a sample of quasars candidates from the sub-sample of objects without spectroscopy that we pre-selected in section 2.3. In section 2.4.1 we will describe in general what is a neural network and how it works. Afterwords, in section 2.4.2, we will describe in particular our ANN algorithm.

2.4.1 Artificial Neural Networks

Machine-learning algorithms are broadly divided in supervised and unsupervised methods, also known as predictive and descriptive, respectively. Supervised methods make use of objects for which the required identification, for example a classification (galaxy, star, quasar, etc.), is known with confidence. The machine is trained on this set of objects, and the result of this learning is applied to further objects for which the identification is not available. Typically in astronomy, the target property is spectroscopic, and the input attributes are photometric. In contrast to supervised methods, unsupervised methods do not require a training set. This is an advantage because the data can speak for themselves without preconceptions such as expected classes. On the other hand, if there is prior information, it is not necessarily incorporated.

Artificial neural networks (ANN) are a powerful kind of machine learning algorithm and their conception was inspired by the attempt of simulating biological neural systems. In fact, the human brain consists primary of nerve cells called *neurons*, linked together with other neurons via strands of fibers called *axons*, the contact point between a neuron and the axon of other neutrons is called *synapse*. Neurologists have discovered that the human brain learns by changing the strength of the synaptic connection between neurons upon repeated stimulation by the same impulse.

Analogous to the human brain structure, ANN is composed of an interconnected assembly of nodes with weighted connections. As shown schematically in Fig. 2.4, a multilayer ANN consists of an "input layer" where the *i*-attributes (or separation variables) of the objects represent the input nodes. Each input node

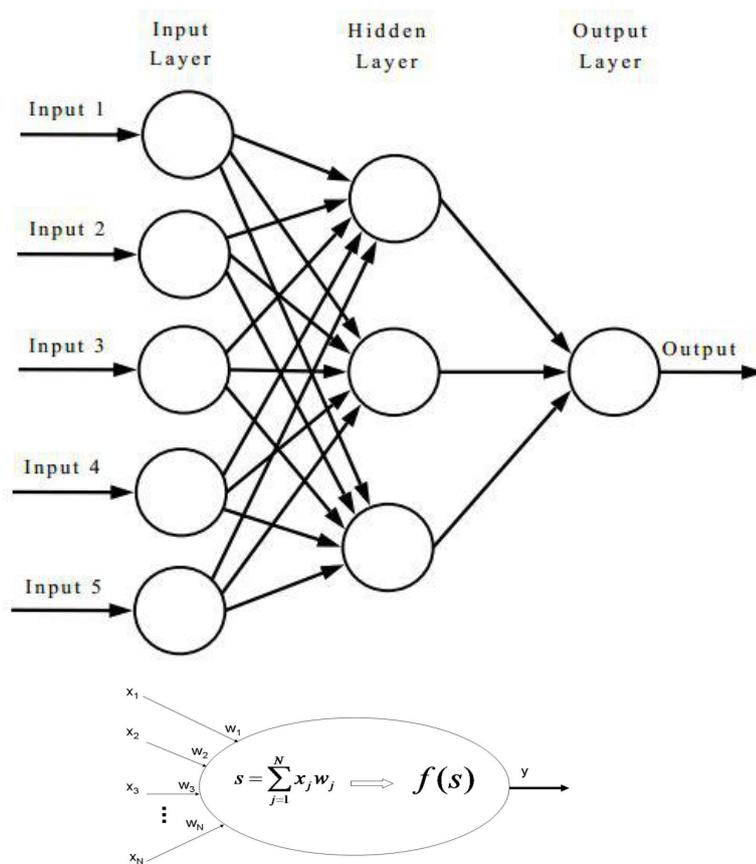


Figure 2.4: Schema of a multilayer feed-forward artificial neural network for an object with n attributes, a hidden layer of size p , and a single continuously-valued output. In the bottom how works each single neuron

is connected via a weighted link, $w_{i,j}$, to the nodes (called *neurons*) of the next *hidden layer*. In a *feed-forward* ANN the nodes in one layer are connected only to the nodes in the next layer, in a *recurrent* ANN the links may connect nodes within the same layer, or from one layer to the previous one. A neuron in the hidden layer forms a weighted sum of its inputs and of an optional bias b , then it passes this information through a *nonlinear transfer function* f , also called *activation function* that it is typically a step or a sigmoid function. The output from the j^{th} hidden node is therefore expressed by:

$$p_j = f\left(\sum_i (w_{i,j}x_i) + b\right) \quad (2.1)$$

These values pass to a possible second hidden layer that performs a similar process. The number of hidden layers, like other parameters of the ANN, has to be optimized for the problem that the ANN aims to solve. The weighted information passes through the layers of the ANN until the last 'output' layer, which performs a sum of its inputs, giving the output \vec{y} . The use of non-linear functions f in neurons of the hidden layer gives the neural network greater computational flexibility than a standard linear regression model. Many nonlinear functions are possible and the choice depends on the nature of the particular problem.

As in the biological system, training a ANN amounts to adapt the weights, \vec{w} , of the links until they fit the input-output relationships of the underlying data. In a *supervised* ANN this is done by training the network on a set of input vectors for which the ideal outputs (or *targets*) are already known. The training uses an algorithm to minimize a cost function c , that it is commonly of the form of the mean-squared deviation between the actual and desired output:

$$c = \frac{1}{N} \sum_l \beta_l (y_l^{\{n\}} - T_l^{\{n\}})^2 \quad (2.2)$$

where N is the number of inputs, T_l is the target output value for the l^{th} output node. The β_l terms are optionals and eventually allows us to assign different weights to different outputs, and thereby give more priority to determining certain outputs correctly.

The biggest advantage of neural networks is that they are general, they can handle problems with many parameters and that they can work well even when the distribution of objects in the N-dimensional parameter space is very complex. The major disadvantage of ANN is that they are slow especially in the training phase. Another disadvantage is that it is very difficult to determine the details on how the net is making its decision.

2.4.2 Our Neural Network

We used a supervised Artificial Neural Network algorithm of feed-forward type, suitable for solving classification problems, and programmed using the *Matlab Neural Network Toolbox software*®. A supervised ANN is trained with samples of known classification, in order to learn how to distinguish between the classes. Only after the training has been carried out, can the trained ANN be used to classify a new problem sample. In our case the classification task was formulated as a binary problem, the two classes being: (a) the target class, i.e. QSOs in the redshift range $3.6 \leq z \leq 4.4$, and (b) the 'non-target' class, i.e. all other types of object. Our ANN was composed of the input layer, just one hidden layer, and an output layer y . The output y for the i th object, with values in the range $(0, 1)$, is given by the non-linear function:

$$y^i = \frac{1}{1 + e^{-a^i}} \quad (2.3)$$

with $a^i = w_0 + \sum_{j=1}^d w_j x_j^i$, where $(x_1, x_2, \dots, x_d)^i$ are the input variables for object i . w_0 and (w_1, w_2, \dots, w_d) , called bias and weights respectively, are the parameters fitted during the training. This ANN model is known as *logistic linear discriminant*. The adopted error function was the variance of the outputs:

$$e = \frac{1}{m} \sum_{i=1}^m (y^i - T^i)^2 \quad (2.4)$$

where m is the number of objects used for the training and T is the target value, set to 1 for the class of high-redshift QSOs and 0 for the remaining sources, during the training. The optimal parameters for the net, i.e. those minimising the error, were obtained using the Levenberg-Marquardt algorithm. This is a simple but robust function for optimisation and it appears to be the fastest for the training of moderate-sized ANNs (Hagan & Menhaj 1994).

The set of input variables adopted for the ANN is the best set obtained in Carballo et al. (2006), and it is a combination of optical magnitudes and colours $(r, u - g, g - r, r - i, i - z)$ and radio-optical separation. The input variables were pre-processed, normalizing their values to the range $(-1, 1)$. No outliers required trimming. For this step the whole pre-selected sample was used, regardless of whether the source was spectroscopically classified and thus suitable as a training object, or not. In fact, the input variables of the new objects (i.e. the problem objects) presented to the trained net are expected to be normalized in the same way as the ones used in the training process.

2.5 QSOs candidates

When using a machine learning for classification, an important and often critical step is constituted by the so-called *pre-processing* of the data. Data preprocessing in general consist in analyzing the data and in the transformation of the set of attributes to another set of attributes in order to enable the particular data mining or machine learning method to achieve its results. In our case this step involve the construction of the training subsample, the preparation of the data and the analysis of ANN training. These steps are discussed in section 2.5.1 and 2.5.2. After those, in section 2.5.3, we will finally present the list of the quasar candidates selected with our methodology.

2.5.1 Spectroscopic classification and pre-processing

As discusses in the previous section, our ANN is of supervised type and therefore it needs to be trained with a sample having well known target property. Therefore we need to extract from the sample pre-selected in section 2.4, the sub-sample having spectroscopical identification, and hence use the their attributes (photometrical and radio-information) to train the ANN to distinguish between quasars having $3.6 \leq z \leq 4.4$ and other sources. Considering that the number of known RL QSOS in this range of redshift it is very small, this step it is particularly critical and we needed to be very careful since the inclusion or the exclusion of a QSOS from the training sample could have severe consequences on the performance of the ANN.

We verified that of the sample of 9,139 sources passing initial pre-selection (Table 2.1), 6,091 have spectra in the SpecObj view of SDSS-DR7. 5,348 of the latter are included in the 5th edition of the SDSS Quasar Catalogue (DR7 QSO Catalogue; Schneider et al. 2010), which is also based on SDSS-DR7, but uses more stringent criteria for the classification of the objects as QSOS, in order to exclude dubious cases. 71 of the QSOS in this catalogue have redshifts in the range $3.6 \leq z \leq 4.4$.

The remaining 743 sources with spectra in DR7-SpecObj but not included in the DR7 QSO Catalogue are classified by SDSS as stars, galaxies, quasars and sources of 'unknown' type. For all of these sources, a search was made in NED (the NASA/IPAC Extragalactic Database), and none of them was classified there as a $z \geq 3.6$ QSO. We also visually inspected the DR7-SpecObj spectra of these sources to check if any of them could be a QSO in the redshift range of interest here, but none of the objects, which include those with 'unknown' spectra, had spectral features consistent with a high- z QSO. We notice that many quasars classified in SDSS DR7 as quasars at $z \geq 3$ are indeed quasars at lower redshift (for confusion of the Mg II with the $Ly\alpha$ emission line), but all these

misclassifications (but one classified as having $z = 3.3$, $J153148.92 + 114220.3$) were corrected in the 5th Quasar catalogue, that reports the correct redshift.

3,048 of the 9139 sources in the sample lack spectroscopic classification in DR7-SpecObj. However, the DR7 QSO catalogue was compiled by inspecting all the SDSS spectra, not just those of the quasar candidates, and identified 115 of these as quasars (one of them being a high- z QSO in the redshift range of interest for our work) despite them not being automatically identified as such. Another 17 sources were identified as QSOs by C08, comprising 8 high- z QSOs and 9 QSOs with redshift below 3.6. The remaining 2,916 sources were checked in NED, and none of them had been spectroscopically classified as of March 2012.

A total of 6,223 ($= 6091 + 115 + 17$) sources thus have a reliable spectroscopic classification. Of these, 80 ($= 71 + 1 + 8$) are QSOs in the redshift range $3.6 \leq z \leq 4.4$. These sources form the training sample, i.e. the sample used to train the neural network to distinguish high- z QSOs in our redshift range of interest from other objects. Quasars with $3.6 \leq z \leq 4.4$ play the role of the target class. The non-target sources include stars, galaxies, QSOs with other redshifts and objects with spectra classed as 'unknown' but lacking the features expected for our target sources.

The 2916 sources without available spectra from SDSS-DR7 or from the literature (as of March 2012), form the sample from which new QSO candidates in our redshift range of interest are selected using the trained ANN.

2.5.2 Training and testing of the ANN

Having defined the training sample, with 6223 spectroscopically classified sources, of which 80 are high-redshift QSOs, we are ready to train the ANN and to test its performance as a classifier.

The classification algorithm fitted by the ANN provides for each source an output $0 \leq y \leq 1$. The extreme values of 1 and 0 correspond respectively to sources with input variables more similar or less similar to those of the high- z QSO class. Objects with measured y greater than some threshold value y_c are candidate high- z QSOs.

The performance of the trained ANN can be expressed in terms of two basic parameters: efficiency (or reliability) and completeness. In our case the efficiency is the fraction of candidate high- z QSOs selected by the ANN which are true high- z QSOs. The completeness is the fraction of true high- z QSOs with $y \geq y_c$, i.e., the fraction of true high- z QSOs selected as such by the ANN.

The performance of the ANN is ideally tested with a sample of objects not used during the training. In our case, since the target sample has only 80 objects, the 'leave one out' method was applied, using all but one of the objects for the training, and the remaining one for the test. In total 6223 ANNs were run, each

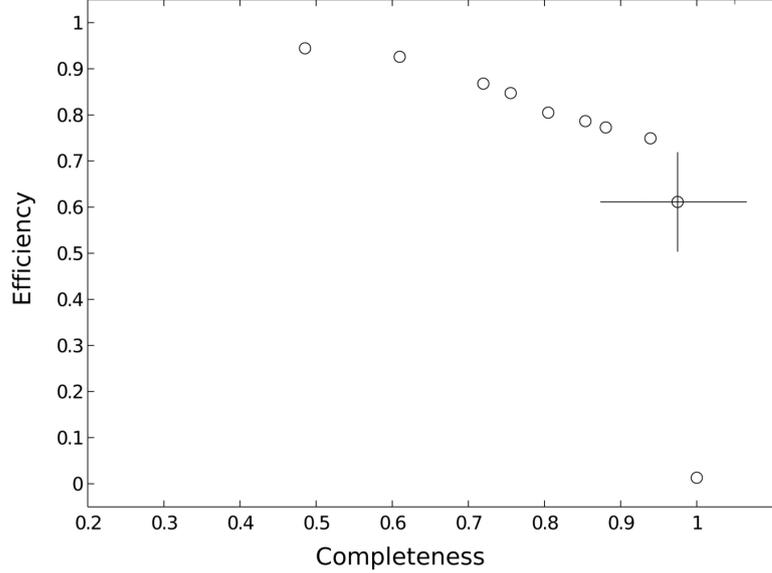


Figure 2.5: *Efficiency versus completeness, measured for the test sample. Each symbol corresponds to a given value of the threshold y_c , ranging from 0.9 to 0, in increments of 0.1, from left to right. The symbol with the error bars correspond to the adopted threshold $y_c = 0.1$.*

of them providing the output for a test object. These output values, for a sample of 6223 test objects, were used to compute efficiency and completeness as a function of y_c , and the results are shown in Fig. 2.5. Since our purpose is to build a sample appropriate for statistical analysis, priority is given to completeness, accepting lower y values at the cost of lower efficiency. Choosing $y_c = 0.1$ our ANN classifier has an efficiency of $60\% \pm 9$ and a completeness of $97\% \pm 11$ (errors assume Poisson statistics).

2.5.3 High-redshift QSO candidates

The set of 6223 trained ANNs used for the testing was applied to the sample of 2916 sources without spectra, in order to find candidate high- z QSOs. For each source we adopted the median of the 6223 output values, and we selected as high-redshift QSO candidates those sources with $y_{med} > 0.1$. In this way 15 QSO candidates in the range $3.6 \leq z \leq 4.4$ were found and are listed in Table 2.2. These candidates come out of an original set of 2,916 sources lacking spectroscopic identification, and therefore we notice that are just the $\sim 0.5\%$ of the whole sample that is significant enough on the sensibility of our network network.

Table 2.2: Sample of 15 FIRST-SDSS DR7 high- z QSO candidates selected by our ANN.

RA (J2000) (1)	DEC (2)	r_{AB} (2)	$S_{1.4\text{GHz}}$ (mJy) (3)	y_{med} (4)
08:15:55.02	+46:53:21.4	19.89	2.97	0.12
08:33:16.91	+29:22:28.0	20.13	12.63	0.32
08:57:24.33	+11:05:49.2	19.81	1.91	0.98
09:09:53.85	+47:49:43.2	19.90	373.29	0.22
09:14:36.23	+50:38:48.5	20.19	47.98	0.15
09:26:40.29	-02:30:41.5	19.82	1.9	0.12
10:29:40.93	+10:04:10.9	19.47	2.81	0.22
10:34:20.43	+41:49:37.5	20.12	2.17	0.30
11:33:00.71	-04:11:58.5	19.96	9.64	0.12
11:51:07.42	+50:15:58.6	20.09	1.69	0.32
12:05:31.73	+29:01:49.2	20.17	1.51	0.50
12:13:29.43	-03:27:25.7	19.64	23.37	0.76
12:28:19.97	+47:40:30.4	19.32	2.24	0.46
12:44:43.07	+06:09:34.6	19.78	1.29	0.21
15:43:36.59	+16:56:21.8	18.97	10.85	0.12

The columns give the following: (1) SDSS J2000 coordinates; (2) SDSS dereddened PSF r magnitude; (3) FIRST peak radio flux density; (4) ANN output.

2.6 Conclusions

In this chapter we used a feed-forward neural network to obtain a sample of 15 RL QSOs candidates in the redshift range of $3.6 \leq z \leq 4.4$ out of a sample of 2,916 radio-optical sources lacking spectroscopic identification. On the basis of the ANN training based on sources with known spectroscopy, the selection has an overall high completeness of 97% and a 60% efficiency. The ANN gives for each candidate an output $0 \leq y_i \leq 1$, listed in column 4 of table 2.2, that gives the degree of similarity to the class of targets. Adopting for the 15 candidates a weighted efficiency based of their value of y_i , we expect that at least 4 of the candidates are actual RL QSOs in the desired range of redshift.

Chapter 3

Data analysis

3.1 Introduction

The ultimate test and purpose of the QSOs selection method discussed in the previous chapter is constituted by the spectroscopic follow-up of the candidates. Once the spectroscopic types and the redshifts of the observed objects are identified, it is possible to confirm the predictive accuracy of the selection and to use the new discoveries for further analysis and interpretation. In this chapter we present the results of the spectroscopic observation of the 15 NN-selected high- z -QSO candidates (in section 3.2) and of 3 further QSOs candidates selected in a previous investigation of our research group (Carballo et al. 2008, hereafter called "C08"). We use the results of these observations to assess efficiency and completeness of our selection. In particular we use SDSS DR9 to check the completeness of our selections. In addition, we use our new data to assess the efficiency of other selections of high- z -QSO: before testing the selection by C08 and finally checking the completeness of the spectroscopic identification of high- z -QSOs in SDSS DR7 (section 3.3).

3.2 Spectroscopic follow-up

At the time of their observations, none of the 15 QSO candidates selected by our NN (see previous chapter) had spectra available in the literature. Therefore long-slit spectra of all 15 were obtained with the 2.5-m Nordic Optical Telescope (NOT) during the night of 25th March 2012, using ALFOSC (Andalucía Faint Object Spectrograph and Camera) with grism #4, which provides a dispersion of 3\AA per pixel. The spectral coverage was $4000 - 9000\text{\AA}$ and the resolution was 15\AA . The exposure time was about 1000s per source, delivering signal-to-noise ratios $\gtrsim 8$ per pixel. The seeing was typically better than $1''.3$ FWHM. A

spectrophotometric standard star was observed to correct for spectral response. After each target spectrum, an exposure of an arc-lamp was taken for wavelength calibration. Data were reduced using standard IRAF¹ routines.

All 15 candidates were confirmed as QSOs, 7 of them (Fig. 3.1) in the desired range of redshift, 5 in the nearby range $3.1 \leq z < 3.6$ (Fig. 3.2), and the remaining 3 at lower redshifts (Fig. 3.3). For the last 3 quasars at lower redshift, the NN-misclassification is caused most likely by confusion of Mg II emission with Ly α emission. Table 3.1 lists the coordinates, r magnitudes, y_{med} values and redshifts for the 15 candidates. A note in column 6 indicates whether there is a spectrum of the object in the latest SDSS release DR9 (see next subsection). We report also the discovery of five new broad absorption line (BAL) QSOs (see Table 3.1), 4 of them within $3.6 \leq z \leq 4.4$, and 1 with $z = 1.4$

The observed efficiency of this selection is therefore $\sim 47\%$ (7 $z > 3.6$ QSOs out of 15 candidates).

During the same night of observation, we also obtained the spectra 3 QSOs candidates listed in Table 3.2 and selected in C08. Those candidates were observed in order to complete the observation program of that research and therefore assess its efficiency.

3.3 Check of completeness and efficiency

The spectroscopic results obtained from the observation of the QSOs candidates have been used to assess the completeness and the efficiency of our selection and to compare it with other methodologies. In section 3.3.1 we discuss the completeness of our selection making use of the new spectroscopic data from SDSS DR9. In section 3.3.2 we assess the efficiency of the selection presented in C08. Finally in section 3.3.3 we discuss the completeness of the SDSS target algorithm in comparison with our methodology .

3.3.1 Check of completeness from SDSS DR9

As discussed in chapter 2, our selection-methodology is applied to a sample of sources detected in both SDSS and FIRST using a match radius of $1''.5$, and it makes use of photometric data in SDSS and of the radio-optical separation. Our pre-selection and selection methods are therefore based on the same variables as the SDSS target-selection algorithms. Therefore we checked if the more recent

¹IRAF is distributed by the National Optical Astronomy Observatories, which is operated by Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

Table 3.1: Sample of 15 FIRST-SDSS DR7 high- z QSO candidates selected by our NN.

RA (J2000) (1)	DEC (2)	r_{AB} (3)	$S_{1.4\text{GHz}}$ (mJy) (4)	y_{med} (5)	redshift (6)	DR9 (7)	C08 (8)	Notes (9)
08:15:55.02	+46:53:21.4	19.89	2.97	0.12	3.20		yes	BAL
08:33:16.91	+29:22:28.0	20.13	12.63	0.32	3.30	yes	yes	
08:57:24.33	+11:05:49.2	19.81	1.91	0.98	3.71			
09:09:53.85	+47:49:43.2	19.90	373.29	0.22	3.64		yes	
09:14:36.23	+50:38:48.5	20.19	47.98	0.15	3.62		yes	
09:26:40.29	-02:30:41.5	19.82	1.9	0.12	3.76	yes		
10:29:40.93	+10:04:10.9	19.47	2.81	0.22	3.40		yes	
10:34:20.43	+41:49:37.5	20.12	2.17	0.30	4.00	yes	yes	
11:33:00.71	-04:11:58.5	19.96	9.64	0.12	3.39			BAL
11:51:07.42	+50:15:58.6	20.09	1.69	0.32	3.40		yes	BAL
12:05:31.73	+29:01:49.2	20.17	1.51	0.50	3.44		yes	BAL
12:13:29.43	-03:27:25.7	19.64	23.37	0.76	3.67		yes	
12:28:19.97	+47:40:30.4	19.32	2.24	0.46	1.40		yes	
12:44:43.07	+06:09:34.6	19.78	1.29	0.21	3.76	yes	yes	
15:43:36.59	+16:56:21.8	18.97	10.85	0.12	1.40			LoBAL

The columns give the following: (1) SDSS J2000 coordinates; (2) SDSS dereddened PSF r magnitude; (3) FIRST peak radio flux density; (4) NN output; (5) QSO redshift determined in this work; (6) indicates if the source has a spectrum in SDSS-DR9; (7) indicates if the source was previously selected as a high- z candidate by C08; (8) BAL - broad-absorption-line QSO, LoBAL - low-ionization broad-absorption-line QSO.

Table 3.2: Spectra of three high- z QSO candidates from C08, not selected by our NN, but observed to complete the C08 sample.

RA (J2000) (1)	DEC (2)	r_{AB} (3)	$S_{1.4\text{GHz}}$ (mJy) (4)	redshift (5)	DR9 (6)
08:48:18.88	+39:38:06.0	20.15	1.28	1.4	
10:58:07.47	+03:30:59.6	19.91	4.18	3.44	yes
12:04:07.84	+48:45:48.0	19.97	3.96	M-star	

The columns give the following: (1) SDSS J2000 coordinates; (2) SDSS dereddened PSF r magnitude; (3) FIRST peak radio flux density; (4) QSO redshift or spectral classification; (5) indicates if the source has a spectrum in SDSS-DR9.

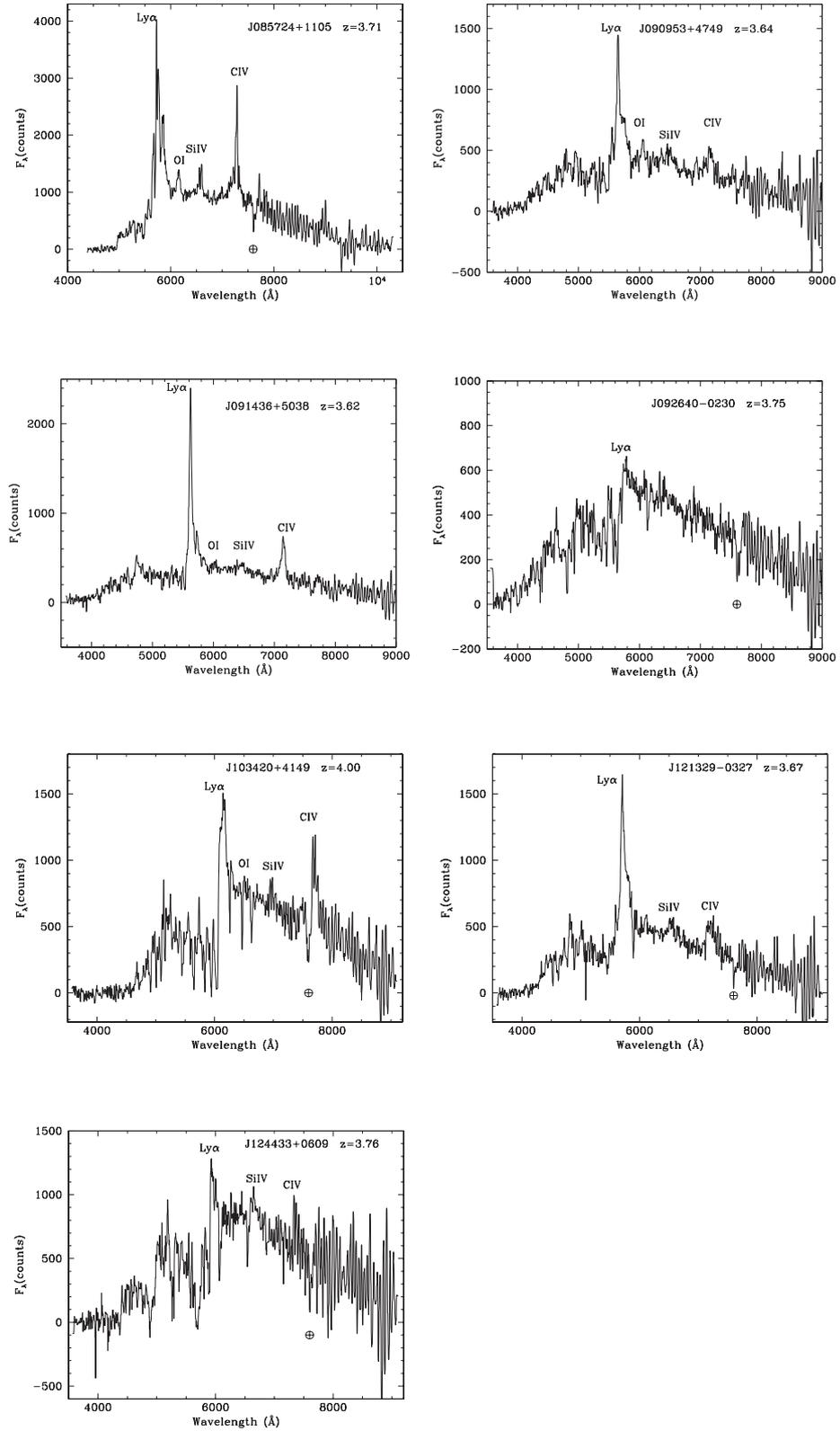


Figure 3.1: NOT/ALFOSC spectra of ANN-selected high- z candidates: QSOs with $3.6 \leq z \leq 4.4$ (section 3.2).

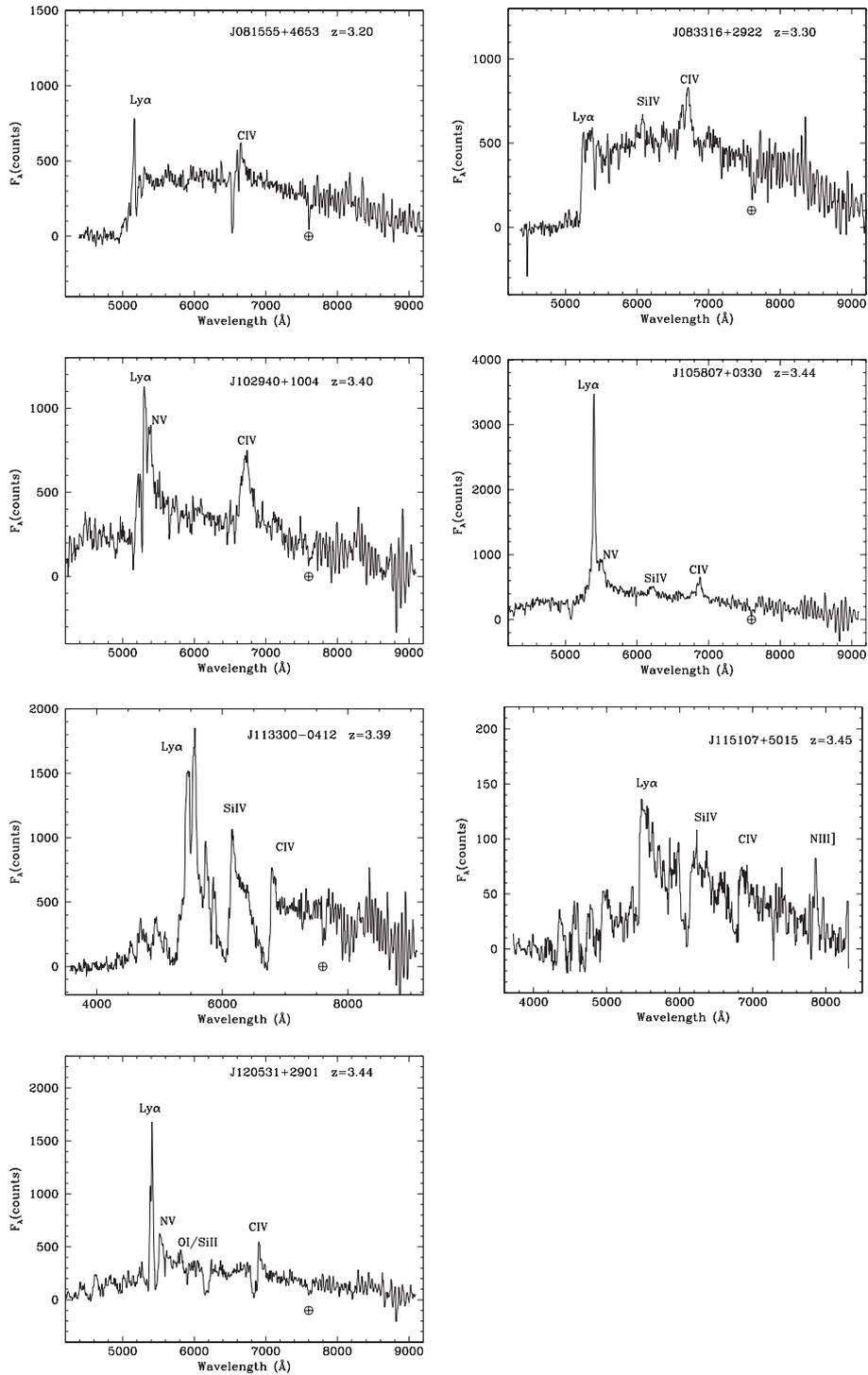


Figure 3.2: NOT/ALFOSC spectra of ANN-selected high- z candidates: QSOs with $3.0 \leq z < 3.6$ (section 3.2).

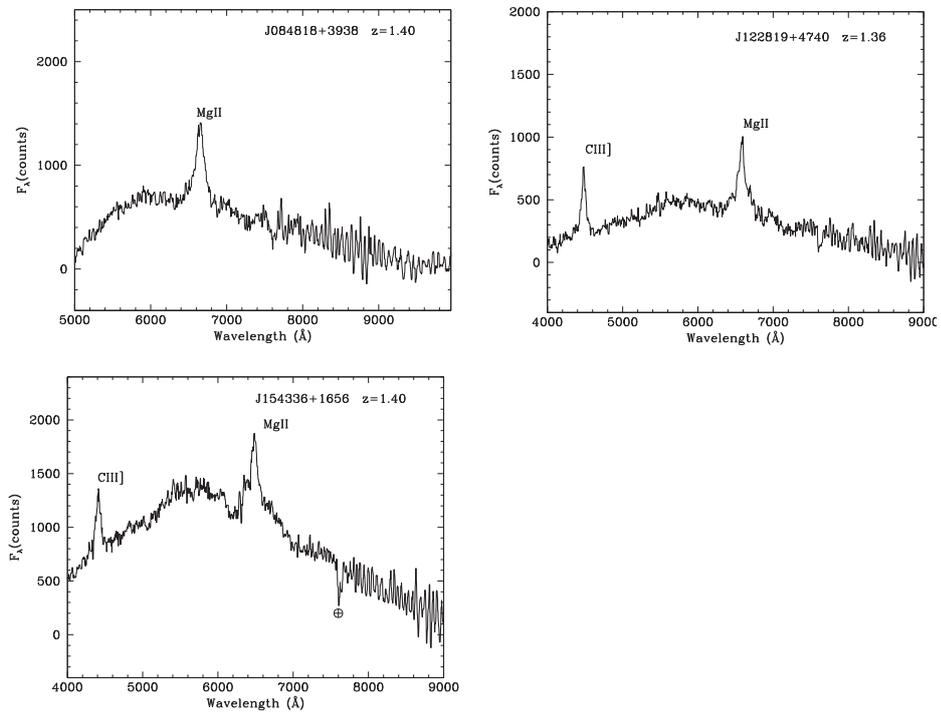


Figure 3.3: NOT/ALFOSC spectra of NN-selected high- z candidates: QSOs with $z < 3$ (section 3.2).

SDSS-DR9 spectroscopic catalogue provides spectra of any of the 2,916 sources lacking a spectral classification after SDSS-DR7 was released.

The 'SpecObj' view of SDSS-DR9 was used for this purpose, giving the following results. 4 of the 15 candidates have spectra in DR9 (see Table 3.1) with redshifts very similar to those reported in this work. Of the remaining 2,901 sources rejected by the NN as high- z QSO candidates, 451 have spectra in DR9 and 4 of them are classified as QSOs with $3.6 \leq z \leq 4.4$. However, examination of the spectra reveals that all four are actually lower-redshift objects or stars. These objects are J075757.87+095607.56, J101403.75+451053.27, J112742.74+363429.5, and J222758.13+003705.45.

The fact that none of these 451 objects are classified in DR9 as high- z QSOs is consistent with the estimated high completeness of our NN algorithm: $\sim 97\%$.

3.3.2 Assessment of the sample of high- z candidates in C08

The selection of high- z QSOs in the present work is based on more recent SDSS data releases than used by C08. However the FIRST-SDSS pairs were obtained in this work in same way as by C08, with similar criteria for the magnitude limits, the maximum radio-optical separation, optical morphology, and photometric quality. Also the classification procedure, aimed at the identification of new high- z QSOs, uses similar NN architecture, input parameter-set, NN training and output parameter.

In C08 we used a training sample comprising 52 QSOs with $z \geq 3.6$, all from the DR5 spectroscopic catalogue, and selected 58 new candidates. In that paper 24 of the candidates were confirmed as $z \geq 3.6$ QSOs (17 from observations obtained for the paper and 7 from the literature or from the DR6 spectroscopic catalogue). 16 sources were classified as other types of object, on the basis of observations by C08, or spectra from the literature or DR6. 18 high- z QSO candidates remained unclassified at C08.

11 of the 18 C08 candidates overlap with the 15 candidate-QSOs identified in this work. As result of our observations, these 11 sources were classified as: 5 QSOs with $3.6 \leq z \leq 4.4$ QSOs, 5 QSOs with $3.2 \leq z \leq 3.5$ and a QSO at $z = 1.40$ (see Table 3.1). The remaining seven candidates of C08 consist of three QSOs included in the DR7 QSO Catalogue and in our training sample (SDSS 110946.44+190257.6 with $z=3.67$, SDSS 123128.22+184714.3 with $z=3.33$ and SDSS 124323.16+235842.2 with $z=3.49$) and four sources not selected by our classifier as being high- z QSOs. The spectra of three of these four sources were obtained in our observing programme at the NOT, yielding a QSOs at $z = 3.44$ and $z = 1.4$ and a late-type star (see Table 3.2, and Fig. 3.2 and 3.3 for spectra) .

The spectroscopic observations of the 58 high- z candidates at C08 are now almost complete (57 out of 58). They yield an efficiency of 52 ± 9 per cent

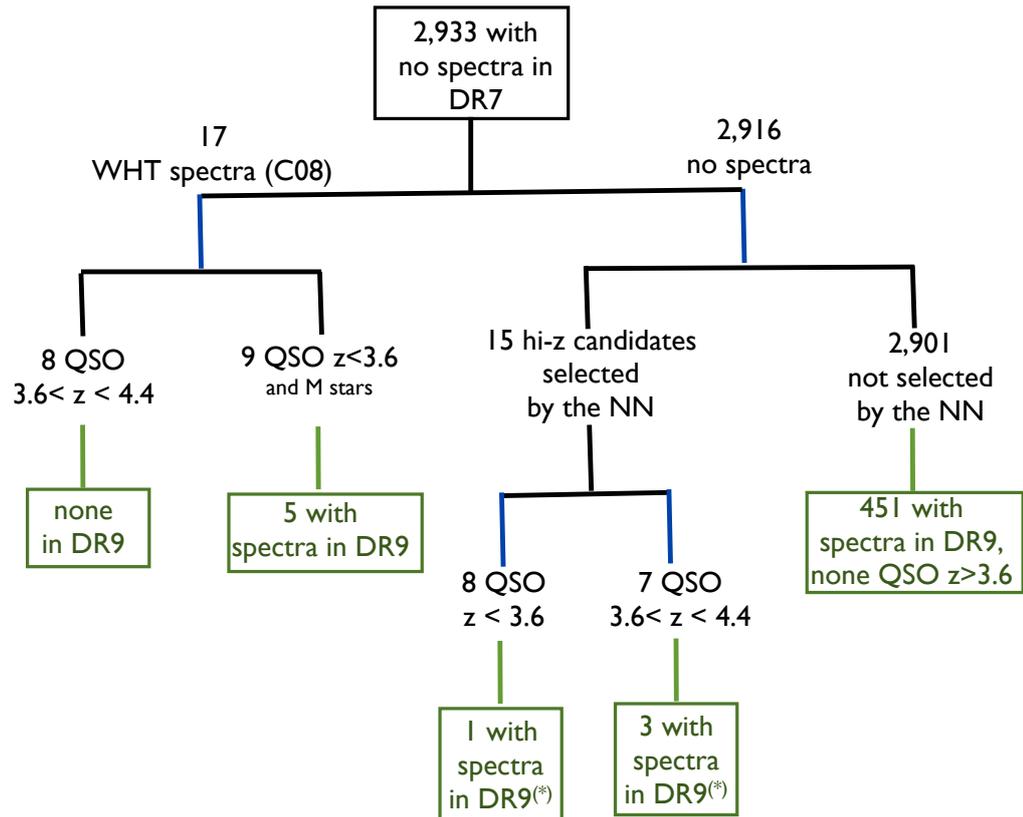


Figure 3.4:

Schematic overview of a check of the efficiency of our selection (see chapter 2) via comparison with DR9 spectroscopic information. The 2916 objects without spectra in DR7, or in the literature, were classified by our NN, with 15 being selected by the NN as high-redshift candidates. () = targets from BOSS (Baryon Oscillations Spectroscopic Survey).*

(30/58), highlighting the value of simple neural networks for this classification task. This efficiency is in reasonable agreement with the value obtained with the training sample, i.e. with the expectation we had from the objects with known classification, 62 ± 9 per cent (C08). In addition, we note that a large fraction of the contaminants, 15 out of 28, are QSOs with $3.1 \leq z < 3.6$, close to the redshift threshold we adopted.

3.3.3 Spectroscopic completeness of SDSS for high- z QSOs

Several studies of the SDSS selection of QSOs (Richards et al. 2002; Croom et al. 2004; Richards et al. 2006; McGreer et al. 2009) suggest an overall completeness above 90%. The completeness in general decreases with increasing redshift and decreasing brightness, and it is particularly inefficient for $2.2 < z < 3$ where quasar and star colours are very similar. In particular, McGreer et al. (2009) studied the completeness of quasar selection at redshift $z > 3.5$ and magnitude $i < 20.2$, and found for the SDSS target algorithm a completeness of $\approx 86\%$, in good agreement with the 85% derived in Richards et al. (2006).

The analysis in this paper and in C08 identifies 15 QSOs with $3.6 \leq z \leq 4.4$, missed by SDSS-DR7 (two have spectra in SDSS-DR9, which uses BOSS). This allows us to estimate the incompleteness of the SDSS-DR7 selection, taking into account the fact that the spectroscopic area of the SDSS survey is $\sim 95\%$ of the imaging area. 4 of our 15 QSOs $3.6 \leq z \leq 4.4$ lie outside the SDSS spectroscopic plates, so the estimated incompleteness in SDSS is 11 QSOs out of 83 (87 minus 4) in this redshift range, i.e. $\sim 13\% \pm 4$, in good agreement with estimates by McGreer et al. (2009) and Richards et al. (2002).

In Fig. 3.5 we plot the $g - i, r - z$ colour-colour diagram of the entire training sample together with the 15 $3.6 \leq z \leq 4.4$ QSOs missed by SDSS and identified by us. Splitting the sample between QSOs with $3.6 \leq z \leq 4.4$ and all other sources, we note that our QSO-selection method is sensitive even at the boundary of the two samples, demonstrating the effectiveness of learning-machine techniques, when compared with simple colour-cut criteria.

3.4 Conclusions

In this chapter, once obtained the spectra of the 15 candidates, we were able to compare the theoretical efficiency and completeness of our NN-selection (based on the training sample) with the actual results on objects without previous spectra. In this way we confirm the predicted results and we show that a simple neural network can be used to select high-redshift QSOs from radio-optical surveys, with 97% completeness and 60% efficiency. We compare our methodology with the

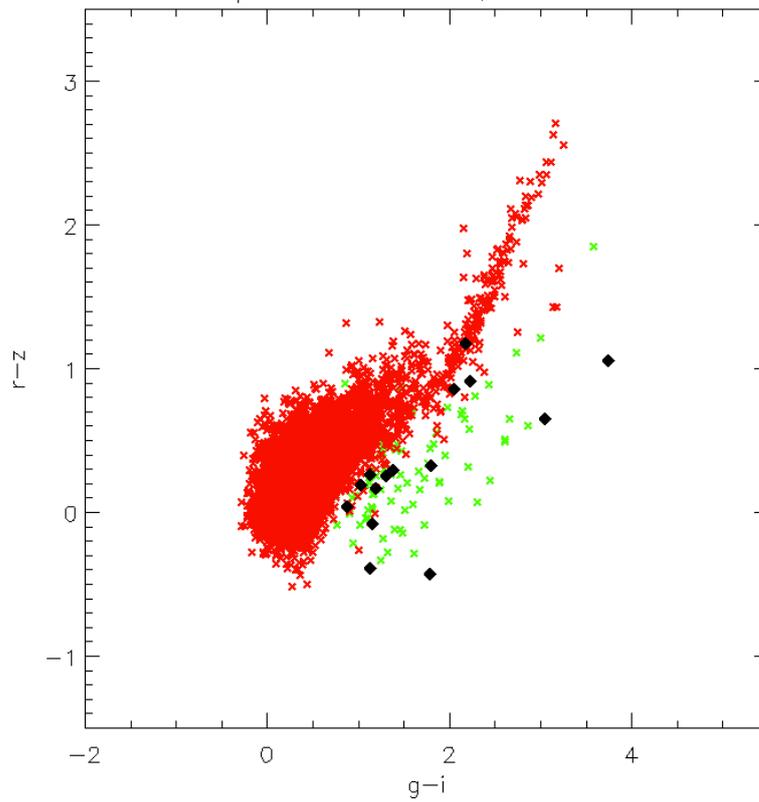


Figure 3.5: Colour-colour diagram of the 6233 sources with spectra in DR7, with green crosses representing QSOs with $3.6 \leq z \leq 4.4$, and red crosses representing other objects. The new 15 QSOs with $3.6 \leq z \leq 4.4$ (8 from C08, 7 from this work, all missed by SDSS) are plotted as black diamonds.

SDSS target algorithm and we obtain that although the efficiencies of the two methodologies are comparable, our neural-network algorithm detects $\sim 97\%$ of the high- z QSOs, while SDSS only detect $\sim 85\%$ of them.

Chapter 4

RL QSOs luminosity function at

$z \sim 4$

[...] aunque sea tamaño como un grano de trigo; que por el hilo se sacará el ovillo¹.

Don Quijote de la Mancha, Miguel de Cervantes

4.1 Introduction

Historically, as the number of known AGNs increased, it became possible to divide them in terms of their physical and cosmological properties, in particular, it became possible to determine the number density of objects in a certain luminosity range and at a given redshift interval, i.e. to calculate their *luminosity function*. The luminosity function and its redshift dependence is highly useful for studying the cosmological evolution of AGNs and for tests of AGN unification models. The accuracy in the determination of the luminosity function depends on the statistics of the considered sample, and therefore it depends on the number of known objects and on the precise determination of its incompleteness.

The sample of radio-loud quasars with $3.6 \leq z \leq 4.4$ and $M_{1450} < -27$ accounts for only 72 known QSOs, but this small sample has been increased to 87 objects with our ANN-selection methods (Carballo et al. 2008; Tuccillo et al. 2015). The completeness of our selection strategy and the identification of new sources offer the possibility to carry out the most accurate determination of the optical luminosity function for these objects known at the time.

¹This sentence, from the Chapter 2 of Don Quijote, has been translated in english as: "by a small sample we may judge of the whole piece"

In this chapter we first give a general overview on the quasar luminosity function and on the models of evolution for the AGN population (section 4.2). In section 4.3 we present the data that we will use for our determination of the luminosity function and we discuss the K-correction. We estimate the completeness of our sample (section 4.4) and we illustrate our new methodology for the calculation of the luminosity function (section 4.6). Finally, we calculate the luminosity function and we discuss (section 4.7) the results of our determination, by comparing it with other results and models from literature.

4.2 Quasar Luminosity Function

Understanding the true distribution of objects in the Universe has always been a basic objective in astronomy. The techniques developed previous to the quasars discovery for determining the distribution of stars in our Galaxy (Trumpler & Weaver 1953) could not be applied directly to quasars, because cosmological equations that relate distance to redshift are much more complex than the euclidian geometry used by galactic astronomers. Therefore it has been necessary to build a discussion of statistical astronomy specific for quasars.

We can say that, in general, the abundance of any class of galactic objects (galaxies, quasars, etc) can be characterized in two different ways. The first is a 2-dimensional information and provides a count of sources as a function of their brightness (apparent flux or magnitude) per solid angle in the sky. This count is comparatively easy to make and it is essentially limited only by the limiting flux in a given waveband and by the angular resolution of the instrument. In this way, if we call $n(F_\nu) \equiv dN/d\Omega dF_\nu$, the number of observed sources per solid angle in the sky between F_ν and $F_\nu + dF_\nu$, then the total intensity of the radiation is calculated as:

$$I_\nu = \int_0^\infty F_\nu n(F_\nu) dF_\nu \quad (4.1)$$

A more detailed description of the abundance of sources is provided by the *luminosity function* $\phi(L, z)dL$ (whose common acronymous is LF) which gives the number of sources per unit comoving volume at redshift z in the luminosity interval $(L, L + dL)$.

The analytic description of the luminosity function is obtained from the observations. For galaxies the most common is the *Schechter luminosity function*, that is a function of a single variable:

$$\phi(L) = \frac{\phi^*}{L^*} \left(\frac{L}{L^*} \right)^{-\alpha} \exp \left(\frac{-L}{L^*} \right) \quad (4.2)$$

For what it concern quasars, the LF is not in general well fitted by the Schechter function, instead, for deep quasar surveys it is usually well fitted by a double-power-law parametrisation that takes into account the redshift and thus, evolution (e.g. Pei 1995; Peterson 1997; Boyle et al. 2000; Croom et al. 2004; Richards et al. 2006):

$$\phi(L, z) = \frac{\phi^*/L^*}{(L/L^*(z))^{-\alpha} + (L/L^*(z))^{-\beta}} \quad (4.3)$$

where α , β , ϕ^* , L^* are the faint-end slope, the bright-end slope, the normalization of the luminosity function, and the characteristic break luminosity, respectively. This model, if $\alpha > \beta$, can be approximated by

$$\phi \propto \begin{cases} L^\alpha & \text{if } L \ll L^* \\ L^\beta & \text{if } L \gg L^* \end{cases} \quad (4.4)$$

The overall shape is therefore a shallow power law at faint luminosities, steepening at luminosities around L^* to a steeper power law at high luminosities.

Unlike the galaxy LF, the quasar LF is strongly dependent in redshift, as already shown in Fig. 1.5. It appears that quasars were far more common at redshift $z \approx 2$ than at the present time. At highest redshift the quasar number density appears to drop quickly, known as the **redshift cut-off** of quasars. This evolution was first parametrized in terms of evolution in ϕ^* only (**pure density evolution** or PDE) or in L^* (**pure luminosity evolution** or PLE). In particular, if the shape of the intrinsic luminosity function of quasars did not change with time but the comoving number density varied with z , then the LF drops before at lower redshift than at higher z ; this is the pure density evolution model. The other scenario corresponds to the assumption that the space density of quasars is constant in time but the luminosity of the population changes with redshift, this is the pure luminosity evolution model. There are some indications that the PLE model fits better the observations (see Fig. 4.1, Croom et al. 2009), suggesting a long lived population of quasars that dimmed over cosmological times. But these results are in contradiction from the ones on the radio-loud QSOs, that although showing the same PLE, they were deduced to be relative young objects from studies of their synchrotron spectra. Nowadays the real underlying causes of this sudden increase and decline in quasar activity are still under debate. Interesting clues are found comparing quasar activity with the galaxy merger rate, but there are no physical basis to expect that either extreme assumption should hold, and probably the PDE and PDL scenarios merely bracket a range of more complicated possibilities.

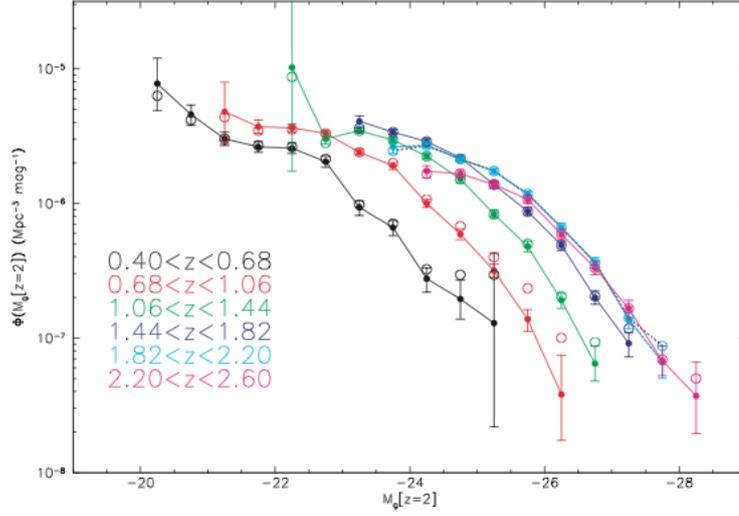


Figure 4.1: *The 2QZ luminosity function. The LF shows the approximate characteristic of Pure Luminosity Evolution (PLE), i.e. a fixed LF shape with an evolving luminosity (Croom et al. 2009).*

4.3 Data

In this section we discuss our derivation of the K-correction (subsection 4.3.1) in order to calculate the absolute magnitudes presented in subsection 4.3.2) that we will use to calculate the optical LF for RL QSOs within $3.6 \leq z \leq 4.4$.

4.3.1 K-correction

Calculation of rest-frame luminosity (or absolute magnitude) from the observed apparent magnitude of an object at non-zero redshift requires correction for the difference between the observed and rest-frame wavelengths and bandwidths. This is known as the K-correction (Humason et al. 1956).

Using the convention defined by Hogg et al. (2002), the K-correction between a bandpass R used to observe a source at redshift z and the same rest-frame bandpass of interest R , is defined by

$$m_R = M_R + D_M(z) + K_R(z) \quad (4.5)$$

where $D_M(z)$ is the distance modulus calculated from the luminosity distance D_L as $D_M = 5 \log_{10} \left(\frac{D_L}{10 \text{ pc}} \right)$, m_R is the apparent magnitude, and M_R is the absolute magnitude.

The K-correction for quasars takes the form $K(z) = -2.5(1 + \alpha_\nu) \log_{10}(1 + z)$

z_{em}	K correction
0.01	0.0075
0.02	0.0151
0.03	0.0241
0.04	0.0412
0.05	0.0720
0.06	0.1236
0.07	0.1619
0.08	0.1851
0.09	0.1959
0.10	0.1999
0.11	0.1985
0.12	0.1939

Table 4.1: K -correction in the SDSS r band (Fig. 4.2). A portion of the table is shown here. The full table is available in the on-line version of Tuccillo et al. (2015).

when the QSO spectrum is approximated by a power-law continuum, of spectral index α_ν ($S_\nu \propto \nu^\alpha$). Traditionally, most quasar luminosity functions (QLF) were computed using this K -correction.

However, although QSOs have an underlying power-law continuum, several studies (Cristiani & Vio 1990 ; Wisotzki 2000 ; Richards et al. 2005) have demonstrated that the broad-emission lines in their spectra have a significant impact on the total flux in a given band.

More accurate K -correction including the contribution of emission lines can be computed by convolving a typical QSO spectrum at different redshifts with the filter response (e.g. Cristiani & Vio 1990, and Wisotzki 2000, the latter based on optical/UV spectra from Elvis et al. 1994).

Following this approach, we calculated the K -correction by convolving the Vanden Berk et al. (2001) composite quasar spectrum with the SDSS r -filter:

$$K = 2.5 \log_{10} \left[(1+z) \frac{\int_0^\infty F(\lambda) S(\lambda) d\lambda}{\int_0^\infty F((\lambda)/(1+z)) S(\lambda) d\lambda} \right] \quad (4.6)$$

where $F(\lambda)$ is the measured intensity per unit wavelength, and $S(\lambda)$ is the r -band filter response. The resulting K -correction, shown in 4.2, allows calculation of the rest-frame absolute AB magnitude M_r from observed apparent magnitude r .

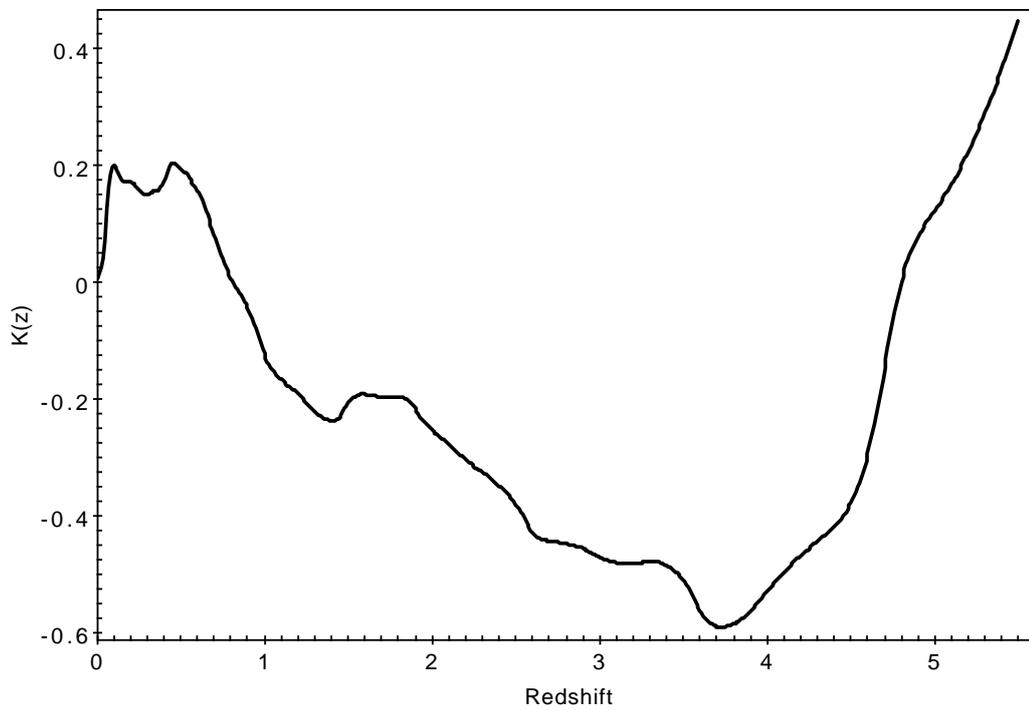


Figure 4.2: Computed K -correction in r as function of the redshift. A tabulation is available on-line, and a sample is shown in Table 4.1

4.3.2 Sample

In Table 4.3 we present the final sample of 87 QSOs with $3.6 \leq z \leq 4.4$ satisfying all our selection criteria. This sample includes 72 QSOs from the 5th quasar catalogue (Schneider et al. 2010) plus 15 QSOs revealed by our neural-network (8 from C08 and 7 from this work). Our sample is defined in r_{AB} with limiting magnitude 20.2, therefore we calculated in this band the absolute magnitude M_r , using the K -correction discussed in the previous section. Then to convert from M_r to the commonly used monochromatic absolute AB magnitude at 1450\AA , M_{1450} , we assume the canonical power-law spectral energy distribution with spectral index $\alpha_\nu = -0.5$ and

$$M_{1450} = M_r + 2.5\alpha_\nu \log_{10} \left(\frac{1450 \text{\AA}}{6231 \text{\AA}} \right) = M_r + 0.791 \quad (4.7)$$

where 6231\AA is the effective wavelength of the SDSS r filter. We use this spectral index instead of the spectral index derived by Vanden Berk et al. (2001) ($\alpha_\nu = -0.44$) to allow a direct comparison with other authors. We also avoid to use the composite spectrum in Vanden Berk et al. (2001) to convert from 6231\AA to 1450\AA rest-frame flux density since above $H\beta$ there is a high contribution from stellar light from low-redshift quasar hosts, which is subtracted off by extrapolating the power-law to red wavelengths.

Table 4.2: Final sample of 87 radio-loud QSOs with $3.6 \leq z \leq 4.4$ (continued on next page).

NAME	RA	DEC	r_{AB}	σ_r	Redshift	$S_{1.4\text{GHz}}$	M_r	$\log_{10} P_{1.4\text{GHz}}$	ID
(1)	(J2000)		(3)	(4)	(5)	(mJy)	(7)	(W/Hz)	(9)
J0153.40-001105	01:53:39.61	-00:11:05.0	18.83	0.022	4.19	4.82	-28.59	26.42	SDSS
J0300.25+003224	03:00:25.23	+00:32:24.2	19.68	0.025	4.18	7.75	-27.74	26.62	SDSS
J0725.18+370518	07:25:18.27	+37:05:18.4	19.60	0.020	4.33	26.72	-27.94	27.19	NED
J0747.11+273903	07:47:11.15	+27:39:03.4	18.35	0.025	4.15	1.08	-29.04	25.76	SDSS
J0747.38+133747	07:47:38.49	+13:37:47.3	19.35	0.015	4.17	7.18	-28.05	26.59	NED
J0751.13+312038	07:51:13.05	+31:20:38.0	19.73	0.020	3.76	5.84	-27.29	26.42	SDSS
J0751.22+452334	07:51:22.36	+45:23:34.2	20.20	0.033	3.61	1.18	-26.73	25.69	SDSS
J0807.11+131739	08:07:10.74	+13:17:39.4	20.00	0.026	3.73	48.20	-27.00	27.32	SDSS
J0810.10+384757	08:10:09.95	+38:47:57.1	19.62	0.018	3.94	26.68	-27.57	27.11	SDSS
J0823.23+155207	08:23:23.32	+15:52:06.8	19.30	0.018	3.78	74.93	-27.74	27.53	SDSS
J0833.23+095941	08:33:22.50	+09:59:41.2	18.69	0.016	3.73	122.52	-28.31	27.73	SDSS
J0838.08+534810	08:38:08.46	+53:48:09.9	19.94	0.032	3.61	8.47	-27.00	26.54	SDSS
J0839.46+511203	08:39:46.22	+51:12:02.9	19.31	0.016	4.39	40.50	-28.28	27.38	SDSS
J0840.44+341102	08:40:44.19	+34:11:01.6	19.78	0.020	3.89	13.64	-27.35	26.81	SDSS
J0852.57+243103	08:52:57.12	+24:31:03.2	19.46	0.016	3.62	157.30	-27.48	27.81	SDSS
J0855.02+182438	08:55:01.82	+18:24:37.7	19.96	0.020	3.96	9.38	-27.25	26.66	SDSS
J0857.24+110549	08:57:24.33	+11:05:49.2	19.81	0.017	3.71	1.91	-27.17	25.92	NOT
J0859.44+212511	08:59:44.06	+21:25:11.2	18.74	0.015	3.70	23.54	-28.24	27.01	SDSS
J0902.54+413507	09:02:54.17	+41:35:06.5	20.12	0.023	3.69	1.12	-26.85	25.68	WHT
J0909.54+474943	09:09:53.85	+47:49:43.2	19.90	0.020	3.64	373.29	-27.05	28.19	NOT
J0914.36+503849	09:14:36.23	+50:38:48.5	20.19	0.028	3.62	47.98	-26.75	27.30	NOT
J0918.24+063653	09:18:24.38	+06:36:53.4	19.76	0.022	4.19	25.87	-27.66	27.15	SDSS

The columns give the following: (1) SDSS object-ID (2) SDSS J2000 coordinates; (3) SDSS dereddened PSF r magnitude; (4) error in PSF r magnitude as given in SDSS; (5) QSO redshift determined in this work or from SDSS; (6) FIRST peak radio flux density; (7) absolute r magnitude ; (8) radio luminosity at 1.4 GHz; (9) indicates where the source redshift was first obtained (the two NED sources are from Benn et al. 2002).

Table 4.3: *Final sample of 87 radio-loud QSOs with $3.6 \leq z \leq 4.4$ (continued on next page).*

NAME	RA	DEC	r_{AB}	σ_r	Redshift	$S_{1.4\text{GHz}}$	M_r	$\log_{10} P_{1.4\text{GHz}}$	ID
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
J0926.40-023041	09:26:40.29	-02:30:41.5	19.82	0.021	3.76	1.90	-27.20	25.93	NOT
J0928.33+184824	09:28:32.88	+18:48:24.4	17.54	0.016	3.77	8.67	-29.49	26.59	SDSS
J0937.14+082859	09:37:14.49	+08:28:58.5	18.58	0.015	3.70	3.47	-28.40	26.18	SDSS
J0940.03+511603	09:40:03.03	+51:16:02.7	19.00	0.014	3.60	12.90	-27.94	26.72	SDSS
J1000.12+102152	10:00:12.26	+10:21:51.9	19.54	0.023	3.64	21.87	-27.41	26.96	SDSS
J1017.48+342738	10:17:47.76	+34:27:37.9	20.00	0.031	3.69	2.83	-26.98	26.08	SDSS
J1030.56+432038	10:30:55.95	+43:20:37.7	19.84	0.025	3.70	31.29	-27.14	27.13	SDSS
J1034.20+414938	10:34:20.43	+41:49:37.5	20.12	0.029	4.00	2.17	-27.12	26.03	NOT
J1034.47+110214	10:34:46.54	+11:02:14.5	18.81	0.025	4.27	1.15	-28.68	25.81	SDSS
J1037.18+182303	10:37:17.72	+18:23:03.1	19.81	0.023	4.05	13.59	-27.48	26.84	SDSS
J1051.21+612038	10:51:21.37	+61:20:38.1	18.92	0.020	3.69	6.90	-28.06	26.47	SDSS
J1057.56+455553	10:57:56.26	+45:55:53.1	17.45	0.022	4.14	1.10	-29.93	25.77	SDSS
J1101.48+001039	11:01:47.89	+00:10:39.4	20.18	0.026	3.69	192.10	-26.79	27.92	SDSS
J1105.44+255343	11:05:43.87	+25:53:43.1	20.09	0.026	3.75	1.69	-26.92	25.87	SDSS
J1109.46+190258	11:09:46.44	+19:02:57.6	20.05	0.024	3.67	7.22	-26.91	26.49	SDSS
J1110.55+430510	11:10:55.22	+43:05:10.1	18.58	0.024	3.82	1.21	-28.50	25.74	SDSS
J1117.02+131115	11:17:01.90	+13:11:15.4	18.28	0.018	3.62	28.00	-28.66	27.07	SDSS
J1117.36+445656	11:17:36.33	+44:56:55.7	20.05	0.026	3.85	24.33	-27.05	27.05	SDSS
J1123.40+291711	11:23:39.60	+29:17:10.8	19.46	0.017	3.77	3.68	-27.57	26.22	SDSS
J1125.30+575723	11:25:30.50	+57:57:22.7	19.43	0.036	3.68	2.52	-27.54	26.03	SDSS
J1127.49+051141	11:27:49.45	+05:11:40.6	19.14	0.012	3.71	2.34	-27.84	26.01	SDSS
J1129.39+131232	11:29:38.73	+13:12:32.3	18.78	0.026	3.61	1.77	-28.16	25.86	SDSS

The columns give the following: (1) SDSS object-ID (2) SDSS J2000 coordinates; (3) SDSS dereddened PSF r magnitude; (4) error in PSF r magnitude as given in SDSS; (5) QSO redshift determined in this work or from SDSS; (6) FIRST peak radio flux density; (7) absolute r magnitude ; (8) radio luminosity at 1.4 GHz; (9) indicates where the source redshift was first obtained (the two NED sources are from Benn et al. 2002).

Table 4.4: (continued)

NAME	RA	DEC	r_{AB}	σ_r	Redshift	$S_{1.4\text{GHz}}$	M_r	$\log_{10} P_{1.4\text{GHz}}$	ID
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
J1133.31+380638	11:33:30.91	+38:06:38.2	19.71	0.025	3.63	1.39	-27.23	25.76	SDSS
J1137.29+375224	11:37:29.43	+37:52:24.2	20.18	0.032	4.17	2.70	-27.22	26.16	SDSS
J1150.46+424001	11:50:45.61	+42:40:01.1	19.88	0.019	3.87	1.82	-27.23	25.93	SDSS
J1154.49+180204	11:54:49.36	+18:02:04.4	19.61	0.024	3.69	37.99	-27.36	27.21	SDSS
J1204.47+330939	12:04:47.15	+33:09:38.8	19.23	0.025	3.62	1.10	-27.71	25.66	SDSS
J1213.29-032726	12:13:29.43	-03:27:25.7	19.64	0.027	3.67	23.37	-27.33	27.00	NOT
J1220.28+261904	12:20:27.96	+26:19:03.6	18.13	0.017	3.70	34.34	-28.85	27.17	SDSS
J1231.42+381659	12:31:42.17	+38:16:58.9	20.18	0.030	4.14	20.44	-27.19	27.03	SDSS
J1240.55+543652	12:40:54.91	+54:36:52.2	19.74	0.023	3.94	14.89	-27.44	26.86	SDSS
J1242.10+372006	12:42:09.81	+37:20:05.6	19.34	0.018	3.84	644.79	-27.75	28.47	SDSS
J1244.43+060935	12:44:43.07	+06:09:34.6	19.78	0.024	3.76	1.29	-27.24	25.76	NOT
J1246.59+120855	12:46:58.83	+12:08:54.7	20.01	0.024	3.80	1.07	-27.05	25.69	SDSS
J1249.44+152707	12:49:43.67	+15:27:07.1	19.34	0.019	3.99	1.75	-27.90	25.94	SDSS
J1303.49+002011	13:03:48.94	+00:20:10.5	18.89	0.019	3.65	1.08	-28.06	25.66	SDSS
J1307.39+150752	13:07:38.83	+15:07:52.1	19.72	0.027	4.11	3.44	-27.63	26.26	SDSS
J1312.43+084105	13:12:42.86	+08:41:05.0	18.52	0.014	3.74	4.41	-28.49	26.29	SDSS
J1315.37+485629	13:15:36.58	+48:56:29.1	19.76	0.025	3.62	9.94	-27.18	26.61	SDSS
J1325.12+112330	13:25:12.49	+11:23:29.8	19.32	0.022	4.41	69.39	-28.28	27.61	SDSS
J1348.54+171150	13:48:54.37	+17:11:49.6	19.13	0.021	3.62	2.10	-27.81	25.94	SDSS
J1354.07-020603	13:54:06.90	-02:06:03.2	19.17	0.018	3.72	709.05	-27.82	28.49	SDSS
J1355.55+450421	13:55:54.56	+45:04:21.1	19.36	0.021	4.09	1.48	-27.98	25.89	SDSS
J1406.36+622543	14:06:35.67	+62:25:43.4	19.73	0.020	3.89	11.03	-27.41	26.72	WHT

The columns give the following: (1) SDSS object-ID (2) SDSS J2000 coordinates; (3) SDSS dereddened PSF r magnitude; (4) error in PSF r magnitude as given in SDSS; (5) QSO redshift determined in this work or from SDSS; (6) FIRST peak radio flux density; (7) absolute r magnitude ; (8) radio luminosity at 1.4 GHz; (9) indicates where the source redshift was first obtained (the two NED sources are from Benn et al. 2002).

Table 4.5: (continued)

NAME	RA	DEC	r_{AB}	σ_r	Redshift	$S_{1.4\text{GHz}}$	M_r	$\log_{10}P_{1.4\text{GHz}}$	ID
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
J1408.51+020523	14:08:50.91	+02:05:22.7	19.08	0.017	4.01	1.47	-28.17	25.87	SDSS
J1422.10+465933	14:22:09.70	+46:59:32.5	19.72	0.022	3.81	10.56	-27.35	26.68	SDSS
J1423.26+391226	14:23:26.48	+39:12:26.3	20.15	0.024	3.92	6.63	-27.01	26.50	SDSS
J1434.13+162853	14:34:13.06	+16:28:52.7	19.86	0.022	4.19	4.90	-27.56	26.42	SDSS
J1435.34+543559	14:35:33.78	+54:35:59.2	20.04	0.025	3.81	93.26	-27.02	27.63	SDSS
J1445.43+490249	14:45:42.76	+49:02:48.9	17.32	0.009	3.87	2.51	-29.80	26.07	SDSS
J1446.43+602714	14:46:43.37	+60:27:14.4	19.79	0.033	3.78	1.87	-27.25	25.92	SDSS
J1453.29+481725	14:53:29.01	+48:17:24.9	20.12	0.030	3.77	4.42	-26.91	26.30	WHT
J1503.29+041949	15:03:28.89	+04:19:49.0	17.96	0.017	3.66	122.70	-29.00	27.72	SDSS
J1506.44+533134	15:06:43.81	+53:31:34.5	18.94	0.022	3.79	13.97	-28.11	26.80	SDSS
J1511.47+252424	15:11:46.99	+25:24:24.3	19.95	0.024	3.72	1.24	-27.04	25.73	SDSS
J1520.28+183556	15:20:28.14	+18:35:56.2	19.82	0.021	4.12	6.26	-27.54	26.52	SDSS
J1533.36+054357	15:33:36.14	+05:43:56.5	19.84	0.020	3.94	27.55	-27.34	27.13	SDSS
J1617.16+250208	16:17:16.49	+25:02:08.2	19.98	0.023	3.94	1.01	-27.20	25.69	SDSS
J1619.34+302115	16:19:33.65	+30:21:15.1	19.53	0.025	3.81	3.88	-27.53	26.25	SDSS
J1637.08+091425	16:37:08.30	+09:14:24.6	19.57	0.018	3.75	9.51	-27.45	26.62	WHT
J1639.51+434004	16:39:50.52	+43:40:03.7	17.95	0.017	3.99	23.78	-29.28	27.07	SDSS
J1643.26+410343	16:43:26.24	+41:03:43.4	20.10	0.025	3.87	63.40	-27.02	27.47	SDSS
J2228.14-085526	22:28:14.40	-08:55:25.7	20.18	0.034	3.64	1.99	-26.77	25.92	WHT
J2235.36+003602	22:35:35.59	+00:36:02.1	20.15	0.027	3.87	5.06	-26.97	26.38	NOT
J2350.22-095144	23:50:22.40	-09:51:44.4	19.67	0.021	3.70	6.37	-27.31	26.44	WHT

The columns give the following: (1) SDSS object-ID (2) SDSS J2000 coordinates; (3) SDSS dereddened PSF r magnitude; (4) error in PSF r magnitude as given in SDSS; (5) QSO redshift determined in this work or from SDSS; (6) FIRST peak radio flux density; (7) absolute r magnitude ; (8) radio luminosity at 1.4 GHz; (9) indicates where the source redshift was first obtained (the two NED sources are from Benn et al. 2002).

4.4 Completeness

Below we estimate the completeness of the sample of 87 radio-loud QSOs at $3.6 \leq z \leq 4.4$. There are several sources of incompleteness: exclusion of optical images of poor quality (section 4.4.1); radio-survey incompleteness and missed radio-optical identifications (section 4.4.2); and incompleteness of selection by the neural-network algorithm (section 4.4.3).

4.4.1 Optical image quality

Due to the sensitivity of our NN to data of poor photometric quality, in our pre-selection (see chapter 2), we discarded all the objects with 'fatal' errors, i.e. having 'fatal' error flags or magnitude errors larger than 0.2 in all five bands (Richards et al. 2002) or flagged CHILD.

Incompleteness due to the exclusion of fatal and non-fatal photometric errors during the SDSS selection of QSOs-candidates, was discussed by Richards et al. (2006) who applied a global 5% correction. Richards et al. (2002) define as 'non fatal' errors some empirical combination of SDSS flags generally associated with poor de-blends of complex objects and possibly indicating unreliable photometric colors. However this selection criterion is not applied in our pre-selection, nor by the SDSS selection for objects matched in FIRST. The incompleteness due to the exclusion of objects flagged with fatal and non fatal errors in SDSS was studied also by Vanden Berk et al. (2005), who estimated an incompleteness of 3.8% for point-like objects with $i < 19.1$. Previous evaluations of this selection effect (Croom et al. 2004) suggest 6% incompleteness for objects with $17.5 \leq i \leq 18.5$.

We assume that our incompleteness due to exclusion of CHILD objects is the fraction (27%) of such objects amongst $15 \leq r \leq 20.2$ QSOs in the 5th SDSS Quasar Catalogue. 96% of the QSOs in this catalogue were discovered by SDSS, and CHILD objects were not excluded.

These fractions are in approximate agreement with the statistics of Table 2.1 of chapter 2, which indicate 4.8% incompleteness due to 'fatal' errors (669 rejected out of a total 13956 of sources), and 31% incompleteness due to the exclusion of sources flagged as CHILD (4148 out of 13287).

The total completeness regarding the quality of the optical photometry is of 69% (0.95×0.73) from the references above and 66% from the simple analysis in this work (0.95×0.69) and we adopt the intermediate value of 68%.

We do not apply any correction for objects misclassified in SDSS as having galaxy morphology but being star-like objects. In fact we tested several random samples of sources with available spectra, and we quantified this source of incompleteness to be $< 0.03\%$.

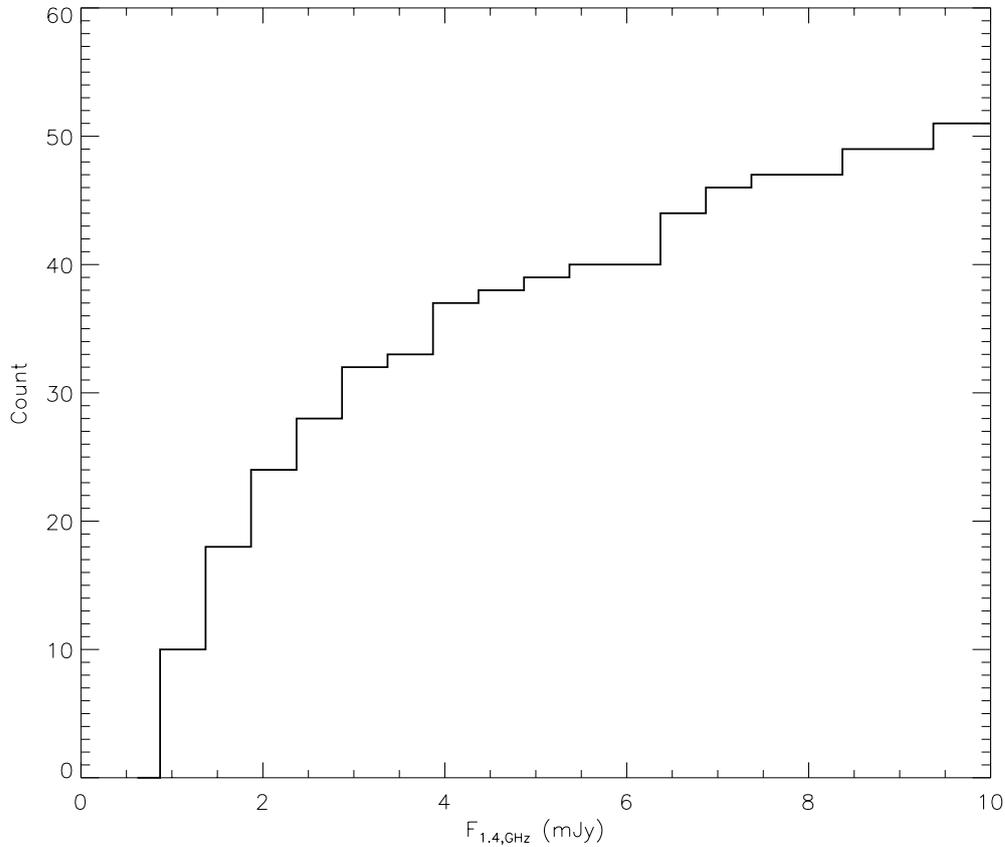


Figure 4.3: *Cumulative distribution of FIRST fluxes for our sample of 87, $3.6 \leq z \leq 4.4$ QSOs. For clarity we show only objects fainter than 10 mJy.*

4.4.2 Radio incompleteness

We have to consider two sources of incompleteness arising from radio selection using the FIRST survey.

Incompleteness of FIRST survey

In Fig. 4.3 we show for our sample the cumulative distribution of integrated radio flux densities, for objects fainter than 10 mJy.

The completeness of the FIRST survey as a function of flux density has been estimated for SDSS quasars and is given in Fig. 1 of Jiang et al. (2007): at integrated FIRST flux density > 3 mJy FIRST is $> 96\%$ complete and declines with decreasing flux density. The high incompleteness of the FIRST catalogue at faint fluxes is studied also in other work (e.g. Prandoni et al. 2001). Taking into account the results of these studies we assume the following completeness function,

$q(S)$, where S is the FIRST integrated flux density:

$$q(S) = \begin{cases} 0.50, & S \leq 1.25 \\ 0.75, & 1.25 < S \leq 2 \\ 0.85, & 2 < S \leq 3 \\ 0.95, & 3 < S \leq 5 \\ 1, & S > 5 \end{cases} \quad (4.8)$$

Applying $q(S)$ to our final sample we obtain a completeness of 84 per cent (87 QSOs detected and ~ 16 missed), for the flux density limit of 1 mJy. A subsample considering only the sources with $S > 2$ mJy would result in a completeness of 93 per cent (76 QSOs detected and ~ 6 missed).

Match radius

To obtain our radio-optical sample we used a simple one-to-one match using a $1''.5$ radius. We used this criteria for consistency with C08, originally based on the estimation that more than 99% of FIRST-APM QSOs with $3.8 \leq z \leq 4.5$, $E \leq 18.8$ and $S_{1.4GHz} > 1$ mJy fall within this radius Vigotti et al. (2003).

On the other hand, the use of a simple one-to-one match between FIRST and SDSS will miss double-lobe QSOs without detected radio cores. de Vries et al. (2006) found for a sample of 5, 515 FIRST-SDSS QSOs with radio morphological information within $450''0$, the fraction of FIRST-SDSS double-lobe QSOs with undetected cores is 3.7%. Since the starting samples of SDSS QSOs in de Vries et al. (2006) and in this work obey similar SDSS selection criteria, we applied this value to correct for this source of incompleteness.

4.4.3 Incompleteness of selection by the neural-network

The completeness of our neural network classifier was estimated as 97 per cent, from the testing on known high- z QSOs presented in Chapter 2. The classifier selected 15 high- z QSO candidates and rejected 2901. Of the later, 451 have now spectroscopy from DR9 and none of them was identified as a high- z QSO, confirming this value completeness for our algorithm.

4.4.4 Net completeness

The net completeness of our sample of 87 RL-QSOs at $3.6 \leq z \leq 4.4$, is the product of each of the completeness terms discussed in this section. Considering the required quality of the optical detection (68% completeness), the selected matching radius and the exclusion of extended sources (99% and 96.3%) and the

completeness of the NN selection algorithm (97%), we obtained a combined value of 63%.

Considering the flux-dependent completeness of the FIRST survey (given by function $q(S)$), our sample ranges from completeness of 84 per cent for $S > 1$ mJy, or 93 per cent for $S > 1.25$ mJy to the higher values of 97% or 100% for $S > 3$ or $S > 5$ mJy respectively. The net completeness of the whole sample is therefore of 53 per cent ($S > 1$ mJy) and increases to 58 per cent for $S > 1.25$ mJy and to 61 per cent for $S > 3$ mJy.

4.5 Binned Luminosity function: Method

Using the final sample of 87 QSOs listed in Section 4.4.2 and correcting for incompleteness as discussed in Section 4.5, we compute the binned Quasar Luminosity Function (QLF) in the redshift range $3.6 \leq z \leq 4.4$.

The binned QLF is usually calculated using the classical $1/V_{max}$ method (Schmidt 1968; Maccacaro et al. 1991; Ellis et al. 1996), or its generalized version (usually known as $\sum V_a^{-1}$) applied to samples comprising subsamples with different flux limits (Avni & Bahcall 1980).

The V_{max} method is an unbiased (Felten 1976) non-parametric estimator of the space density. It is commonly used to fit models of the LF, since it has the advantage that it does not assume any underlying model. Even when the model LF is fitted to the unbinned data (for example in the maximum likelihood technique of Marshall et al. (1983)) it is often used before performing the fit to observe the overall behavior of the LF.

However Page & Carrera (2000) demonstrated that the $\sum 1/V_{max}$ estimator introduces significant errors for objects close to the flux limits of the survey. An alternative method proposed by Page & Carrera (2000) is superior and partially corrects for this source of error, although implicitly assumes a uniform distribution of the sources within each bin (Croom et al. 2009; Miyaji et al. 2001). The variation of the LF within a bin can be particularly critical at the steep bright end of the QSO LF. Instead, we used a modified version of the Page & Carrera (2000) method that does not make use of the uniform-distribution assumption and is still model-independent.

To illustrate the method used in this paper, we start with a brief overview of the $1/V_{max}$ and the Page & Carrera (2000) methods. The luminosity function is defined as the number of objects per unit of comoving volume, per unit of luminosity. A naive approach to the calculation of space density in an interval $[L_1, L_2] \times [V_1(z_1), V_2(z_2)]$ of luminosity and redshift, centered upon values L^* and z^* , would be to simply count the number of objects N within the interval considered:

$$\Phi(L^*, z^*) = \frac{N}{\Delta V \Delta L} \quad (4.9)$$

The $1/V_{max}$ method, first proposed by Schmidt (1968) takes into account the fact that in flux-limited samples there is a higher probability to observe a bright source than a faint one. Thus, the count of sources N is replaced with a sum of probabilities:

$$\sum_1^N \frac{V_0}{V_{max,i}} \quad (4.10)$$

where V_0 is the volume over which we are computing the luminosity function, and $V_{max,i}$ is the maximum volume at which the source could be observed and still be included in our sample. In this way, the computation of the LF becomes

$$\Phi(L, z) = \frac{1}{\Delta L} \sum_1^N \frac{1}{\Delta V_{max,i}} \quad (4.11)$$

Page & Carrera (2000) noted that the limit in apparent magnitude of the survey bounds the region of integration. In particular, for a given bin in redshift, L_1 and L_2 should be replaced by the actual luminosity limits (L_{min}, L_{max}) as determined by the intersections with the limiting-magnitude curves of the survey. Therefore, $\Phi(L, z)$ is calculated as

$$\Phi(L, z) = \frac{N}{\int_{L_{min}}^{L_{max}} \int_{z_{min}}^{z_{max}(L)} (dV/dz) dz dL} \quad (4.12)$$

where z_{min} is the bottom of the redshift interval and $z_{max}(L)$ is the highest possible redshift for an object of luminosity L within the considered bin Δz . This approach takes into account the real area of integration but implicitly assumes uniform distribution of sources over the bin.

In order to minimise this bias, we calculate the maximum actual integration area determined as in Page & Carrera (2000) but for each source in the bin, so as not to lose the V_{max} information for individual sources. Then we sum over all the sources in the bin. In this way we do not count the number of sources over an area larger than that of the actual survey and at same time, within a single bin, we weight sources by luminosity. Finally, $\Phi(L, z)$ is calculated as

$$\Phi(L, z) = \sum_{i=1}^N \frac{1}{\int_{L_{min}}^{L_{max,i}} \int_{z_{min}}^{z_{max}(L),i} (dV/dz) dz dL} \quad (4.13)$$

The difference between this methodology and that of Page & Carrera (2000) is negligible in the case of a large sample of QSOs uniformly distributed in (M, z)

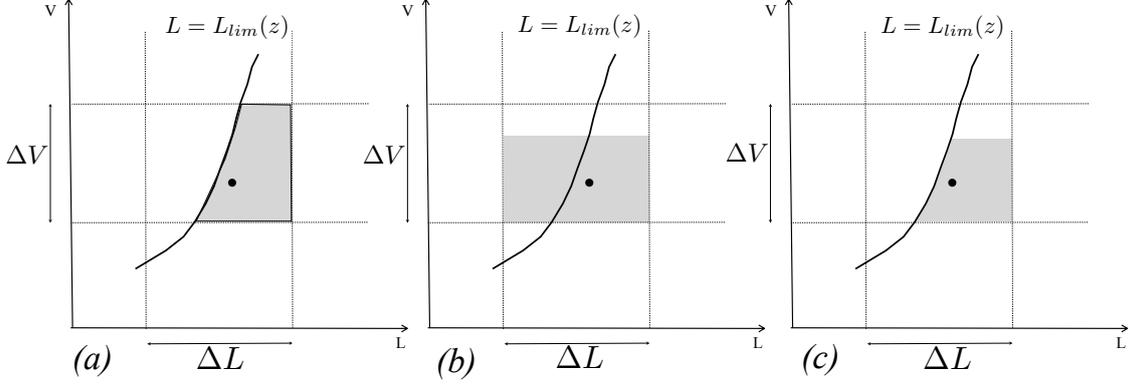


Figure 4.4: Volume-luminosity space used to calculate the contribution to the LF from a single object (the black dot) in a given bin intersected by the line $L = L_{lim}(z)$, i.e. the minimum detectable luminosity of an object at redshift z . The available space (grey shaded area) is shown for a binned LF calculated using: (a) the Page & Carrera (2000) implementation, (b) the classical $1/V_{max}$ method, (c) our methodology.

space, but becomes critical for small samples not distributed uniformly in each bin, as is the case here. In Fig. 4.4 we show the volume-luminosity space available to an object in a bin intersected by a limiting-magnitude curve, (a) in the Page & Carrera (2000) approach, (b) in the classical $1/V_{max}$ case, and (c) for the methodology used here.

The statistical uncertainty $\delta\Phi$ is calculated for each bin i as

$$\delta\Phi_i = \frac{\Phi_i}{\delta N_0} \quad (4.14)$$

Where N_0 is the actual number of objects in the bin and $\delta N_0 = \sqrt{N_0}$ (Poisson statistics). The formula is easily derived, when the space density Φ_i is assumed to be approximately:

$$\Phi_i \approx \frac{N}{V_e L_e} = \frac{N_0 \cdot f}{V_e L_e} \quad (4.15)$$

where $(V_e L_e)$ is an equivalent space-luminosity area, and N is the corrected number of quasars in bin i . The error in the completeness factor f is assumed to be ≈ 0 .

4.6 The QSO Luminosity Function

The QLF was computed using two bins in redshift, 3.6 - 4.015 and 4.015 - 4.415, and 11 bins in optical absolute magnitude starting with $M_r = -26.6$ and with

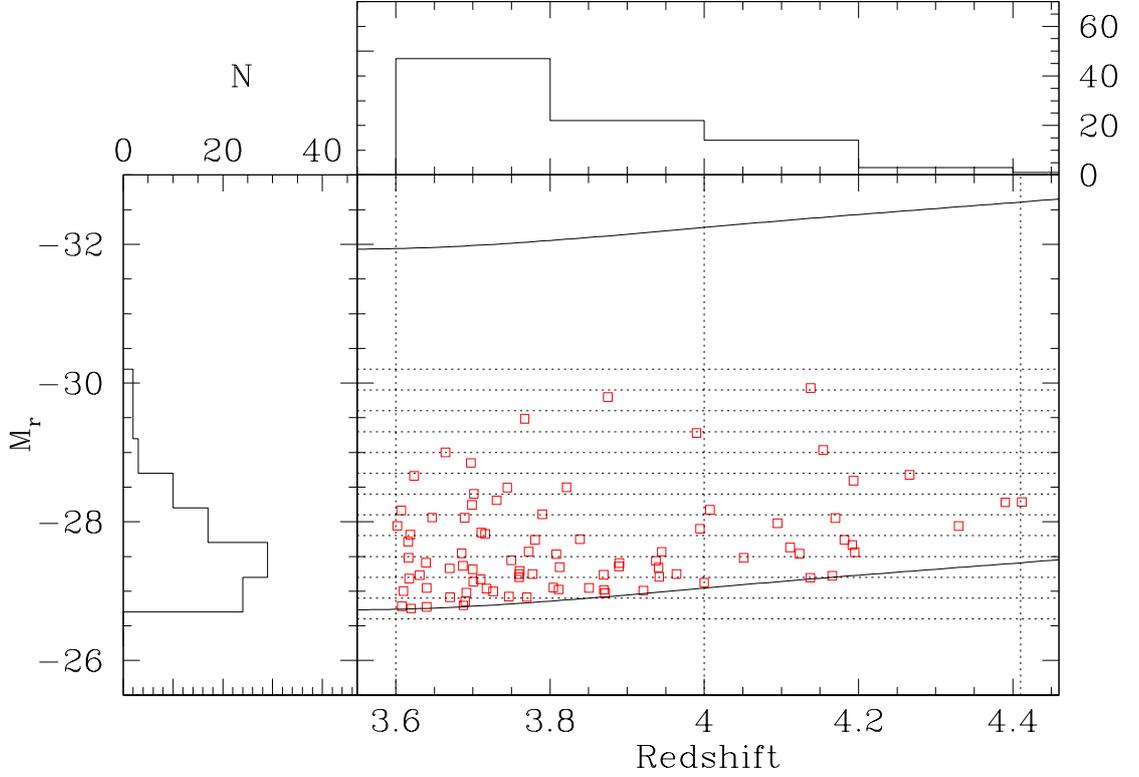


Figure 4.5: Luminosity-redshift diagram for the complete sample of 87 radio loud QSOs. The dotted lines show the limits of the bins ($\Delta L \Delta z$) used to compute the luminosity function. Solid lines represent the upper ($r = 15$) and lower ($r = 20.2$) limits of the survey. Upper and left panels show marginal histograms of redshift and absolute magnitude, respectively (see section 4.6).

$\Delta M = 0.3$. In Fig. 4.5 we plot M_r vs. redshift for our sample of QSOs; the dotted grid shows the bins in magnitude and redshift used to compute the QLF. The curves show the upper and lower limiting apparent magnitude r of our selection. The top and side panels show the marginal distributions in redshift and absolute magnitude, respectively.

Since our complete sample results from two surveys with different flux limits, the maximum redshift at which a source can be observed may be different in the two surveys. The most efficient way to combine areas with different flux limits is to assume that each object could be found in any of the survey areas for which it is brighter than the corresponding flux limit. This is ‘coherent’ addition of samples, in the language of Avni & Bahcall (1980). Therefore, since our survey is a radio-optical survey, for each source we chose the smaller of $z_{max,optical}$ and $z_{max,radio}$.

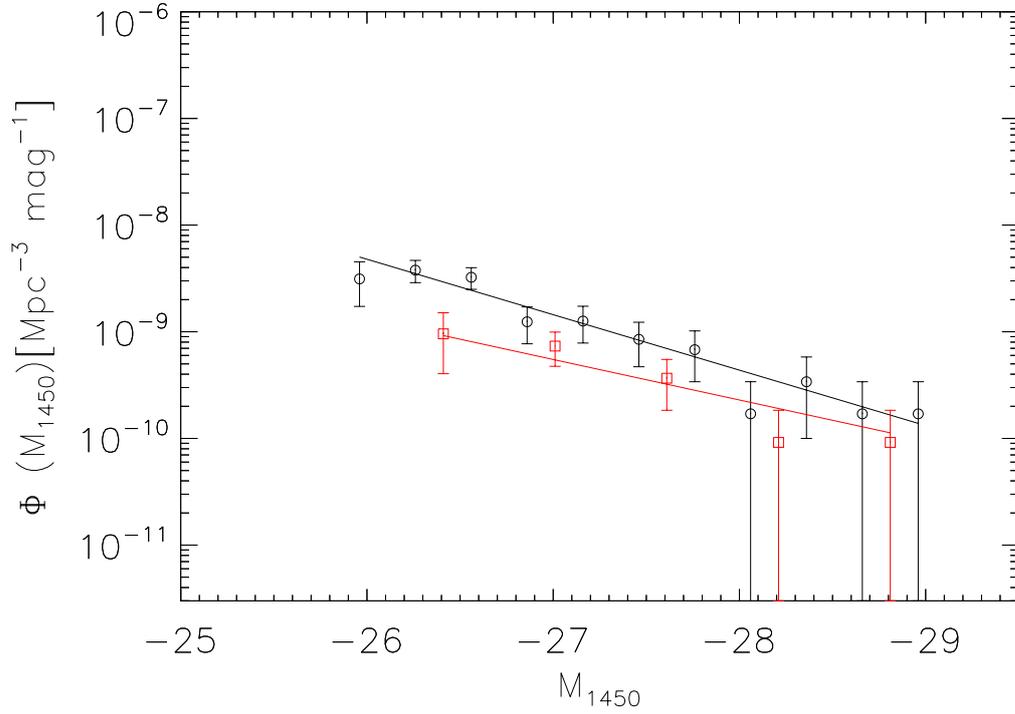


Figure 4.6: *The luminosity function derived from our sample of 87 radio-loud QSOs with $\log_{10}P_{1.4,\text{GHz}}(\text{W}/\text{Hz}) > 25.7$. The luminosity function for $z \sim 3.8$ is shown with black points, while red squares show the luminosity function for $z \sim 4.2$. Black and red lines are the best fit-slopes.*

For each bin we applied the completeness corrections as explained in Section 6 and computed the weighted number of QSOs in the bin. Table 4.6 and Fig. 4.6 show the resulting QLF for radio-loud QSOs. In Fig. 4.6 we plot separately the luminosity function for the two bins of redshift, at $z \sim 3.8$, and $z \sim 4.2$. We also show the best-fit slopes which will be discussed in section 4.7.3.

Table 4.6: *Binned luminosity function for FIRST-SDSS quasars at redshift ~ 4 .*

Redshift	M_r	$\phi_{RL}(\times 10^{-9})$	$\sigma_\phi(\times 10^{-9})$	RLF	$\phi(\times 10^{-8})$	fill	N_Q	N_{Qcorr}	N_r
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
3.8	-26.75	3.129	1.400	3.17(± 1.60)%	9.870	1	5	9.43	0
3.8	-27.05	3.778	0.890	3.58(± 1.82)%	10.559	1	18	33.96	0
3.8	-27.35	3.242	0.744	4.04(± 2.07)%	8.025	0	19	35.85	1
3.8	-27.65	1.240	0.469	4.56(± 2.36)%	2.717	0	7	13.21	1
3.8	-27.95	1.260	0.476	5.15(± 2.69)%	2.447	0	7	13.21	1
3.8	-28.25	0.849	0.379	5.80(± 3.01)%	1.463	0	5	9.43	0
3.8	-28.55	0.679	0.339	6.53(± 3.50)%	1.040	0	4	7.55	0
3.8	-28.85	0.170	0.170	7.35(± 4.00)%	0.231	0	1	1.89	0
3.8	-29.15	0.340	0.240	8.26(± 4.58)%	0.411	0	2	3.77	0
3.8	-29.45	0.170	0.170	9.26(± 5.24)%	0.183	0	1	1.89	0
3.8	-29.75	0.170	0.170	10.39(± 6.01)%	0.163	0	1	1.89	0
4.2	-27.05	1.920	1.920	3.04(± 1.60)%	6.316	1	1	1.89	0
4.2	-27.35	0.657	0.464	3.44(± 1.81)%	1.910	1	2	3.77	0
4.2	-27.65	0.917	0.410	3.89(± 2.07)%	2.358	0	5	9.43	0
4.2	-27.95	0.550	0.318	4.39(± 2.36)%	1.254	0	3	5.66	0
4.2	-28.25	0.367	0.260	4.95(± 2.69)%	0.741	0	2	3.77	0
4.2	-28.55	0.367	0.260	5.58(± 3.07)%	0.658	0	2	3.77	0
4.2	-29.15	0.183	0.183	7.07(± 4.01)%	0.260	0	1	1.89	0
4.2	-30.05	0.183	0.183	10.02(± 6.03)%	0.183	0	1	1.89	0

The columns give the following: (1) median redshift of the bin, (2) median absolute magnitude M_r of the bin, (3) space density ϕ_{RL} ($\text{Mpc}^{-3} \text{mag}^{-1}(\times 10^{-9})$) of the radio-loud QSOs, (4) error on the space density, $\sigma_\phi(\times 10^{-9})$, (5) radio-loud fraction calculated using Jiang et al. (2007), (6) space density ϕ ($\text{Mpc}^{-3} \text{mag}^{-1}(\times 10^{-8})$) of the QSOs (RQ+RL), (7) indication if the bin is intersected by the limiting magnitude curve (1 = yes and 0 = not), (8) actual number of QSOs in the bin, (9) the corrected number of QSOs in the bin after applying the completeness corrections, (10) Number of QSOs in the bin limited by the radio-flux limit.

4.7 Discussion

Below we use the binned LF calculated from the previous section to derive the space density of RL QSOs at $3.6 \leq z \leq 4.4$ (section 4.7.1). We then derive the space density of the entire population (i.e. radio-loud + radio-quiet) of QSOs at this redshift (section 4.7.2) using two independent estimates of the radio-loud-fraction (sections 4.7.2, 4.7.2). Finally we derive the slope of the LF of RL QSOs at $z \sim 3.8$ and at $z \sim 4.2$, and compare our results with those of other authors (section 4.7.3).

Our sample includes QSOs with optical luminosities $M_r < -26.6 \Leftrightarrow M_{1450} \lesssim -25.8$ (see equation 4.5). We adopt the definition of radio-loudness used by Gregg et al. (1996), i.e. $\log P_{1.4,\text{GHz}} > 25.5$. All the sources included in our sample (i.e. $S_{1.4\text{GHz}} > 1$ mJy) meet this criterion and are therefore radio-loud. Due to the limit in flux density, the minimum radio luminosity of sources included in our sample depends on redshift. In particular, using a radio spectral index of $\alpha_r = -0.3$, at the lowest redshift of our sample, i.e. $z = 3.6$, the FIRST flux-density limit corresponds to radio luminosity $\log P_{1.4,\text{GHz}}(W/Hz) > 25.61$. For $z = 4$, it corresponds to $\log P_{1.4,\text{GHz}}(W/Hz) > 25.7$ and for $z = 4.4$ to $\log P_{1.4,\text{GHz}}(W/Hz) > 25.77$.

4.7.1 The space density of RL QSOs at $3.6 \leq z \leq 4.4$

Starting with the binned luminosity function determined in Section 8.2 we calculate the space density of RL QSOs with optical luminosity $M_{1450} \lesssim -25.8$ and radio luminosity $\log P_{1.4,\text{GHz}}(W/Hz) > 25.5$, in two shells of redshift. The first shell has median $z \approx 3.8$ ($3.6 \leq z \leq 4.015$) and the second shell has median $z \approx 4.2$ ($4.015 \leq z \leq 4.415$). Integrating the binned LF (Table 6), the space densities of QSOs are therefore:

$$\begin{aligned} \rho(z \approx 3.8, M_{1450} < -25.8)_{\text{RL}} &= 4.51 \pm 0.61 \text{ Gpc}^{-3} \\ \rho(z \approx 4.2, M_{1450} < -25.8)_{\text{RL}} &= 1.54 \pm 0.63 \text{ Gpc}^{-3} \end{aligned}$$

From a sample of radio QSOs obtained by cross-matching the FIRST radio survey and the Automatic Plate Measuring Facility catalogue of POSS I, Vigotti et al. (2003) measured the space density at $3.8 \leq z \leq 4.5$ of optically-luminous ($M_{1450} < -26.9$) radio-loud QSOs and obtained $\rho(z \approx 4.1, M_{1450} < -26.9)_{\text{RL}} = 0.99 \pm 0.28 \text{ Gpc}^{-3}$. We recalculated the space density and optical luminosities using our adopted cosmology (noted at the end of Section 1), obtaining $\rho(z \approx 4.1, M_{1450} < -27.1)_{\text{RL}} = 0.66 \pm 0.18 \text{ Gpc}^{-3}$. By integrating our binned LF in the interval $M_{1450} \lesssim -27.0$, we obtain:

$$\begin{aligned}\rho(z \approx 3.8, M_{1450} < -27.0)_{\text{RL}} &= 1.09 \pm 0.24 \text{ Gpc}^{-3} \\ \rho(z \approx 4.2, M_{1450} < -27.0)_{\text{RL}} &= 0.50 \pm 0.16 \text{ Gpc}^{-3}\end{aligned}$$

in good agreement with Vigotti et al. (2003) (see Fig. 4.7) and consistent with a linear decrease of space density with increasing redshift.

Using a sample of QSOs obtained by cross-matching FIRST and SDSS-DR6, McGreer et al. (2009) calculated a binned luminosity function in the redshift range $3.5 \leq z \leq 4.0$. These authors used the same starting surveys as we did and a similar range of redshift, but they calculated the LF only for QSOs with radio-loudness parameter $R > 70$. The R parameter is another common criterion for distinguishing between radio-quiet and radio-loud AGN. It is defined (Kellermann et al. 1989; Stocke et al. 1992) as the rest-frame ratio of the monochromatic 6-cm (5 GHz) and 2500Å flux densities. Generally, objects are considered to be RL for $R > 10$.

The space density calculated in McGreer et al. (2009) for $M_{1450} < -26.1$ is $\rho(z = 3.75, M_{1450} < -26.1)_{R>70} = 1.38 \pm 0.59 \text{ Gpc}^{-3}$. The cosmology parameters used by McGreer et al. (2009) are the same that we use.

At redshift ≈ 4 our definition of radio-loudness is very close to the common definition $R > 10$, but we needed to re-calculate the LF using a subsample of RL-QSOs with $R > 70$ in order to compare our LF with McGreer et al. (2009). To calculate the R -parameter for our sample of QSOs, we used $\alpha_\nu = -0.5$ (in agreement with McGreer et al. 2009) to transform the flux from $S_{1.4\text{GHz}}$ to $S_{5\text{GHz}}$. We follow Oke & Gunn (1983) when converting from magnitude to luminosity (2500 Å). In this way, we obtain

$$\rho(z = 3.8, M_{1450} < -26.1)_{R>70} = 2.49 \pm 0.36 \text{ Gpc}^{-3} \quad (4.16)$$

which is a factor 1.8 (2σ) higher than the value $\rho = 1.38 \pm 0.59 \text{ Gpc}^{-3}$ found by McGreer et al. (2009). This difference may in part be ascribed to the higher completeness of our NN selection, and in part to the smaller FIRST-SDSS matching radius used by McGreer et al. (2009), which will exclude some quasars.

In Fig. 4.8 we show the cumulative luminosity functions for the two redshift bins (i.e. $z = 3.8$ and $z = 4.2$). Each point of the cumulative function is the space density $\rho(< M_{1450})$ as a function of absolute magnitude. The two functions can be compared with previous results by Vigotti et al. (2003) and Carballo et al. (2006), at redshifts $z \sim 4.1$ and $z \sim 4$, respectively. As expected, due to the evolution of space density with redshift, these last two values lie between our determinations.

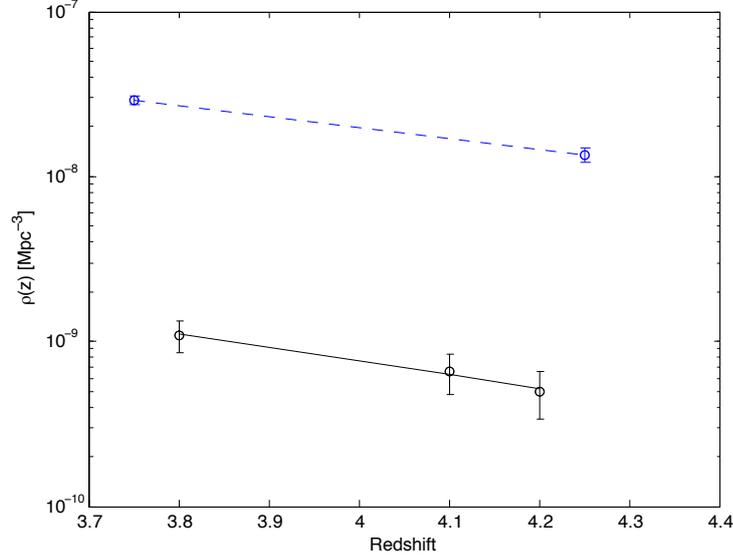


Figure 4.7: The black points show the integrated luminosity function for RL QSOs with $M_{1450} < -27.0$. The points for $z \sim 3.8$ and $z \sim 4.2$ were obtained in this work, showing good agreement with the point obtained for $z \sim 4.1$ in Vigotti et. (2003). For comparison, blue points show the space density for the entire population of QSOs and for $M_{1450} < -27.6$, as found in Richards et al. (2006). See section 4.7.1

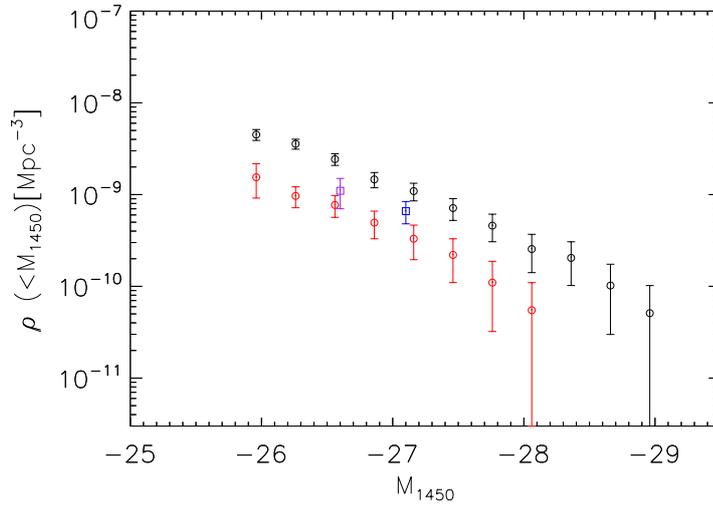


Figure 4.8: Cumulative luminosity functions at $z = 3.8$ (black circles) and $z = 4.2$ (red circles). Squares represent densities derived by Vigotti et al. (2003) (blue square) and Carballo et al. (2006) (purple square) at redshifts $z \sim 4.1$ and $z \sim 4$, respectively.

4.7.2 Total space density of QSOs at $3.6 \leq z \leq 4.4$

From the space density of the radio-loud QSO population we can roughly test the predictions of Radio-Loud-Fraction (RLF) over this range of redshift, by comparing the implied total space densities with measurements of space density from the literature.

It has long been known that between 5% and 15% of all quasars are radio-loud (e.g., Kellermann et al. 1989; Urry & Padovani 1995; Ivezić et al. 2002). However, some authors conclude that the RLF does not change significantly with redshift (e.g., Goldschmidt et al. 1999; Stern et al. 2000; Cirasuolo et al. 2003) or luminosity (e.g., Bischof & Becker 1997; Stern et al. 2000; Vigotti et al. 2003), while others find that the RLF decreases with increasing redshift (e.g., Peacock et al. 1986; Miller et al. 1990; Visnovsky et al. 1992; Schneider et al. 1992) and decreasing optical luminosity (Padovani 1993; La Franca et al. 1994); or that it evolves non-monotonically with redshift and luminosity (e.g. Hooper et al. 1995).

We therefore derive below the space density of all QSOs in two different ways: assuming constant RLF; and using a recently-determined redshift- and luminosity-dependent RLF (Jiang et al. 2007).

For constant RLF

From our binned luminosity function we derive the space density for $M_{1450} < -26.1$ (equivalent to the limit of $M_i < -27.6$ in Richards et al. 2006) as:

$$\begin{aligned}\rho(z = 3.8, M_{1450} < -26.1)_{\text{RL}} &= 3.57 \pm 0.44 \text{ Gpc}^{-3} \\ \rho(z = 4.2, M_{1450} < -26.1)_{\text{RL}} &= 1.54 \pm 0.63 \text{ Gpc}^{-3}\end{aligned}$$

In Vigotti et al. (2003) the RLF is assumed almost constant and is estimated as:

$$\text{RLF}(M_{1450} < -26.9) = 13.3\%$$

Therefore, for the total QSO population we obtain:

$$\begin{aligned}\rho(z = 3.8, M_{1450} < -26.1) &= 26.8 \pm 3.3 \text{ Gpc}^{-3} \\ \rho(z = 4.2, M_{1450} < -26.1) &= 11.6 \pm 4.7 \text{ Gpc}^{-3}\end{aligned}$$

in good agreement with the space densities derived from Richards et al. (2006):

$$\begin{aligned}\rho(z = 3.75, M_{1450} < -26.1) &= 29.0 \pm 2.0 \text{ Gpc}^{-3} \\ \rho(z = 4.2, M_{1450} < -26.1) &= 13.6 \pm 1.3 \text{ Gpc}^{-3}\end{aligned}$$

For redshift- and luminosity-dependent RLF

Jiang et al. (2007) use a sample of more than 30,000 optically selected QSOs from the SDSS to study the evolution of the RLF as a function of redshift and luminosity. They find that the RLF of QSOs decreases with increasing redshift and decreasing luminosity, according to:

$$\log \frac{RLF}{(1 - RLF)} = b_o + b_z \log(1 + z) + b_M (M_{2500} + 26) \quad (4.17)$$

where M_{2500} is the absolute magnitude at rest frame 2500 Å. The parameters b_o , b_z and b_M depend on the value of radio loudness and for $R > 10$; they are $b_o = -0.132 \pm 0.116$, $b_z = -2.052 \pm 0.261$, $b_M = -0.183 \pm 0.025$. M_{2500} is calculated from M_r as

$$M_{2500} = M_r + 2.5\alpha_\nu \log \left(\frac{2500 \text{ \AA}}{6231 \text{ \AA}} \right) \quad (4.18)$$

We used the Jiang et al. (2007) formula to obtain for each bin the corresponding value of the RLF (column (5) in Table 5). The RLF lies in the range 3%-10%, and increases with decreasing ϕ_{RL} . Applying the corresponding RLF to each bin and integrating we obtain:

$$\begin{aligned} \rho(z \approx 3.8, M_{1450} < -26.1) &= 81.7 \pm 31.7 \text{ Gpc}^{-3} \\ \rho(z \approx 4.2, M_{1450} < -26.1) &= 41.0 \pm 31.1 \text{ Gpc}^{-3} \end{aligned}$$

This is a factor ~ 3 higher than the results from Richards et al. (2006), but still within 2σ , due to the large errors in the luminosity function and in the RLF. In particular, the errors on the RLF at this redshift and magnitude are $\sim 50\%$. Given that our determination of the radio-loud luminosity function agrees reasonably well with McGreer et al. (2009) and with Vigotti et al. (2003), this discrepancy cannot be attributed solely to a possible overestimation of our luminosity function but may also be due to a systematic underestimation of the RLF in Jiang et al. (2007). The large quoted errors invite caution when using the Jiang et al. (2007) formula to determine the fraction of radio-loud quasars at high redshifts.

4.7.3 The bright-end slope of the luminosity function for RL QSOs

As already mentioned we calculated the LF in terms of optical luminosity in two bins of redshift. We compare our results and the best-fit slope with those of Richards et al. (2006) for the entire population of QSOs, and with the results

of McGreer et al. (2009) for RL QSO with $R > 70$. The limiting magnitude of the QSOs samples used by these authors ($M_{1450} < -26.1$ for Richards et al. 2006 and McGreer et al. 2009) was considered bright enough and far from the break luminosity to approximate the LF by a single power law $\propto L^\beta$.

This kind of approximation led in recent years to a long debate about an apparent flattening of the bright-end slope for $z > 4$, after it was noticed in early high-redshift surveys (Schmidt et al. 1995; Fan et al. 2001). These authors showed that the slope at $z > 4$ had a value $\beta \approx -2.5$, much shallower than the one seen at $z < 2.2$ ($\beta = -3.3$, Croom et al. 2004). This flattening was then confirmed by Richards et al. (2006) who used a large, homogeneous QSO sample from the SDSS-DR3 extending to $z = 5$. At higher redshift the constraints are weaker as they come from small samples, but in general they do not confirm a continued flattening of the slope with increasing redshift. In fact, Willott et al. (2010), combining the CFHQS (Canada-France High- z quasar survey) with the more luminous SDSS sample, derived the QLF from a sample of 40 QSOs at redshifts $5.74 < z < 6.42$ and found $-3.8 < \beta < -2.3$. At redshift $z \sim 6$, Jiang et al. (2008) find $\beta = -3.1 \pm 0.4$ using QSOs from SDSS Stripe 82.

Evolution of the shape of the QLF with redshift (changes in the slopes or in the location of the break luminosity) provides one of the fundamental observational constraints to the growth of super-massive black holes (SMBHs) over cosmic time. Assuming that brighter AGN have more-massive black holes, the flattening of the bright-end would be a remarkable indication of a downsizing of the SMBHs at high redshift. Downsizing was reported also by X-ray surveys (Ueda et al. 2003; Hasinger et al. 2005; see also Brusa et al. 2009).

On the other hand, recent work by Shen & Kelly (2012) and McGreer et al. (2013) aims to fill the gap in the QLF between $z \sim 3.5$ and $z \sim 6$, with the purpose of testing the flattening of the bright-end slope at $z > 3$. Shen & Kelly (2012) constrain the luminosity function by Bayesian modeling and using an homogeneous sample of SDSS-DR7 QSOs at $z = 0.3 - 5$. The results of Shen & Kelly (2012) and Richards et al. (2006) are, in general, in good agreement, finding that the curvature of the LF changes significantly beyond $z = 3$. However, Shen & Kelly (2012) suggest that the apparent flattening of the slope appears to be more related to a strong evolution of the break luminosity than a change in the bright-end slope. A similar conclusion is drawn by McGreer et al. (2013), who find no evidence for an evolution in the bright-end slope at $M_{1450} < -26$ for a sample of QSOs with $4.7 \leq z \leq 5.1$. On the other hand, McGreer et al. (2013) find evidence of strong evolution in the break luminosity, as it brightens from $M_{1450}^* \approx -25.4$ at $z = 2.5$ to $M_{1450}^* \approx -27.2$ at $z = 5$. They conclude that this evolution could flatten the bright-end slopes for surveys where the faint limit is near the break luminosity. McGreer et al. (2013) compared different models for the evolution of the QLF normalization and break luminosity. Eventually they found a good

fit of their data with recent results from the literature, using a modified version of a luminosity evolution and density evolution (LEDE) model proposed by Ross et al. (2012). In particular the evolution of the break luminosity in this model is log-linear (up to $z \sim 5$), with a break luminosity that brightens with redshift. This modified LEDE model predicts that for $z \sim 3.8$ the break luminosity would be $M_{1450}^* \sim -26.2$ and for $z \sim 4.2$ it would be $M_{1450}^* \sim -26.4$.

If we approximate the LF by a single power law $\propto L^\beta$, we find that in the first bin of redshift, $z \sim 3.8$, our best-fit slope is $\beta = -2.3 \pm 0.2$. As shown in Fig.4.9 our best-fit is in good agreement with the slope found in Richards et al. (2006) ($\beta = -2.4 \pm 0.1$), and with that found by McGreer et al. 2009 ($\beta = -2.2 \pm 0.2$). For the RL-QLF calculated in our second bin of redshift, i.e. $z \sim 4.2$, we rebinned the LF using $\Delta M = 0.6$, in order to reduce the statistical noise. In this way, as shown in Fig.4.10 the best-fit slope is $\beta = -2.0 \pm 0.4$. This result is consistent with the result found by Richards et al. (2006) for the entire population of QSOs, i.e. $\beta = -2.2 \pm 0.1$.

Our determinations of the bright-end slope for the RL population of QSOs at $z \sim 3.8$ and $z \sim 4.2$ are consistent with the flattening (between these redshifts-bins) of the bright-end slope found in Richards et al. (2006) for $z \geq 4$, which McGreer et al. (2013) suggest is due to a bias resulting from a single-fit power law in a region near the break luminosity.

As we have quasars with luminosities near or below the predicted break luminosity, we repeat the fit but excluding those points. In the first bin at redshift $z \sim 3.8$, we exclude the two fainter points. In this way we obtain a slightly steeper best-fit slope and a larger error: $\beta = -2.4 \pm 0.3$. This fit is shown in Fig.4.9 as a dashed line. In the second bin of redshift, $z \sim 4.2$, we exclude the faintest point, obtaining again a small increase of the slope: $\beta = -2.1 \pm 0.4$. This fit is shown again as a dashed line in Fig.4.10.

In light of the results from McGreer et al. (2013), our data do not strongly constrain the slope of the bright end nor the exact location of the break luminosity, especially considering the large errors of the brighter bins of the LF. Nevertheless this simple derivation is consistent with the results of McGreer et al. (2013).

In summary, our results are in good agreement with those of Richards et al. (2006) and McGreer et al. (2009). This result in itself is not trivial, because we are comparing different populations of QSOs in this range of redshift. In particular, we are comparing our results with the whole population of QSOs (by comparing with Richards et al. 2006) and with a population of RL QSOs where the radio-loudness is defined differently (being $R^* > 70$ for the RL sample of McGreer et al. 2009). We therefore have indications of a certain homogeneity of the QLF regardless of the differences in radio-loudness.

On the other hand, since we can't constrain the bright and the faint-end slopes, and we don't have an estimate of the break luminosity, we can't conclude that

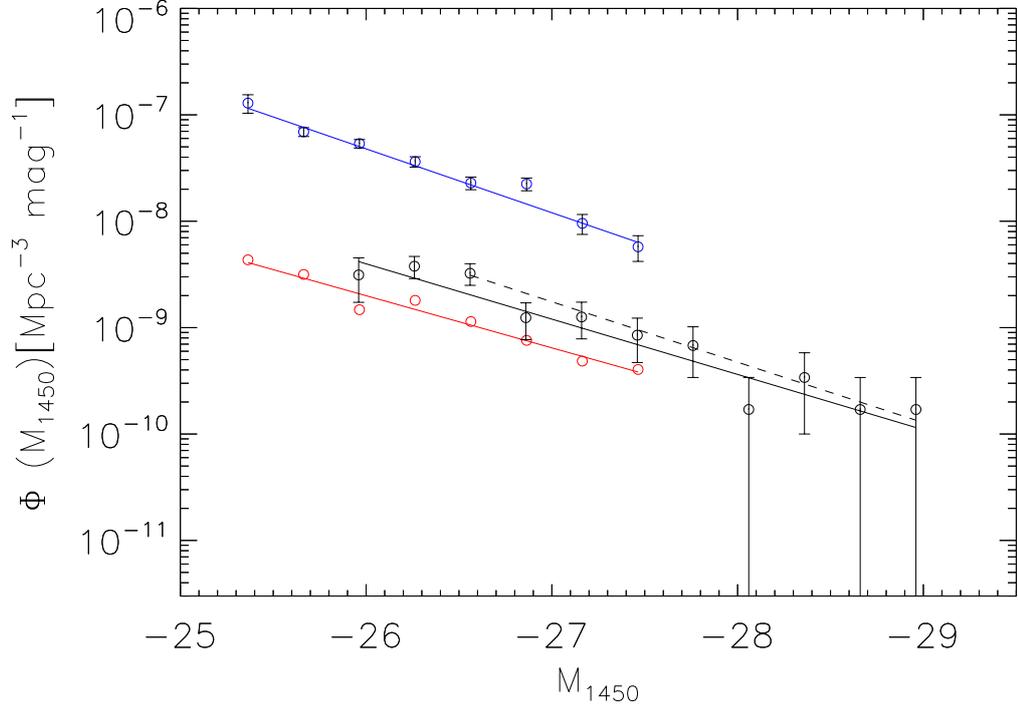


Figure 4.9: *Black points show the luminosity function derived for $z \sim 3.8$, for RL QSOs with $\log_{10} P_{1.4, \text{GHz}} (W/Hz) > 25.7$. For comparison, red points show the luminosity function calculated by McGreer et al. (2009), for RL QSOs with radio-loudness $R > 70$, and blue points the LF as calculated by Richards et al. (2006) for the entire population of QSOs. Best-fit slopes are $\beta = -2.3$ for our LF (black line); $\beta = -2.2$ for McGreer et al. (2009, red line), $\beta = -2.4$ for Richards et al. (2006, blue line). For our LF we obtain $\beta = -2.3$ (black line) and $\beta = -2.4$ after excluding the two fainter points (dashed line). See Section 4.7.3.*

the noted consistency of the slopes imply also consistency of all the parameters. Any differences of the slopes, or a different value of the break luminosity, could point to different density evolution of the RL and the RQ populations. In fact, Jiang et al. (2007) express the dependence of the RLF on optical luminosity as $\approx L^{0.5}$, implying $\beta_{RLQSO} \approx \beta_{QSO} + 0.5$, which is consistent with the differences of slopes that we find between the first ($z \sim 3.8$) and the second bin ($z \sim 4.2$) of redshift. Also, Baloković et al. (2012) and Kratzer (2014) find evidence that, at high redshift, the radio-loudness distribution of quasars is not a universal function, and likely depends on redshift and/or optical luminosity. We therefore need a sample of RL QSOs at fainter luminosities to constrain the faint-slope, and a larger survey area to extend the bright end of the luminosity function and thus, determine the break luminosity.

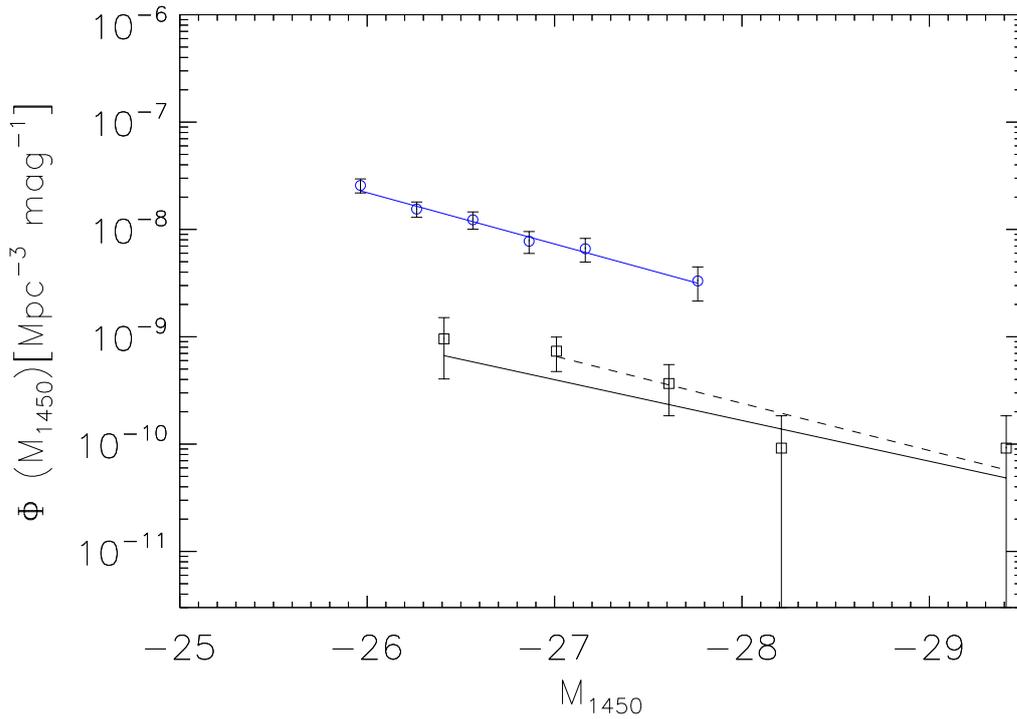


Figure 4.10: Black points show the luminosity function derived for $z \sim 4.2$, for RL QSO with $\log_{10} P_{1.4, \text{GHz}} (\text{W}/\text{Hz}) > 25.7$. For comparison, blue points show the LF as calculated by Richards et al. (2006) for the entire population of QSOs in the same bin of redshift. Best-fit slope is $\beta = -2.2$ for Richards et al. (2006, blue line). For our LF we obtain $\beta = -2.0$ (black line) and $\beta = -2.1$ after excluding the faintest point (dashed line) See section 4.7.3.

4.8 Conclusions

We construct a sample of high-redshift radio-loud QSOs $3.6 \leq z \leq 4$, and use it to measure the luminosity function and space density of QSOs in this range of redshift. Our principal conclusions are:

1. With the aid of the neural-network (see chapter 2 and 3), we construct a sample of 87 radio-loud QSOs at redshift ~ 4 . Of the various sources of incompleteness in the optical and radio surveys (Section 4.4), exclusion of SDSS 'CHILD' images is the main cause of our incompleteness relative to the SDSS selection of QSO candidates. But when applied to non-CHILD objects, our neural-network algorithm detects $\sim 97\%$ of the high- z QSOs, while SDSS only detect $\sim 85\%$ of them.
2. We determine the optical luminosity function for radio-loud QSOs in two redshift bins, $3.6 \leq z < 4.0$ and $4.0 \leq z \leq 4.4$ (Section 4.6), and measure the total comoving density of QSOs in these two redshift ranges (Fig. 4.7), obtaining a result consistent with that of Vigotti et al. (2003) at luminosities $M_{1450} < -27.6$. We also find good agreement between our cumulative luminosity functions (Fig. 4.8) and that measured by Vigotti et al. (2003) and Carballo et al. (2006), which determine the space density at intermediate redshifts.
3. Assuming a radio-loud fraction of 13.3% Vigotti et al. (2003) we estimate the total comoving density of QSOs (Section 4.7.2). The derived density of QSOs at $z \sim 4$ is consistent with that of Richards et al. (2006). Alternatively (Section 4.7.2), using the redshift- and luminosity-dependent radio-loud fraction found by Jiang et al. (2007), we measure a total comoving density of QSOs a factor 3 higher than measured by Richards et al. (2006). However, this result is significantly affected by the large error bars on the formula assumed for the radio-loud fraction (RLF).
4. We determine the slope of the luminosity function in two bins of redshift (Section 4.7.3). In the lower-redshift bin ($z = 3.8$) we found $\beta = -2.3 \pm 0.2$, consistent with Richards et al. (2006) and McGreer et al. (2009). In the higher redshift bin ($z = 4.2$) we find a slope $\beta = -2.0 \pm 0.4$ consistent with Richards et al. (2006). Values of the slope consistent with our determination, have been interpreted as a flattening of the bright end slope for the high- z QSOs population, but has recently been re-interpreted as the result of a strong evolution of the break luminosity for high- z QSO McGreer et al. (2009). The consistency of our results with Richards et al. (2006) and McGreer et al. (2009) suggests a similar evolution for both radio-loud and

radio-quiet populations. Our results can be also interpreted as suggestive of a flattening of the bright-end slope from $z \sim 2$ to $z \sim 4$, for the radio-loud population only. If confirmed, this implies an evolution of the density of super-massive black holes associated with radio-loud QSOs, in the sense that they were more abundant at $z \sim 4$. However, to clarify the evolution of the RL population relative to that of the whole population of QSOs, more observational constraints are needed, especially at redshifts above 4.

Chapter 5

High-z BALs QSOs

*Back then, the visions...most of the time I was convinced that I'd lost it.
But there were other times, I thought I was main-lining the secret truth
of the universe*

Rust Cohle, True Detective

5.1 Introduction

Broad absorption line (BAL) quasars exhibit broad, blue-shifted absorption lines from transitions in highly-ionized species such as C IV, Si IV, and N V. This type of absorption is present in $\approx 15\%$ of optically selected quasars, and it is caused by the outflow of gas with high velocity (up to $\approx 0.2c$; Hewett & Foltz 2003), as part of the accretion process. It is known that quasars may have broad impacts on the evolution of the host galaxy and its surrounding environment via feedback effects (for example, the well established M-sigma relationship; Ferrarese & Merritt 2000; Gebhardt et al. 2000), and BAL outflows likely play a role in these relationships. Establishing the demographics, geometry and other properties of these intrinsic BAL outflows is essential to properly incorporate their effects into feedback models relevant to galaxy evolution. Nevertheless, the nature and origin of these objects is still under debate. The two principal theoretical models proposed to explain their origin and properties, the *orientation scenario* (Elvis 2000; Murray et al. 1995) and the *evolutionary model* (Lípári & Terlevich 2006), can be efficiently tested studying the radio properties of subsample of radio-detected BAL QSOs. In fact, in the last years several works have studied this class of objects through multi-frequency radio-observations (Montenegro-Montes et al. 2008; Liu et al. 2008; DiPompeo et al. 2010; DiPompeo et al. 2011; Bruni et al. 2012), significantly enhancing the understanding of their synchrotron emission properties,

providing key indications about the nature of these objects. In spite of this, there is a lack of observational radio-information about BAL QSOs at extreme redshifts. In all the cited works they used samples of BAL QSOs within $1.2 \leq z \leq 3.4$ in studies where the emphasis was on the global difference in properties between BAL and non-BAL, not testing possible evolutionary effects at extreme redshift.

In the work presented in this chapter, we selected a sample of 22 high- z RL BAL QSOs within $3.6 \leq z \leq 4.8$, extending the known number of objects of this class. Afterwards we observed the brightest objects of this sample in multiple radio frequencies (at the EVLA radio-array and at the 100-m Effelsberg telescope), together with a comparison sample of unabsorbed RL QSOs matched in redshift and radio-optical fluxes. For 16 BAL and 14 unabsorbed QSOs we collected enough frequency coverage to reconstruct the SED and analyze the shape of their radio spectra. With these data we tested the orientation of the jet with respect to the line of sight (using the spectral indexes), and thus helping in verifying a possible preferred orientation. The shapes of the radio spectra also allowed us to evaluate statistically the radiative age of the sources (counting the number of CSS/GPS sources, Murgia et al. 1999) as has been successfully done in the previous-cited studies on BAL QSOs (Montenegro-Montes et al. 2008; DiPompeo et al. 2011; Bruni et al. 2012). The motivation of this study was to understand whether the radio-characteristics found in the BAL quasars at lower redshift are representative of the radio-loud BAL QSO population at higher redshift or if there are indications of evolutionary trends in BAL properties.

5.2 BAL Quasars: definition and theoretical models

Intrinsic absorption lines are a common feature in quasar spectra, and they are likely produced by outflowing winds along the line of sight that are launched from the accretion disk or other structures around the central supermassive black hole (Murray et al. 1995; Proga et al. 2000). Broad absorption lines (BALs) in ultraviolet and visible spectra of quasars are the most spectacular manifestation of such outflows, in which the absorption lines, shaped by the geometry and kinematic structure of the outflows absorptions lines, are blue-shifted as much as $\Delta v > 2000 \text{Kms}^{-1}$. Quasars that present BAL troughs in their spectra are often classified into three subtypes depending on the presence of absorption lines in specified transitions: (1) High-ionization (Hi) BAL quasars show absorption lines in high-ionization transitions including Si IV, and C IV ; (2) Low-ionization (Lo) BAL quasars possess Mg II and/or Al III absorption lines, in addition to the high-ionization transitions. (3) Iron low-ionization BAL quasars show additional

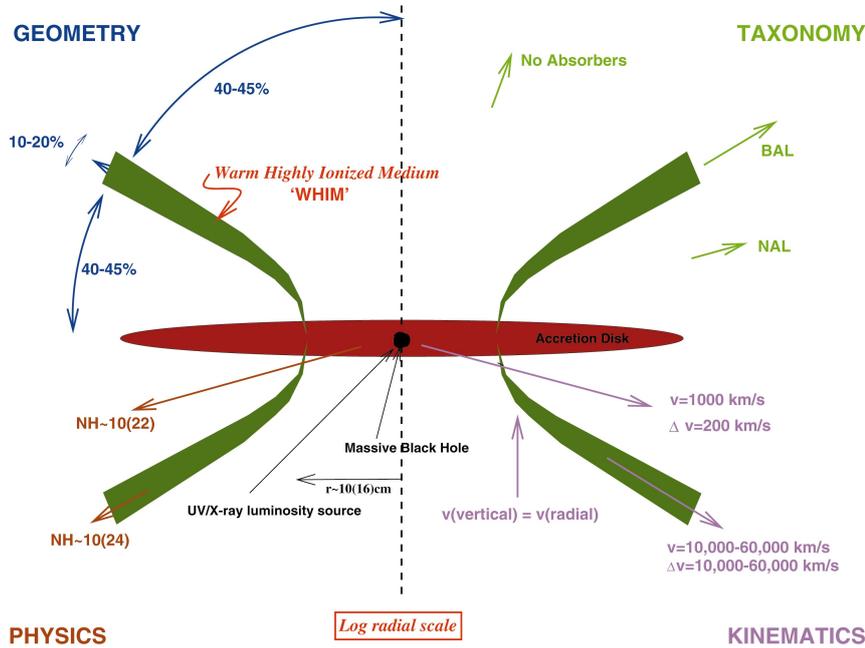


Figure 5.1: Illustrating the orientation model that explains the origin of the BAL in quasars. BALs are seen in the spectrum high inclination angles when the outflow intercept the line of sight (figure from Elvis 2000).

absorption from excited states of Fe II and Fe III (e.g., Hall et al. 2002). Around 10-15% of the whole population of quasars are classified as BAL.

Powerful outflows can enrich the quasar host galaxy interstellar medium, contributing to a variety of feedback effects such as regulating host galaxy star formation rates (Hopkins & Elvis 2010), and limits the quasar lifetimes by removing fuel from the nuclear regions (Silk & Rees 1998). Despite these important implications, the nature of BALQSOs is still not fully understood.

A theoretical model called *orientation scenario*, proposed by Murray et al. (1995) and Elvis (2000), suggests that BAL QSOs are normal quasars seen at high inclination angles (see Fig. 5.1), and outflows originated near the equatorial plane. The model does work interpreting the small differences in proprieties between BALs and non-BALs, though there has been little direct evidence to support it. In this model slightly redder continua (Reichard et al. 2003) and higher optical polarization (DiPompeo et al. 2010), could be a consequence of a preferred line-of-sight.

On the other hand, Becker et al. (2000) studied the radio properties of BAL QSOs and noted that an edge-on geometry should lead to lobe-dominated emission with a steep radio spectrum, while in their sample a third of the objects show

a flat two-point spectrum. This indicates significantly self-absorbed, relativistically beamed core-emission dominated objects, which would be seen when looking near the jet axis. A few other works have also shown that BAL QSOs have a wide range of spectral indices (α), overall suggesting that radio-loud BAL quasars are not oriented along a particular line of sight (Montenegro-Montes et al. 2008; Fine et al. 2011). It was also noted by Becker et al. (2000) and others that most radio-loud BAL QSOs seem to belong to the same class of compact steep spectrum (CSS) objects and gigahertz-peaked spectrum (GPS) objects. These compact objects with linear sizes of less than 20 kpc, are generally considered to be either young radio sources (Fanti et al. 1990), or radio sources frustrated by interaction with a dense environment (van Breugel et al. 1984).

These results all lead to the second main scenario proposed to explain BAL QSOs: the *evolutionary model*, discussed among others by L ipari & Terlevich (2006). In this view BAL QSOs are young systems with large covering fraction outflows accompanied by absorption clouds. The radio-loud systems may be associated with the later stages of evolution, with the jets responsible to remove the clouds generating of BALs. This model could also explain the rarity of extended radio sources also showing BAL features (Gregg et al. 2006), but in general lack of definitive evidences in its favor.

5.3 Radio studies of BALs QSOS

For quite a long period of time (from the discovery of the first BAL in 1967, Lynds 1967, to the late '90s) it was a common belief that BAL quasars were always radio quiet and never strong radio sources (Stocke et al. 1992). This view had to change when Becker et al. (1997) discovered the first radio-loud BAL quasar using the VLA FIRST survey, and then five other radio-loud BAL quasars were identified in NVSS by Brotherton et al. (1998). Since then, the number of radio-loud BAL QSOs has increased considerably together with the number of radio quiet quasars.

Radio observations are probably the best way to study BAL QSOs and to learn about their orientation and evolutionary status, and therefore test the two major scenario discussed in the previous section.

The few studies in radio have provided fundamental information about the nature of these objects, even by only studying a few examples of BAL quasars. Among them, Montenegro-Montes et al. (2008) studied a sample composed of the 15 radio brightest BAL QSOs, with flux densities $S_{1.4GHz} > 15$ mJy and measured radio flux densities in full polarization using both the 100-m Effelsberg telescope and the VLA array, covering a wide spectral range, from 1.4 up to 43 GHz. The physical properties of the BAL and non-BAL QSO samples were investigated by analyzing the synchrotron spectra spectral index determinations, since

flatter spectral indices imply lines of sight closer to the radio axis (Orr & Browne 1982). The synchrotron turn-over frequency, that allows to make considerations about the age or the frustration of the source, was also analyzed. Many of the radio characteristics of these sources were found to be prototypical of CSS or GPS sources, and a variety of spectral indices were present, allowing different orientations for the BAL producing outflow. These results are difficult to reconcile with a simple orientation scenario.

DiPompeo et al. (2011) increased dramatically the number of BAL QSOs with multi-frequency radio observations, observing a sample of 74 BAL QSOs built from SDSS and the FIRST survey at 4.9 and 8.4 GHz, along with an extremely well matched sample of non-BAL quasars. They identified a significant over-abundance of steep-spectrum BAL sources, with an increasing likelihood of seeing BALs at the largest viewing angles. However, simulations of the spectral index distributions show that there is significant overlap in the viewing angles to BAL and non-BAL QSOs (DiPompeo et al. 2012). This may indicate a situation where BAL outflows evolve as a function of angle from symmetry axis, as radio structures grow out of a cocoon of outflowing material. Bruni et al. 2012 obtained similar results, finding only a mildly preferred equatorial orientation, and similar fractions of young radio sources (GPS-CSS) in the two samples, considering a radio window between hundreds of MHz and 43 GHz. Their polarization properties were also found to be similar. Expanding the SEDs to lower frequencies thanks to GMRT (Giant Metrewave Radio Telescope) observations (235/610 MHz), Bruni et al. (2015) found that a significant fraction ($\sim 70\%$) of sources from the sample of Bruni et al. (2012) is in a GPS or CSS+GPS phase, suggesting a correlation between emerging radio components or restarting radio activity and the BAL phenomenon. In the same work, a study of dust abundance in the mm-band revealed that BAL QSOs are even dust-poorer than normal ones.

However, information about the detailed radio properties of BAL quasars is pretty limited. Interpretation of the full radio spectrum of these sources becomes important due to their compactness; of the small sample of compact BAL quasars that have been observed with VLBI (Jiang & Wang 2003; Kunert-Bajraszewska & Marecki 2007; Liu et al. 2008; Montenegro-Montes et al. 2008; Bruni et al. 2013), about half of them show unresolved radio structures even in the high-resolution observations; the others have core-jet structures indicating complex morphology. Kunert-Bajraszewska et al. (2010) had studied in the details the complex radio morphology of a BAL quasar belonging to a primary sample of 60 candidates for CSS sources selected from the Very Large Array (VLA) FIRST catalogue (White et al. 1997) and initially observed with all the sample with MERLIN at 5 GHz Marecki et al. (2003). This source is dominated by the strong radio jet resolved into many subcomponents and changing its orientation during propagation in the central regions of the host galaxy. They discuss the possible origins of the multi-

ple radio structures observed in this BAL and also suggest that there is no direct correlation between the jet-observer orientation and the possibility of observing BALs. Otherwise they stress that, as already suggested by Elvis (2000), not only equatorial disk wind but also cone outflows can be responsible for BAL features. Thus, an internal orientation effect, as the opening angle of the accretion disk wind, instead of the external orientation effect, as an angle between the jet axis and our line of sight, could be what determines a possibility of observing any BAL features.

All of these results on the one hand indicate the need for further research on BAL population, on the other hand clearly suggest that the phenomena in BAL quasars appears to be more complex than could be depicted by a simple orientation scenario.

5.4 Selection of RL BAL at high- z

In the last 20 years, the interest in BAL QSOs, has grown together with the number of known objects showing their characteristic features. In particular, in 1993 the number of known BAL quasar was relative small, just 72 (Korista et al. 1993), while now the largest BALs catalogue, made by Gibson et al. (2009), counts of 5,039 BAL quasars, obtained from the analysis of the spectra contained in the Sloan Digital Sky Survey (SDSS) Data Release 5 (DR5) QSO catalogue.

In spite of these large numbers, few are the known RL BAL QSOs at extremely high- z and consequently there are few studies on their behavior. The limit in redshift for the work of Montenegro-Montes et al. (2008) was $z = 3.4$ while the samples of DiPompeo et al. (2011) and Bruni et al. (2012) had a maximum of $z = 3.38$ and 3.57 , respectively. Extension of the radio studies to samples at higher redshift is useful in looking for evolutionary trends in BAL properties, and Allen et al. (2011) already found a dependence on redshift of the BAL QSO fraction, the latter decreasing by a factor 3.5 ± 0.4 from redshifts ~ 4.0 down to ~ 2.0 . They interpret this as an evolutionary behaviour, not reproducible by the orientation model alone.

With this motivation on mind, and starting from the sample of radio-optical sources discussed in chapter 2, we selected a complete sample of RL BAL QSOs in the redshift range $3.6 \leq z \leq 4.8$. The range of redshifts was chosen in order to mostly overlap the region used in Tuccillo et al. (2015) and discussed in this thesis in chapters 2,3 and 4. With this choice we search for BAL QSOs within a very complete spectroscopic sample of RL QSOs. Moreover, the careful determination of the quasar LF that we discussed in chapter 4, assures that the RQ and the RL QSOs populations do not evolve differently in this range of redshift. Therefore any possible difference found in this research between BAL and non BAL QSOs,

are not a-priori biased by different evolutions between RL and RQ populations.

We started our selection re-considering the sample of 222,517 sources obtained cross matching each FIRST source, not flagged as possible sidelobe or nearby bright source ($\sim 3.6\%$ of the sources in the FIRST catalogue), with the closest optical object in the "PhotoPrimary" view of the SDSS DR7 catalogue within a $1''.5$ radius (see Table 2.1 of Chapter 2). We remind that in this sample there is no selection by radio flux density or radio morphology other than the requirement that the radio source has at least a weak core component. From this sample we discarded all the sources not classified as "point-like" or tagged with the "fatal" error flags by the SDSS pipelines. Differently from the pre-selection presented in Table 2.1, here we did not select the sample by magnitude nor we excluded the sources tagged with the "CHILD" SDSS-flag. This way we ended-up with a sample of 36,267 sources.

At this point we searched for all the available optical spectra for the sources of this sample, either from the SpecObj view of SDSS-DR7 or from the 5th edition of the SDSS Quasar Catalogue (DR7 QSO Catalogue; Schneider et al. 2010). We integrated the sample with all the new quasars discovered with our neural-network selection strategies (Tuccillo et al. 2015). We visually inspected all the spectra of the sources classified as QSOs at $z > 3$ in the DR7 QSO catalogue or in the SpecObj view of SDSS-DR7, looking for BAL features. Finally we measured the absorption in the Ci IV for a rigorous and homogeneous selection of the final sample of BAL QSOs.

In literature there is more than one metric used to separate BALQSOs and non-BAL quasars on the basis of the measured absorption. The most widely used definitions are: (i) the *balnicity index* (BI, first presented by Weymann et al. 1991); (ii) the *absorption index* (AI) (Hall et al. 2002; Trump et al. 2006), designed to include narrower troughs than the BI; (iii) and the *modified balnicity index* BI_0 (Gibson et al. 2009), which extends the integration region to zero velocity. In this work, to discriminate between BAL and not-BAL QSOs, we choose to apply the same criteria used in Bruni et al. (2012), in order to compare our results with their studies at different redshift. Therefore we calculated the absorption index (AI), as defined in by Hall et al. (2002):

$$AI = \int_{0 \text{ Km/s}}^{25000 \text{ Km/s}} \left(1 - \frac{f(\nu)}{0.9}\right) C d\nu \quad (5.1)$$

where $f(\nu)$ is the continuum-normalized flux (unsmoothed whenever possible) as a function of velocity ν , relative to the line centre. We integrate the spectral region between the peaks of the Ci IV and Si IV emission lines to up to 25000 Km/s from the former. The constant C is equal to unity over contiguous troughs where $f(\nu)$ has been continuously less than 0.9 for at least 1000 Km/s (as in Trump et al.

2006), and zero elsewhere. We considered as genuine BAL QSOs only objects with an $AI > 100$. Applying this criteria we ended-up with a sample of 22 BAL QSOs within $3.6 \leq z \leq 4.8$. Comparing this sample with larger catalogues of BALs from literature, we verified that 9 of the sources included in our sample are included also in the catalogue of Allen et al. (2011), 10 are included also in the Gibson catalogue (Gibson et al. 2009), and 3 in the catalogue from Trump et al. (2006). 9 BALs of our sample are not classified as such in literature and they have been identified as BAL QSOs in this work for the first time. We note that one of the BAL QSOs of our sample (J074738.49+133747.3) has no spectra in SDSS, but it was identified by our group in a previous investigations (Carballo et al. 2008).

5.4.1 BALs and comparison samples

With the aim of repeating at high-z radio-studies similar to those of Montenegro-Montes et al. (2008), DiPompeo et al. (2011) and Bruni et al. (2012) we originally proposed to observe the 22 BAL QSOs of our sample and a "comparison sample" of 22 unabsorbed QSOs in multiple frequency bands, including 1.4 GHz (L-band), 4.9 GHz (C-band), 8.4 GHz (X-band), and 22 GHz (K-band) at the 100-m Effelsberg Telescope and at the EVLA radio-array in the A configuration.

The "matched sample" is composed of 22 non-BAL RL QSOs, selected in order to make meaningful comparisons with the parent populations of quasars. This sample was selected using the same criteria applied in DiPompeo et al. (2011), i.e. searching for each BAL QSOs of our sample, a correspondent non-BAL RL QSOs matched within 20% SDSS i-band magnitude, 20% of 1.4 GHz radio flux, and 10% of redshift. The two samples have similar properties to those used in DiPompeo et al. (2011) and Bruni et al. (2012), but extend the redshift range covered out to $z = 4.8$ and extend to lower optical and radio luminosities. These radio observations also included the frequencies used in the two cited works, in order to significantly increase their sample size and therefore determine whether the spectral index distribution differences identified remain robust. This allows us to test not only the orientation properties for a higher redshift population, but also a larger range of radio brightness.

Mainly due to technical problems related to bad weather, not all the sources of the sample have been observed, but we have been able to collect radio flux densities for 16 (out of the 22) BAL QSOs, and 14 (out of 22) QSOs of the "comparison sample". Although not completed, these observed samples provide sufficient statistic to complete the planned studies. In Table 5.1 we list those 16 BAL QSOs and 14 QSOs for which the radio-data provide significant frequency coverage to analyze the full shape of the radio spectrum. In Table 5.2 we list the 7 BALs sources selected in our sample but eventually not observed. The two tables list various properties of the sample, including the AI of the BAL QSOs, the radio

power and the radio-loudness parameter R (see section 1.5.2 for definition). In Fig. 5.9 we show two plots of the sources in Table 5.1, respectively as magnitude and FIRST radio flux-peak versus redshift.

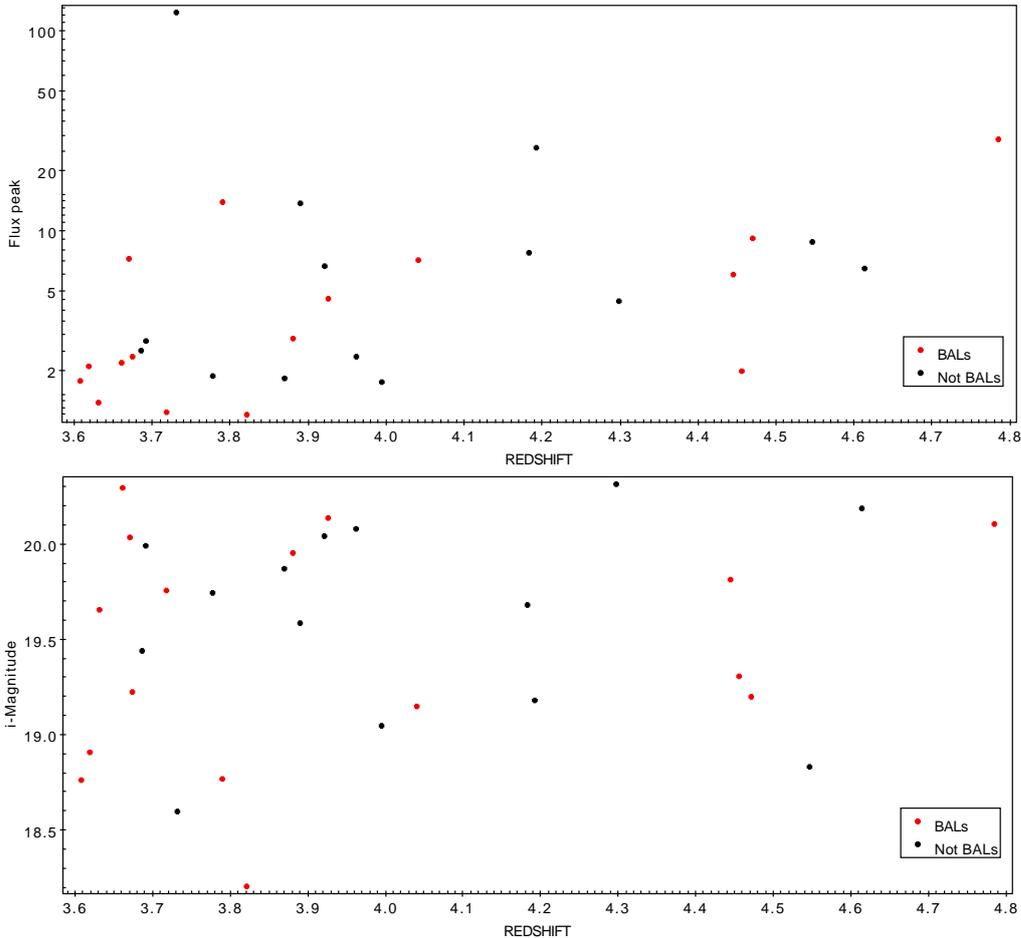


Figure 5.2: Plots of FIRST radio flux-peak versus redshift (upper plot) and SDSS dereddered i_{AB} magnitude versus redshift (lower plot) for the samples of 16 BAL (red dots) and 14 unabsorbed QSOs (blue dots) listed in table 5.1. For all these sources we obtained radio fluxes in multiple frequency bands at the 100-m Effelsberg Telescope and at the EVLA A radio-array.

5.5 Radio observations and data reduction

The observational campaign carried out for this work included both interferometric (JVLA) and single-dish (Effelsberg-100m) observations. In this section we

Table 5.1: BAL and comparison QSOs observed with the Effelsberg-100m single dish and/or the JVLA interferometer.

SDSS ID	ID	RA	DEC	i_{AB}	$S_{1.4\text{GHz}}$	z	$\log_{10} P_{1.4\text{GHz}}$	$\log_{10} R$	AI(*)	Note	
(1)	(2)	(3)	(J2000)	(4)	(mJy)	(5)	(6)	(7)	(8)	(9)	(10)
(BAL Sample)											
J000051.57+001202.5	0000+00	00:00:51.57	+00:12:02.5	19.95	2.64	4.00	26.13	2.10	3762.29	(a),(g)	
J074738.49+133747.3	0747+13	07:47:38.49	+13:37:47.3	19.15	6.62	4.04	26.56	2.13	5879.09		
J100645.59+462717.2	1006+46	10:06:45.59	+46:27:17.2	19.81	6.32	4.44	26.56	2.65	185.31		
J102343.13+553132.4	1023+55	10:23:43.13	+55:31:32.4	19.31	2.80	4.45	26.08	2.25	4820.07	(a),(g)	
J103601.03+500831.8	1036+50	10:36:01.03	+50:08:31.8	19.20	9.22	4.47	26.75	2.65	1737.88	(a),(g)	
J110946.44+190257.6	1109+19	11:09:46.44	+19:02:57.6	20.03	6.95	3.67	26.49	2.46	1678.28		
J111055.22+430510.1	1110+43	11:10:55.22	+43:05:10.1	18.21	1.17	3.82	25.77	1.09	193.72		
J112938.73+131232.3	1129+13	11:29:38.73	+13:12:32.3	18.76	1.33	3.61	25.86	1.22	206.46		
J113330.91+380638.2	1133+38	11:33:30.91	+38:06:38.2	19.66	0.81	3.63	25.77	1.39	176.29	(g)	
J115731.67+225726.4	1157+22	11:57:31.67	+22:57:26.4	20.14	3.81	3.92	26.34	2.32	979.66		
J133234.18+000921.7	1332+00	13:32:34.18	+00:09:21.7	20.30	1.49	3.66	25.97	1.95	149.60	(t)	
J134428.55+625608.2	1344+62	13:44:28.55	+62:56:08.2	19.23	2.58	3.67	26.00	1.77	4967.82	(t),(g)	
J134854.37+171149.6	1348+17	13:48:54.37	+17:11:49.6	18.91	1.89	3.62	25.94	1.52	356.61		
J150643.81+533134.5	1506+53	15:06:43.81	+53:31:34.5	18.77	14.63	3.79	26.80	2.34	166.89	(t)	
J151146.99+252424.3	1511+25	15:11:46.99	+25:24:24.3	19.76	1.39	3.72	25.73	1.72	3034.72	(a)	
J165913.23+210115.8	1659+21	16:59:13.23	+21:01:15.8	20.11	28.81	4.78	27.29	3.37	787.91		
(Comparison Sample)											
J030025.23+003224.2	0300+00	03:00:25.23	+00:32:24.2	19.68	7.69	4.18	26.62	2.31			
J083322.50+095941.2	0833+09	08:33:22.50	+09:59:41.2	18.60	125.76	3.73	27.73	3.18			
J084044.19+341101.6	0840+34	08:40:44.19	+34:11:01.6	19.58	13.59	3.89	26.81	2.64			
J090129.23+104240.4	0912+10	09:01:29.23	+10:42:40.4	20.08	2.14	3.96	26.06	2.01			
J091824.38+063653.4	0918+06	09:18:24.38	+06:36:53.4	19.18	26.50	4.19	27.15	2.88			
J101747.76+342737.9	1017+34	10:17:47.76	+34:27:37.9	19.99	2.64	3.69	26.08	2.02			
J110201.91+533912.6	1102+53	11:02:01.91	+53:39:12.6	20.31	5.57	4.30	26.40	2.51			
J112530.50+575722.7	1125+57	11:25:30.50	+57:57:22.7	19.44	2.99	3.68	26.03	1.85			
J115045.61+424001.1	1150+42	11:50:45.61	+42:40:01.1	19.87	1.51	3.87	25.93	1.72			
J124943.67+152707.1	1249+15	12:49:43.67	+15:27:07.1	19.05	2.01	3.99	25.94	1.61			
J131121.32+222738.7	1311+22	13:11:21.32	+22:27:38.7	20.19	6.53	4.61	26.62	2.84			
J142326.48+391226.3	1423+39	14:23:26.48	+39:12:26.3	20.04	6.51	3.92	26.50	2.46			
J144643.37+602714.4	1446+60	14:46:43.37	+60:27:14.4	19.74	1.80	3.78	25.93	1.77			
J161105.65+084435.4	1611+08	16:11:05.65	+08:44:35.4	18.84	8.82	4.54	26.74	2.31			

The columns give the following: (1) SDSS object-ID; (2) shortened name assigned to source for the radio observations ; (3) SDSS J2000 coordinates; (4) SDSS dereddened PSF i_{AB} magnitude; (5) FIRST peak radio flux density at 1.4GHz; (6) QSO redshift; (7) radio luminosity at 1.4GHz; (8) R-parameter of radio-loudness (Kellermann et al. 1989); (9) Modified absorption index AI(*) (Bruni et al. 2012); (10) note indicating if the QSO is classified as BAL in other catalogues, (a) Catalogue by Allen et al. (2011), (g) Catalogue by Gibson et al. (2009), (t) Trump et al. (2006).

Table 5.2: *BAL QSOs of our selected sample, for which the radio-observations did not provide enough frequency coverage to analyze the full shapes of the radio spectrum.*

SDSS ID	RA	DEC	i_{AB}	$S_{1.4\text{GHz}}$	z	$\log_{10}P_{1.4\text{GHz}}$	$\log_{10}R$	AI	Note
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
J094003.03+511602.73	09:40:03.03	+51:16:02.73	18.77	12.90	3.60	26.76	2.33	166.42	
J120447.15+330938.77	12:04:47.15	+33:09:38.77	18.38	1.10	3.61	25.58	1.25	6534.97	(a), (g)
J124658.83+120854.72	12:46:58.83	+12:08:54.72	19.86	1.07	3.80	25.82	1.76	1519.70	(a), (g)
J130348.94+002010.51	13:03:48.94	+00:20:10.51	18.66	1.08	3.65	25.62	1.15	1032.38	(a), (g)
J135554.56+450421.09	13:55:54.56	+45:04:21.09	19.31	1.48	4.09	26.03	1.62	330.64	(g)
J161716.49+250208.17	16:17:16.49	+25:02:08.17	19.83	1.01	3.94	26.06	1.95	4439.58	(a), (g)

The columns give the following: (1)SDSS object-ID; (2) SDSS J2000 coordinates; (3) SDSS dereddened PSF i -magnitude; (4) FIRST peak radio flux density; (5) QSO redshift; (6) radio luminosity at 1.4GHz; (7) R -parameter of radio-loudness (Kellermann et al. 1989); (8) Modified absorption index AI(*) (Bruni et al. 2012); (9) note indicating if the QSO is classified as BAL in other catalogues, (a) Catalogue by Allen et al. (2011), (g) Catalogue by Gibson et al. (2009), (t) Trump et al. (2006).

Table 5.3: *Observing frequencies, bandwidth and beam sizes (half-power beam-width) for the sample of BAL and matched comparison QSOs listed in Table 5.1*

Telescope	Frequency (GHz)	Bandwidth (MHz)	θ_{HPBW} (arcsec)
Effelsberg-100m	2.64	80	265
Effelsberg-100m	4.85	500	145
Effelsberg-100m	8.35	1100	80
JVLA(A)	1.5	1024	1.3
JVLA(A)	5.5	2048	0.33
JVLA(A)	9.0	2048	0.20

give details about the observing setup and strategy, as well as the data reduction process. In Table 5.3 we give a summary of the observational setups, including the observing frequencies, bandwidth and the beam sizes.

5.5.1 JVLA

During November 2012, we performed observations at the Jansky Very Large Array observations at 1.5, 5.5, and 9 GHz, with the aim of adding high-sensitivity flux-density measurements, and obtain morphological information. All observations were performed in A-configuration, and in dynamic mode, with 1 or 1.5 hours slots. Phase referencing was applied to all sources, and standard flux-density calibrators were observed at least once for every slot. A typical RMS

lower than 0.5 mJy was obtained. Data were reduced with the latest stable version of the CASA¹ software (4.1.0), making use of customized reduction scripts ran on the MPIfR High-Performance Computer cluster (HPC). Flux densities were extracted via bi-dimensional Gaussian fit on the produced maps. Errors were calculated assuming a $\sim 10\%$ uncertainty for the absolute flux-density calibration and adding quadratically the RMS of the map.

5.5.2 Effelsberg-100m telescope

Observations with the Effelsberg-100m single dish were performed in different runs, initially as a granted-time project in September 2012, and later continued as a filler project, to improve frequency coverage, until May 2015. The cross-scan observing mode was used, at three different frequencies (2.6, 4.8, and 8.3 GHz), increasing the number of sub-scan repetition depending on the source faintness. Re-pointing on sources near to the target was performed every time the telescope significantly changed the elevation, to avoid pointing problems due to gravitational deformations of the antenna. Data were reduced with the latest version of the TOOLBOX² reduction package. Flux density scale was calibrated on well-known sources observed every ~ 4 hours (3C286, 3C48, 3C295) using the flux-densities from Baars et al. (1977). Flux densities were extracted via Gaussian fit of the cross-scans. Errors were calculated, as for the JVLA, assuming a $\sim 10\%$ of uncertainty for the absolute flux-density calibration and adding via quadratic sum the cross-scan RMS. Three sigma upper limits are given for non-detections.

5.6 Observational results

The collected flux densities are reported in Tab. 5.4 and 5.5. Considering the whole sample of 30 BAL and non BAL sources listed in Table 5.1, we have that 20 of them were observed with the JVLA, and 27 with the Effelsberg-100m. When no detection at 1.5 GHz was available from our campaign, we used the measurement at 1.4 GHz from the FIRST survey.

¹<http://casa.nrao.edu/index.shtml>

²<https://eff100mwiki.mpifr-bonn.mpg.de/doku.php>

Table 5.4: Flux densities for the 16 high-redshift BAL QSOs. Measurements at 1.5, 5.5, 9.0 GHz are from the JVLA, while at 4.8 and 8.3 GHz are from the Effelsberg-100m single dish. Asterisked values at 1.5 GHz comes from the FIRST survey. For one object (1133+38) we could find a measurement in the Gibson et al. (2009) catalogue at 4.8 GHz (asterisked). In the last two columns, the spectral index and HFP classification are given.

Source	$S_{1.5\text{GHz}}$ (mJy)	$S_{2.6\text{GHz}}$ (mJy)	$S_{4.8\text{GHz}}$ (mJy)	$S_{5.5\text{GHz}}$ (mJy)	$S_{8.3\text{GHz}}$ (mJy)	$S_{9.0\text{GHz}}$ (mJy)	α	HFP
0000+00	1.8±0.2	<11.1	-	0.7±0.1	-	0.7±0.1	0.00±0.33	
0747+13	8.0±0.8*	11.8±3.9	13.1±2.2	-	5.5±0.7	-	-1.58±0.34	✓
1006+46	6.3±0.6*	11.8±1.3	7.6±0.9	-	4.6±0.8	-	-0.92±0.34	✓
1023+55	2.8±0.3*	<23.1	3.7±0.5	1.9±0.2	-	1.6±0.2	-0.35±0.27	
1036+50	9.2±1.0*	7.4±1.0	4.6±0.5	-	2.9±0.5	-	-0.84±0.33	
1109+19	6.4±0.7	5.6±0.8	-	2.8±0.8	-	1.6±0.2	-1.14±0.51	
1110+43	1.2±0.2*	-	3.2±0.5	-	3.3±0.5	-	0.06±0.35	✓
1129+13	2.8±0.3	-	-	0.8±0.1	-	0.5±0.1	-0.95±0.38	
1133+38	0.8±0.2*	43.3±4.5	27.0±0.4*	-	<8.4	-	-0.77±0.17	✓
1157+22	2.8±0.3	45.8±5.2	-	2.5±0.2	-	1.3±0.1	-1.33±0.18	✓
1332+00	2.2±0.2	-	-	0.7±0.1	-	0.5±0.1	-0.68±0.40	
1344+62	2.2±0.2	<5.9	-	0.8±0.1	-	-	-0.78±0.25	
1348+17	2.3±0.2	3.7±0.7	-	3.1±0.3	-	2.6±0.3	-0.36±0.25	✓
1506+53	14.6±1.5*	14.2±2.5	-	-	-	-	-0.04±0.33	
1511+25	1.2±0.2*	-	<12.0	-	7.9±1.9	-	1.06±0.48	✓
1659+21	27.2±2.7	19.8±2.4	16.0±3.8	11.1±1.1	-	7.3±0.7	-0.85±0.22	

Table 5.5: Flux densities for the 14 high-redshift non-BAL QSOs. Measurements at 1.5, 5.5, 9.0 GHz are from the JVLA, while at 4.8 and 8.3 GHz are from the Effelsberg-100m single dish. Asterisked values at 1.5 GHz comes from the FIRST survey. In the last two columns, the spectral index and HFP classification are given.

Source	$S_{1.5\text{GHz}}$ (mJy)	$S_{2.6\text{GHz}}$ (mJy)	$S_{4.8\text{GHz}}$ (mJy)	$S_{5.5\text{GHz}}$ (mJy)	$S_{8.3\text{GHz}}$ (mJy)	$S_{9.0\text{GHz}}$ (mJy)	α	HFP
0300+00	6.8 ± 0.7	5.6 ± 0.8	-	3.8 ± 0.4	-	1.8 ± 0.2	-1.52 ± 0.25	
0833+09	$7.7\pm 0.8^*$	106 ± 11	87.9 ± 9.4	-	77.3 ± 13.5	-	-0.23 ± 0.33	
0840+34	$126\pm 12^*$	14.8 ± 1.6	31.0 ± 5.8	-	7.9 ± 0.9	-	-2.50 ± 0.36	✓
0901+10	1.8 ± 0.2	-	-	0.7 ± 0.1	-	0.5 ± 0.1	-0.68 ± 0.40	
0918+06	$26.5\pm 2.6^*$	42.7 ± 4.3	35.6 ± 3.9	-	23.1 ± 4.0	-	-0.79 ± 0.33	✓
1017+34	$2.6\pm 0.3^*$	7.4 ± 1.4	10.5 ± 1.3	-	-	-	0.57 ± 0.37	✓
1102+53	$5.6\pm 0.6^*$	6.2 ± 0.6	6.2 ± 1.0	-	6.3 ± 1.4	-	0.03 ± 0.45	✓
1125+57	2.5 ± 0.3	-	-	0.7 ± 0.1	-	0.5 ± 0.1	-0.68 ± 0.40	
1150+42	2.9 ± 0.3	-	-	0.5 ± 0.1	-	<0.3	-1.35 ± 0.37	
1249+15	2.0 ± 0.2	-	-	0.6 ± 0.1	-	0.4 ± 0.1	-0.82 ± 0.49	
1311+22	6.0 ± 0.6	6.8 ± 2.0	-	-	-	1.4 ± 0.1	-1.27 ± 0.49	✓
1423+39	8.6 ± 0.9	<21.0	-	10.0 ± 1.0	-	7.0 ± 0.7	-0.72 ± 0.23	✓
1446+60	2.8 ± 0.3	21.6 ± 2.5	-	1.3 ± 0.1	-	0.7 ± 0.1	-1.26 ± 0.26	✓
1611+08	22.5 ± 2.4	-	-	14.2 ± 1.4	-	11.4 ± 1.1	-0.45 ± 0.23	

5.6.1 Morphology

All the observed sources resulted to be unresolved both with the Effelsberg-100m and the JVLA in A configuration. Considering the angular resolution of the latter for the 8 BAL and 8 non-BAL QSOs detected at 9 GHz (0.2 arcsec), we obtain an upper limit for the projected linear size of 1.4 kpc at the mean redshift of 4.02 and 3.91 for our BAL and non-BAL samples. This is compatible with the typical linear size of High Frequency Peakers sources (HFP, 0.01-0.5 kpc, Dallacasa et al. 2000) and also GPS sources (0.5-5 kpc). Further continental interferometry observations would be required to resolve sources at these redshifts.

5.7 SEDs shape and spectral index

In the following we analyze the SEDs in order to get information about the orientation and age of our high-redshift sample.

5.7.1 Synchrotron peak frequencies

The collected flux densities between 1.5 and 9 GHz allow us to reconstruct the SED of our objects in a reasonable way. For most of the sources more than 3 measurements are available, while for the remaining sources we took advantage of the FIRST data to obtain at least 2 measurement. In Fig. 5.3 and 5.4 we present the SEDs for each source, as connected dots. While an exhaustive analysis via SED fitting is not always possible, we can nevertheless get an indication of the presence of a peak in our frequency window, in order to obtain the fraction of young radio sources in our samples. GPS sources at lower redshifts are usually identified as objects peaking in the range $1 \text{ GHz} \leq \nu_{peak} \leq 8 \text{ GHz}$ (Gopal-Krishna et al. 1983; O’Dea et al. 1991) while HFP can peak at frequencies $> 8 \text{ GHz}$ (Dallacasa et al. 2000). Both are generally interpreted as young radio sources, in a evolutionary track towards CSS sources, peaking at hundred of MHz (Saikia & Salter 1988; Fanti et al. 1990), as the emitting plasma adiabatically expands. At the redshift range of our sample (mean redshift ~ 4), the peaking interval of GPS and HFP translates into an interval between 0.2 GHz and 1.6 GHz for the first (outside our observing window) and $> 1.6 \text{ GHz}$ for the latter. Thus, with our data we can only identify the peak of the most young class of radio objects. In our approach, we consider a source peaked when we can see an inverted SED - with flux increasing from the lower to highest frequency - or when a clear maximum in flux density is present among datapoint at the extremes of the interval. Generally, when only two datapoint are present we cannot assess the presence of a peak (the risk is that a peak in between is not detected and source is considered flat or, even worse,

steep). Among our objects, only a few have less than three detections (4 BAL and 1 non-BAL QSOs) but upper limits or vicinity of data points allow to rule out the presence of a peak, except for BAL QSO 1344+62 and non-BAL QSO 1150+42. For these cases we can take advantage of the broad bandwidth offered by the JVLA observations at 1.5 GHz (1024 MHz) to have in-band flux density measurements that constrain better the SED behavior. At higher frequencies signal to noise is not enough to do the same. As shown in their correspondent plots, the datapoint obtained dividing the band into four 250-MHz intervals (circles) clearly suggest a steepening of the SED around 1.5 GHz, ruling out the presence of a peak in-between 1.5 GHz and 5.5 GHz in both cases.

In total, 7 out of 16 BAL QSOs and 7 out of 14 non-BAL QSOs show hints of a peak between 1 and 10 GHz, not suggesting a difference between the two samples. Considering the associated poissonian error (\sqrt{N} , where N is the number of peaking sources), we obtain a HFP fraction of $44\pm 16\%$ and $50\pm 19\%$ for BAL and non-BAL QSOs, respectively. This does not suggest a particular young radio phase for BAL QSOs with respect to ‘normal’ QSOs, even at this redshift range.

5.7.2 Spectral indices

The spectral index (α) of the optically-thin part of the synchrotron spectrum can be a useful orientation indicator (Orr & Browne 1982). This has been used in previous works to characterise the orientation angle distribution of RL BAL QSOs (Montenegro-Montes et al. 2008; DiPompeo et al. 2011; Bruni et al. 2012) finding only a mildly preferred orientation with respect to non-BAL QSOs. Here we repeat the same analysis, to test the orientation scenario also for high redshifts. In order to avoid spectral index calculation across the peak, we consider for each source the two measurements at the two highest frequencies available.

Only sources presenting a negative spectral index can be used to test orientation: these are 13 BAL and 12 non-BAL QSOs. Among these, 9/13 BAL QSOs ($70\pm 23\%$) are steep ($\alpha < -0.5$), also considering the error, while 5/12 non-BAL QSOs ($42\pm 19\%$) are steep. This suggests a preferred equatorial orientation for BAL QSOs, with a discrepancy more significant than the one found in previous works at lower redshifts ($67\pm 17\%$ vs $50\pm 14\%$ from Bruni et al. 2012). This result is in line with the findings by Bruni et al. (2012), i.e. BAL QSOs would be more easily found among sources in a young or restarting radio phase. At this redshift, given the higher rest-frame frequency range explored, this scenario can be tested in younger radio sources (e.g. HFP), where jet has just been collimated and launched. A possible presence of BAL-producing outflows in this kind of objects seems more probable in the light of this work. Later these could be recollimated to form the radio jet, as also proposed by Elvis (2000).

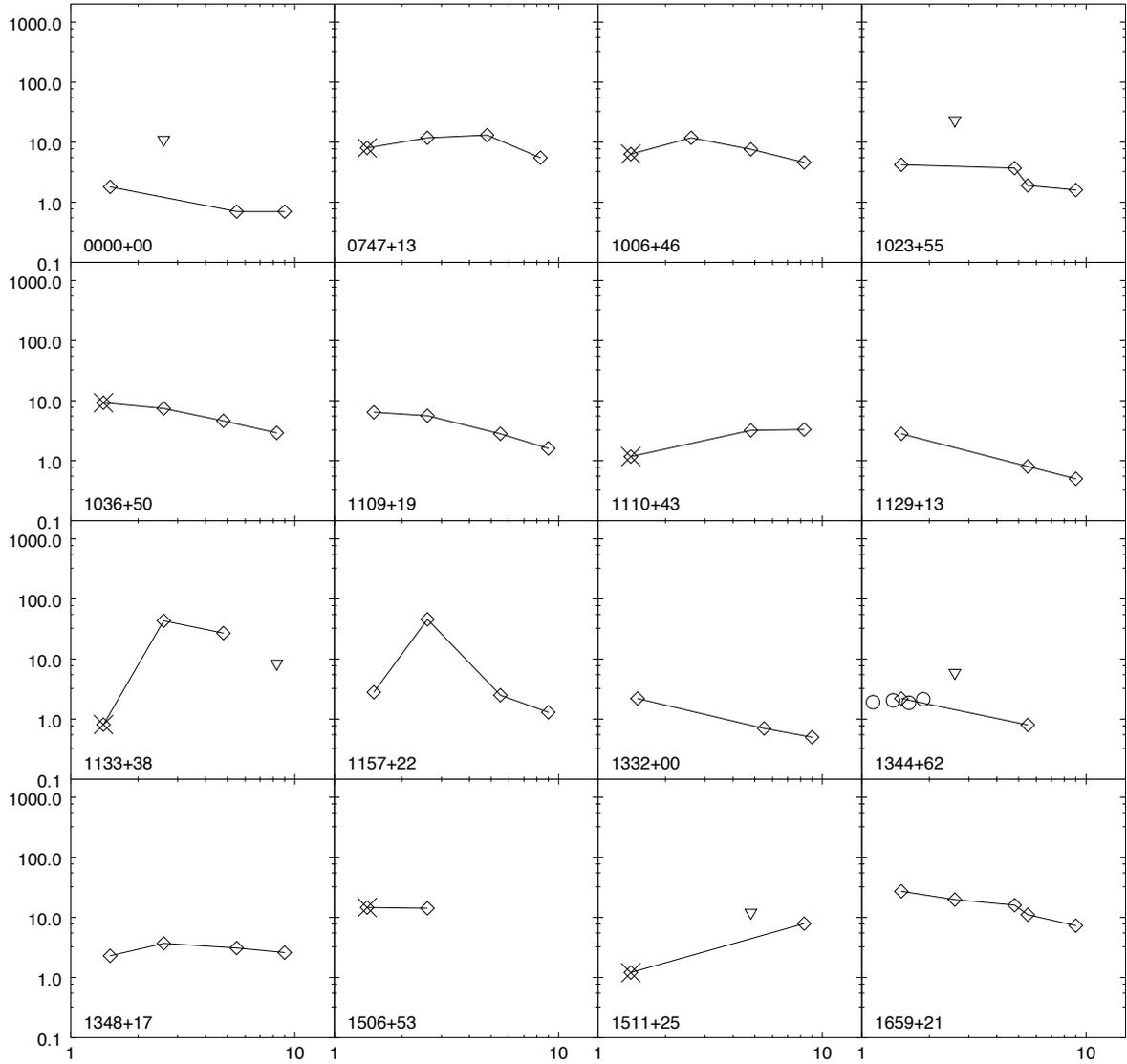


Figure 5.3: SEDs of the 16 high-redshift BAL QSOs studied here (x-axis: GHz, y-axis: mJy). Flux densities from the FIRST catalogue, at 1.4 GHz, are plotted as crosses. Circles indicate the in-band flux densities at 1.5 GHz from the JVLA.

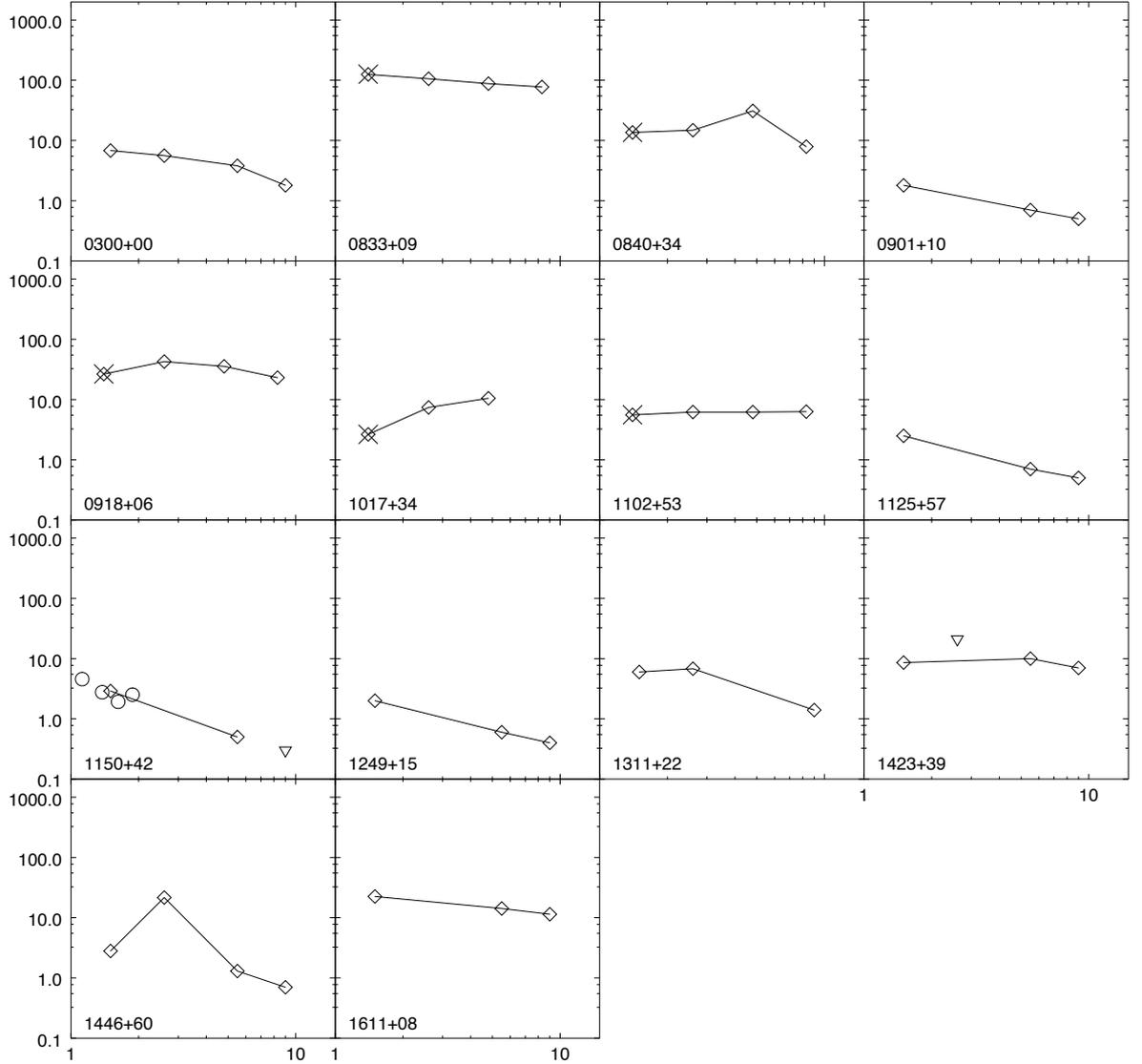


Figure 5.4: SEDs of the 14 comparison non-BAL QSOs (x -axis: GHz, y -axis: mJy). Flux densities from the FIRST catalogue, at 1.4 GHz, are plotted as crosses. Circles indicate the in-band flux densities at 1.5 GHz from the JVLA.

5.8 Characterization optical/uv

Numerous studies on the continuum and emission-line properties of BAL QSOs spectra have been pursued in the last 25 years, with the aim to establish if and how the broad absorption line troughs are correlated with the physical mechanisms that produce the emission in quasars. The differences that have been found in those properties are not consistent with the existence of two parent populations for BAL and not BAL quasars. Nevertheless BAL QSOs tend to differ statistically in some average continuum and emission line properties in the optical and in the near-infrared. In particular it was early noticed (Weymann et al. 1991) that they show redder continua than those of normal quasars. This early claim has been confirmed by a number of different studies (Sprayberry & Foltz 1992; Brotherton et al. 2001; Reichard et al. 2003), that quantify the difference between the continua of the two populations assuming a Small Magellanic Cloud (SMC) extinction law. The exact value of the SMC dust-correction $E(B - V)$ varies from author to author depending on the criteria used to select the BALs of the samples, but there is a general agreement on the fact that the subpopulation of LoBAL (Low-ionization BAL quasars possess Mg II and/or Al III absorption lines, in addition to the high-ionization transitions) are significantly redder than HiBALs (High-ionization BAL quasars show absorption lines in high-ionization transitions including Si IV, and C IV), and that HiBAL are moderately redder than quasars not showing BAL features³. The fact that the reddening is consistent with SMC-like dust does not unveil the origin of the reddening itself, that is still subject of debate and speculation nowadays (see for instance Krawczyk et al. 2015).

The fact that BAL QSOs appear redder than average, posed the question that optical surveys like SDSS, may miss many of these objects because their selection functions are studied to target normal quasars that are intrinsically bluer. This effect, that is well studied in Allen et al. (2011), seems confirmed in Urrutia et al. (2009) that find an unusually high percentage of BAL QSOs in the sample of extremely red quasars that they selected from a cross matched sample of FIRST, 2MASS and SDSS sources. About the 60% of their BALs are not targeted as quasar-candidates by SDSS. At the same time it is worth notice that almost all (except one) of the BAL QSOs of the Urrutia et al. (2009) sample belong to the class of LoBAL.

In this section we use our sample of RL high- z BAL QSOs to compare their broadband optical colors with the remaining population of quasar. For this purpose we use all the 22 BAL QSOs selected in section 2 and listed in Table 5.1 and 5.2. We compare them with the entire sample of RL QSOs within $3.6 \leq z \leq 4.8$

³For instance, in Brotherton et al. (2001), they find $E(B - V) \sim 0.04$ for HiBAL and $E(B - V) \sim 0.1$ for LoBAL

from which we have extracted the 22 BALs. This comparison sample consists of 100 QSOs from the SDSS DR7 QSO catalogue and it is complemented by another 14 QSOs selected with our NN selection-strategies (Tuccillo et al. 2015), for a total of 114 RL QSOs. We note that we excluded from the comparison sample 5 QSOs showing broad absorption in the C IV) emission-line only (but not in C IV) in order to reduce any bias in the comparison between BALs and non-BALs quasars.

The distribution of the radio power, the i -magnitude and the radio-loudness R , for the two samples of 22 BAL and the 114 non-BAL samples is showed in Fig. 6.34.

5.8.1 Broadband optical colors

The continuum emission differences between BALQSOs and non-BALQSOs can be investigated studying the differences between their broadband colors, since they can give an indication of the photometric spectral index.

In Fig. 5.6 we show SDSS color-color diagrams of our samples of BAL and non-BAL QSO. All the colors have been dereddened using the reddening maps of Schlegel et al. (1998). BALQSOs tend to occupy redder sub-regions of the color-color space determined by the distribution of the parent population of quasars. The same trend is analyzed in Fig. 5.7, where we plot the normalized distribution of the colors for BAL and non-BAL QSOs. In the latter plots, the color-distributions for BALQSO clearly peak toward redder colors than the peak of non-BAL QSOs. T-test are conducted to compare the means of all the SDSS colors for the BALs and non BALs samples. In Table 5.6 we report only the colors for which the t-test reject the null-hypothesis that the color-vectors come from normal-distributed samples having equal means, and therefore suggest significant difference in those values. The t-test is performed in the general conditions of not assuming equal variances (Behrens-Fisher problem) and the number of degrees of freedom df is given by the Satterthwaite approximation.

Such differences in the mean colors of BAL and non-BAL QSOs should be more noticeable in the infrared where dust is seen in emission rather than in absorption. Cross matching the two samples with data from UKIDSS-LAS [REF], we find that 5 BAL and 49 non BAL are detected. The plot of the UKIDSS colors for these two small samples suggests that also in the range of the NIR covered by UKIDSS, the BALs tend to be redder than normal quasars.

5.8.2 RL and RQ BALs

Radio selection may be introducing a bias on the colors of the sample, as it has been showed that quasars with higher radio luminosity have a tendency to show

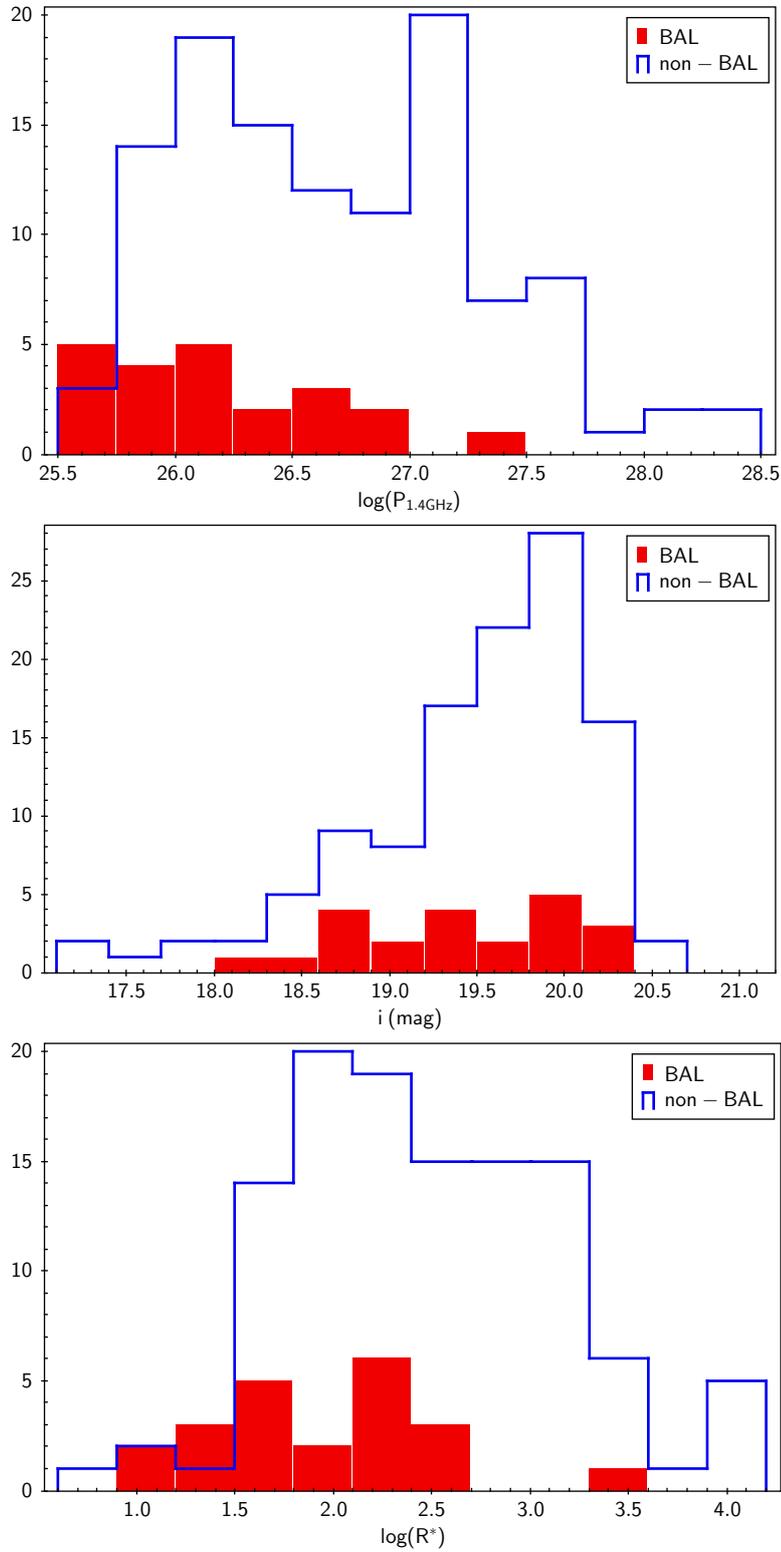


Figure 5.5: Distribution of radio power, SDSS i -magnitude, and radio-loudness R^* for the sample of 22 BALs and 114 non-BALs RL QSOs used in this section.

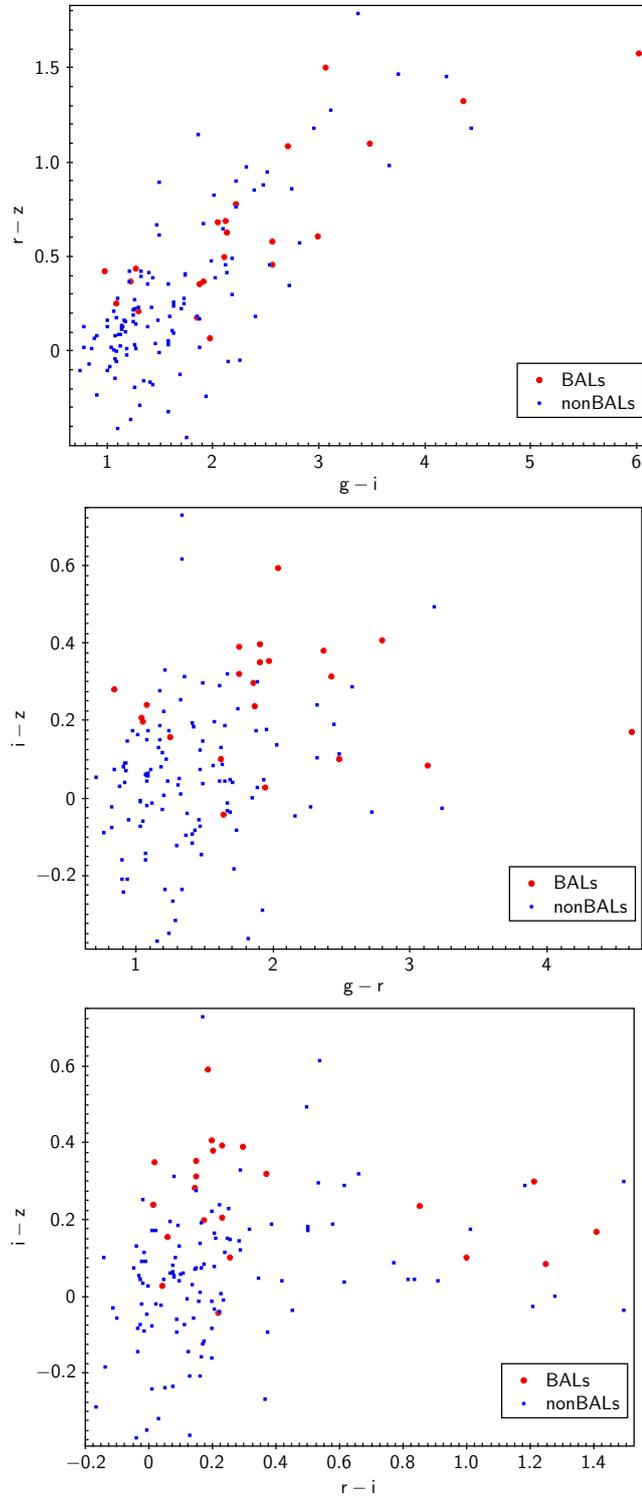


Figure 5.6: Colors of 114 non-BALs (blue squares) and 22 BAL (red points) in SDSS color-color space for RL QSOs with $3.6 \leq z \leq 4.8$. BALQSOs tend to occupy redder sub-regions of the color-color space determined by the distribution of the parent population of quasars (see section 5.8.1)

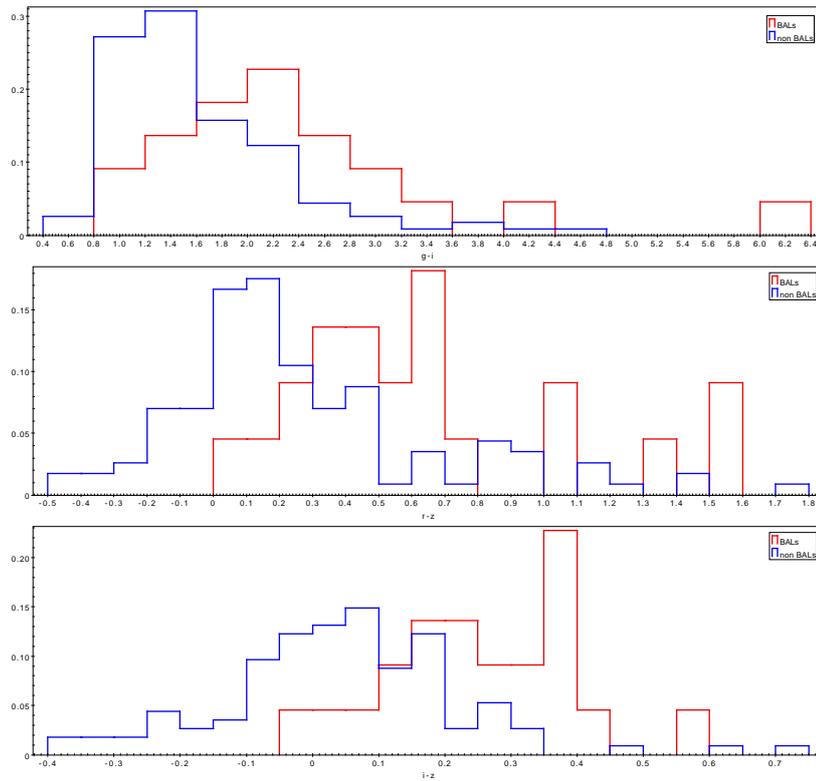


Figure 5.7: Normalized distribution of SDSS colors for our sample of 22 RL BAL and 114 non-BAL QSOs with $3.6 \leq z \leq 4.8$. The peak of the distribution for the BAL population is generally shifted toward redder colors (see section 5.8.1)

Table 5.6: *t*-test on SDSS colors for our sample of 22 RL BAL and 114 non-BAL QSOS with $3.6 \leq z \leq 4.8$

Color (1)	Population (2)	Mean (3)	σ (4)	Median (5)	<i>t</i> (6)	df (7)	<i>p</i> ($\times 10^{-2}$) (8)
(u-i)	BAL	5.13	1.02	5.21	2.3	32.2	3.00
	non BAL	4.58	1.14	4.40			
(u-z)	BAL	5.38	1.04	5.47	3.0	32.1	0.46
	non BAL	4.62	1.17	4.48			
(g-r)	BAL	1.96	0.82	1.87	3.0	23.8	0.55
	non BAL	1.41	0.48	1.31			
(g-i)	BAL	2.36	1.15	2.11	2.8	24.1	1.00
	non BAL	1.64	0.71	1.43			
(g-z)	BAL	2.61	1.15	2.45	3.6	24.8	0.14
	non BAL	1.69	0.77	1.46			
(r-z)	BAL	0.64	0.42	0.54	3.7	29.3	0.08
	non BAL	0.28	0.41	0.18			
(i-z)	BAL	0.25	0.15	0.26	5.75	34.3	0.02×10^{-2}
	non BAL	0.05	0.18	0.05			

The columns give the following: (1) SDSS color; (2) subsample considered; (3) mean of the color; (4) standard deviation; (5) median of the color; (6) statistic *t*; (7) associated degrees of freedom to the *t*-test; (8) the statistic *p* associated to the *t*-test, the significance level α to reject the null-hypothesis, as usually is 5×10^{-2} .

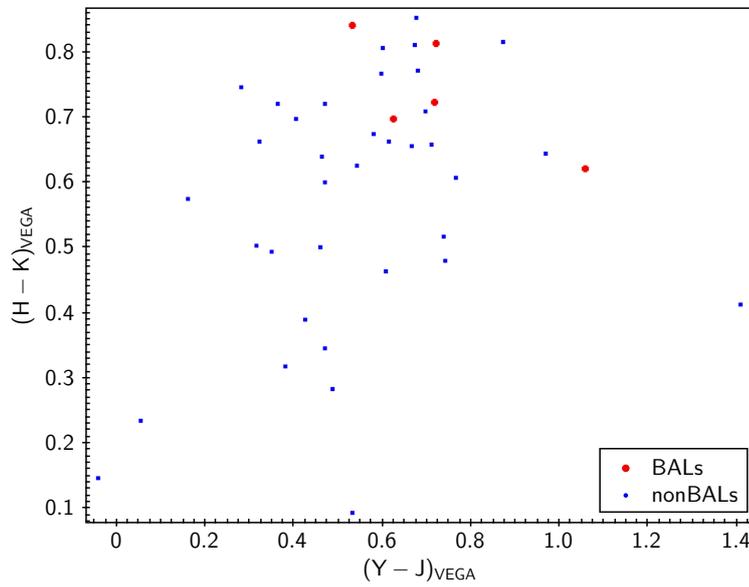


Figure 5.8: Cross matching the two samples of 22 RL BAL and 114 non-BAL QSOs with $3.6 \leq z \leq 4.8$ with data from UKIDSS -LAS, we find that 5 BAL and 49 non BAL are detected. The plot of the UKIDSS colors for these two small sample suggest that also in the range of the NIR covered by UKIDSS, the BALs tend to be redder than normal quasars.

redder colors than quasars with very low radio luminosity (White et al. 2007). Therefore we compare the colors of the RL BAL QSOs of our sample with RQ BALs in the same range of redshift, using the Gibson et al. (2009) catalogue. In the upper plot of Fig. 5.9 we compare the colors of the 10 RL BAL of our sample listed also in the Gibson et al. (2009) catalogue with all the remaining 347 RQ BAL of that catalogue and in the same range of redshift. In this way we compare BAL QSOs classified using the same balnicity index. In the bottom plot of Fig. 5.9, we plot all the BAL QSOs of our sample against the same RQ BALs used previously. In both cases no differences/trends in colors are found and RL BALs do not appear to have average different colors than RQ BALs.

5.8.3 Discussion

The sample of BAL and BAL-QSOs that we used in this section have all been selected by the SDSS target algorithm or by a neural-network trained with quasars of the SDSS QSOs catalogue. The SDSS algorithm use a color-selection criteria based mainly on normal quasars and an independent criterion based on matching stellar SDSS sources to FIRST radio sources: all sources having $15.0 < i < 19.1$ and a FIRST match within $2''0$ are selected for spectroscopic follow-up. However, the bright magnitude limit adopted for this second criteria, make that most of the high-z quasars are identified for their colors only. As such, we can not expect that our BAL QSOs have color very different from those of non-BAL quasars. Extremely red-objects, like the ones selected in Urrutia et al. (2009), may be selected using different selection functions or techniques.

Nevertheless, many of the SDSS colors of our BAL sample, particularly in $i - z$ and $r - z$, is shifted toward redder colors. The $r - z$ difference of the medians is 0.36 mag and the difference in the $i - z$ median is 0.21 mag. This result can not be caused simply by absorption from the BAL troughs themselves, since the trough absorption can make the broadband color of BALQSOs bluer as well as redder, depending on where the redshift of the quasar places the troughs with respect to the filters. Instead, it gives an indication of an overall flux deficit. We obtain this result using our sample of high-z RL BAL, but we did not find differences in the mean color of our RL BAL with the ones of the RQ BAL in the same range of redshift. Therefore we do not find indications of differences, in this range of redshift, between the continua of these two populations.

5.9 Fraction

To determine the true BALQSO fraction, we should investigate carefully the completeness of the SDSS BALQSO sample and apply corrections to our observed

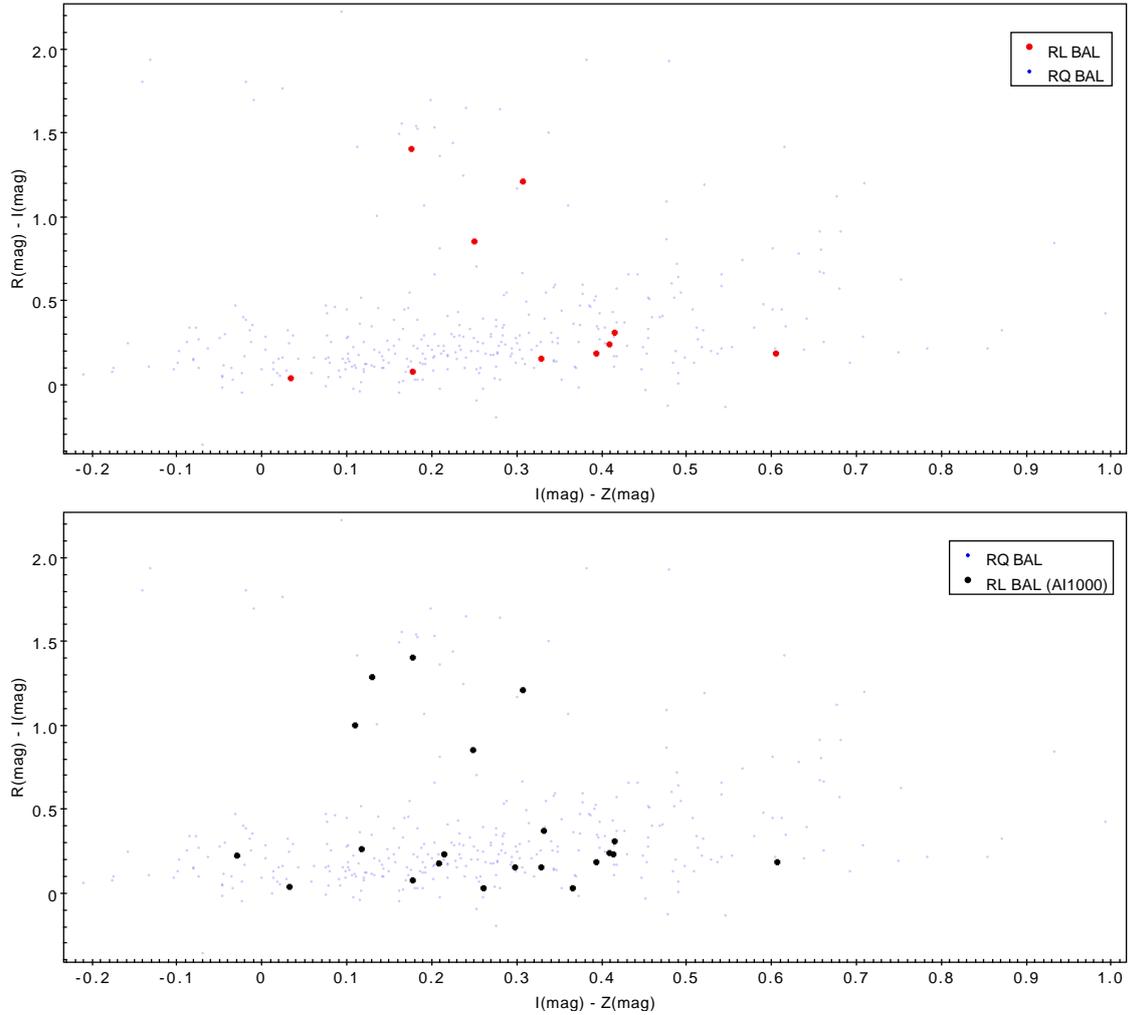


Figure 5.9: Here we compare the colors of the RL and RQ BALs to see if there are differences. For the RQ BALs sample we use the ones from the Gibson et al. (2009) catalogue and within our range of redshift. In the upper plot we compare them with the 10 RL BAL of our sample also present in the Gibson et al. (2009) catalogue (therefore using homogeneous BI criteria). In the second figure, we plot all the 22 BALs of our sample against the RQ BALs of the Gibson et al. (2009) catalogue. In both cases we do not see systematic differences/trends in colors

fraction. Since the SDSS QSOs are selected using an i -band magnitude limit, it should be necessary to correct the BALQSO fraction for absorption by the BAL trough itself in the band that was used to select quasars. Also, if BAL QSOs are generally redder than normal QSOs, a significant fraction of them could lie off the color-selection window usually used to detect quasars. Nevertheless the non-corrected percentage of the RL BAL of our sample on the unabsorbed RL QSOs in the same range of redshift is 22 on $114 + 22 + 5$, i.e. $\sim 15.6\%$. This percentage has to be meant as a lower limit on the real percentage of BAL QSOs. Allen et al. (2011) estimate the C IV BALQSO fraction within the SDSS spectroscopic survey is 14.0 ± 1.6 per cent.

5.10 Conclusions

In this chapter we study a sample of 22 RL BAL QSOs in the redshift range of $3.6 \leq z \leq 4.8$, comparing their properties in the radio and in the optical with a parent population of RL QSOs in the same range of redshift. The conclusions of this study can be summarized as:

- All sources are unresolved, also when observed with the JVLA at 9 GHz (8 BAL vs 8 non-BAL QSOs). This translates into an upper limit for the projected linear size of 1.4 kpc, compatible with GPS-HFP sources.
- We compared the peak frequencies for the BAL and non-BAL QSOs samples, not finding a predominance of HFP in BALs. This does not suggest a particular younger radio phase for BAL QSOs with respect to non-BAL objects, even at this redshift range. Nevertheless half of both samples can be classified as HFP, that are rarely found in CSS and GPS samples (Dallacasa et al. 2000).
- We derived the spectral index for the two samples, finding that a preferred equatorial orientation is present among BAL QSOs, with a more significant discrepancy with respect to previous studies at lower redshifts. This could indicate that, among young radio sources, the outflows responsible for the BAL features can have larger angles with respect to the jet axis than in older radio sources.
- Although our BALQSOs and non-BALQSOs are selected from the SDSS QSO catalogue, using mainly the optical/UV part of the spectrum, the SDSS selection algorithm is sensitive to a wide range of optical/UV properties and we do find differences in the mean colors of the two populations. We find that the RL BAL of our sample tend to be redder in average than the unabsorbed QSOs. If this is due to gas rich systems fueling the central AGN,

such differences should be more noticeable at other wavelengths, particularly the IR and submillimeter (where dust is seen in emission rather than absorption); however submillimeter observations show no significant differences between BALQSOs and non-BALQSOs (Bruni et al. 2015). Those circumstances could indicate that the reddening is mainly caused by orientation rather than for dust host-environment. However, this orientation effect need not be specific to the orientation of the accretion disk (i.e., external orientation), but rather it could be related to the opening angle of a disk wind (Elvis 2000), where some other quasar property is causing the opening angle of the disk wind to change, thus producing an orientation type effect (i.e., internal orientation).

Chapter 6

Selection of Candidates high-z RL Red QSOs

6.1 Introduction

Over the last decades, observations have established that super massive black holes (SMBHs) likely reside in the centers of all galaxies with spheroids (e.g., Kormendy & Richstone 1995; Richstone et al. 1998; Kormendy & Gebhardt 2001). Studies on the mechanism underlying the formation of these SMBHs and their role in the evolution of their host galaxy have therefore been pursued with increasing interest, both on the observational and on the theoretical side. In the context of these studies, the presence of Active Galactic Nucleus (AGN) found in Ultra Luminous Infrared Galaxies (ULIGs, Sanders et al. 1988) have been explained with models relating the co-evolution of galaxies and SMBHs through episodes of major mergers. In these scenarios the merging allows that large amounts of gas can settle into dense clouds which produce extensive starbursts while other gas continues to the center to fuel the accretion of any existing central black hole (Sanders et al. 1988; Barnes & Hernquist 1996). Such scenario is confirmed observationally by the large percentage of ULIGs showing merging and interaction (Sanders & Mirabel 1996). On the other side, luminous blue QSOs does not show sign of merging, and their host galaxies are mostly undisturbed (Dunlop et al. 2003). The possible link between the AGNs hosted in the ULIGs and the blue-luminous QSOs could be a population of "red QSOs", whose reddening is not explained in terms of orientation of the QSO. These red quasars are suggested to represent a relatively short evolutionary-phase in which the dust, produced by the rapid evolution of massive stars, obscures the young quasar resulting from the large funneling of the SMBH. The large intrinsic AGN luminosity is then reradiated in the infrared, while it is dimmed in the UV and in the optical. When

the radiated power of the accretion disk increases, the surrounding dust is blown away, and the red-quasar evolves into an unobscured -luminous quasar (Hopkins et al. 2006; Narayanan et al. 2010).

Since the larger part of the known QSOs are selected in the optical band, these sources have been quite elusive, and the number of known red QSOs is very small compared with the population of blue QSOs. Small populations of red QSOs have been found using selection methods based in wave-bands less sensitive to dust extinction, like in the radio (Gregg et al. 2002; White et al. 2003) and the infrared. Using 2MASS (the Two Micron All-Sky Survey) and FIRST (Faint Images of the Radio Sky at Twenty-cm), Glikman et al. (2012) selected and studied the reddening properties of a significant sample of red QSOs, demonstrating the existence of a population of type-1 QSOs whose physical mechanism for the reddening appears to differ from an orientation-based extinction.

Nevertheless, the near-infrared surveys like 2MASS were too shallow to select large samples of red QSOs, and deeper surveys did not cover enough area. The search for QSOs has found impetus with the new epoch of wide near and mid infrared surveys as UKIDSS and WISE (Wide-Field Infrared Survey Explorer). The use of near infrared data from these surveys has been proved to be effective in selecting extremely red QSOs (Banerji et al. 2012), finding objects considerably fainter and redder than the ones selected with 2MASS.

In this chapter we present a catalogue of sources detected in the radio (FIRST) and in the infrared (UKIDSS) but undetected in the optical (SDSS), and a list of selected red-quasars candidates at high- z . The first part of the work consisted in the methods developed to "clean" the sample from false sources and to associate to each final object its probability to be a real radio-infrared object with optical luminosity under the SDSS detection flux-limit. The second part of the work consist in the selection of the sample of very red QSO candidates, dust-obscured in the optical and with $z > 3$. Beside of the catalogue, the interest of this research lies in the collection of the data mining techniques presented and in the illustration of the problems that may occur in this kind of analysis.

6.2 Astrometry of the surveys

In this section we describe concisely the characteristics of each of the survey used in this work, including the sky coverage, wavelength and flux limit. In particular this section is focused on the discussion and determination of the astrometric accuracy of FIRST (subsection 6.2.1), SDSS (subsection 6.2.2) and UKIDSS (subsection 6.2.3). These determinations will be needed for the determination of the cross-match radii and for the determination of the likelihood ratio.

6.2.1 FIRST astrometry

In this work we have used the 2008 July-16 version of the FIRST catalogue that counts 816,331 sources and covers a total area of 9,055 deg² of the sky: 8,444 deg² in the North Galactic cap and 611 deg² in the South Galactic cap.

The astrometry of FIRST is discussed in details in White et al. (1997). In that work, from the offset of 79 bright point-like FIRST sources with MERLIN calibrators, they find rms positional errors of 0''.22 ($\alpha = 0''.17$ and $\delta = 0''.14$) and a small systematic uncertainty, estimated to be less than 0''.03 after a further comparison with the positions of 24 FIRST sources with another independent set of calibrators (catalogue of Johnson et al. 1995). Nevertheless the individual accuracy on the positions varies from source to source and depends on the brightness and size of the source, and to the noise in the map. In White et al. (1997) they give an empirical expression for the positional accuracy at 90% confidence (i.e. 1.645 σ) as:

$$(B^2 + \phi^2)^{1/2}(1/SNR + 1/20)arcsec \quad (6.1)$$

where $B = 5''.4$ is the beam size, ϕ is either the major or the minor axis FWHM as given in the catalogue, and SNR is the peak flux density signal-to-noise ratio given by $SNR = (S_p - 0.25)/\sigma(S)$; with S_p flux peak and $\sigma(S)$ the rms on its measure. The uncertainty is elliptical for elliptical sources. The limit on positional accuracy for bright sources is $\approx 0''.2$, which is 4% of the synthesized beam FWHM.

We calculate the individual uncertainties on the positions of all the sources of the FIRST catalogue using equation 6.1 rescaled to one σ , for both major and minor axis. The empirical equation does not allow to separate the positional uncertain in right ascension and declination, therefore we sum quadratically the uncertainties on the two axis to have an estimation of the combined accuracy on the positions of the individual sources as $\sigma_{comb,i} = (\sigma_{maj,i}^2 + \sigma_{min,i}^2)^{1/2}$. As resumed in Table 6.1, the median value of (σ_{comb}) is $\sim 0''.66$ for the whole sample. Its value increases to $\sim 0''.98$ for the subsample of the fainter sources (radio flux peak $S_{1.4GHz} < 2$ mJy), and it drops to $\sim 0''.48$ for the remaining sources. However, the errors σ_{comb} are $< 1''.0$ for $\sim 70\%$ of the catalogue, as resumed in the cumulative histograms in Fig. 6.1.

To have an estimation of the average σ_{comb} of the catalogue, we calculate the rms of the individual $\sigma_{comb,i}$ associated to each source. Obtaining $\sigma_{comb} = 0''.94$; while for population with $S_{1.4GHz} > 2$ mJy, the rms is 0''.53. In White et al. (1997) and in Becker et al. (1995) they claim that point-sources have 90% confidence error circles of radius less than 1''.0 at the survey limit, where point-like sources are the ones with quoted major axis of 2''.0 or less. Restricting our calculation to only this population, we obtain for σ_{comb} an rms of 0''.58 at 1σ . At the 90%

Table 6.1: Median values of the global positional errors, σ_{comb} , of the FIRST radio-sources. The medians are given for the whole FIRST catalogue, for the subsample composed by the fainter sources ($S_{1.4GHz} < 2$ mJy), and for the subsample composed by the brighter sources ($S_{1.4GHz} > 2$ mJy)

Sample	Median (σ_{comb})
Whole sample	0'.66
$S_{1.4GHz} < 2$ mJy	0'.98
$S_{1.4GHz} > 2$ mJy	0'.48

confidence errors, we have rms of 0'.95 that is in agreement with FIRST claim for the positional accuracy.

6.2.2 SDSS Astrometry

The optical data used in our analysis were taken from the 7th data release of the Sloan Digital Sky Survey (Abazajian et al. 2009). A detailed description of the astrometric calibration in SDSS is given by Pier et al. (2003) and it is based on its DR1, where the systematic offset in astrometric calibrations are estimated to be $\leq 0'.020 - 0'.30$. All of DR7sources have been recalibrated astrometrically and while the systematic errors are not well characterized, they are thought to be less than 0'.02 (quoted in DR7-SDSS webpage as: Zacharias, private communication). The *photoObjAll* table (described in Chapter 2) and derived views in the Catalogue Archive Server (CAS) database of the SDSS-DR7 include the errors for the right ascension and declination for all objects. These errors are calculated by adding the centroid errors in quadrature with the estimated astrometric calibration errors.

We test the positional accuracy of the SDSS sources in two ways. Comparing the position of 2,301 bright DR7-SDSS radio-sources with the positions of the same sources as given in International Celestial Reference System (ICRS) and measured using very-long-baseline interferometry (VLBI) . The resulting offset scatter-plot is shown in Fig. 6.2, where we estimate systematic uncertainty $\sim 0'.001$ ($RA = 0'.001$ and $DEC = 0'.001$) and a 1σ random uncertain of 0'.07 ($RA = 0'.052$ and $DEC = 0'.051$).

We made an independent estimation of the positional uncertainty using the errors in right ascension and declination as given in the SDSS database. For this estimation we consider the 798,251 primary SDSS sources included in the area $169^\circ \leq RA \leq 191^\circ$ and $0^\circ \leq DEC \leq 1.2^\circ$ (this area is covered also by FIRST). We calculate the rms of these errors in $RA = 0'.19$, and in $DEC = 0'.26$, with a resulting total uncertainty of 0'.32. The cumulative histogram of the total positional uncertainty is given in Fig. 6.3. When we restrict our analysis to the bright point-like sources ($\sim 120,000$ sources, $i < 21$, morphologically classified as

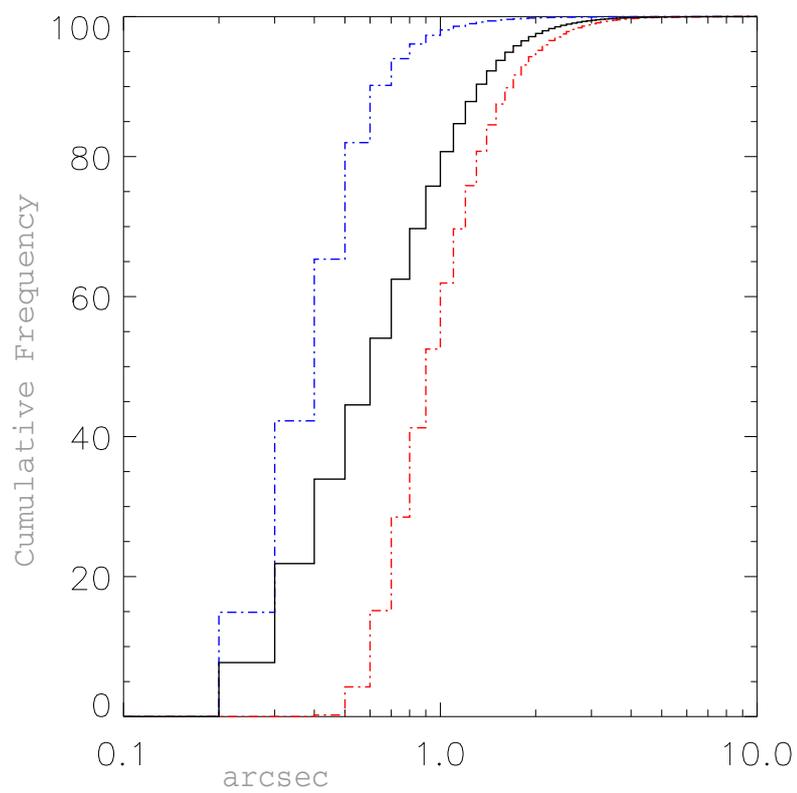


Figure 6.1: *Cumulative histograms of the global positional errors of the sources in the FIRST catalogue. The red line represents the population of the sources with ($S_{1.4\text{GHz}} < 2 \text{ mJy}$), the blue one the population with ($S_{1.4\text{GHz}} > 2 \text{ mJy}$) and the solid black line the whole population.*

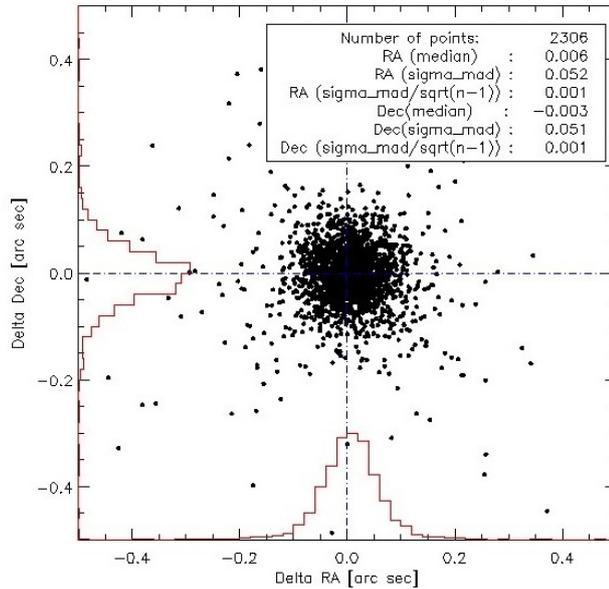


Figure 6.2: *ICRS-SDSS DR7 offset for a sample of 2,301 bright stars*

star), we obtain rms in $RA = 0''.08$, $DEC = 0''.09$, with a resulting total uncertainty of $0''.12$

6.2.3 UKIDSS Survey and astrometry

The UKIRT (United Kingdom Infrared Telescope) Infrared Deep Sky Survey (UKIDSS) is a set of five complementary surveys in the near infrared, which began in 2005 May. Three are the surveys targeting extragalactic fields, different in depth and area covered: the Large Area Survey (LAS), the Deep Extragalactic Survey (DXS), and the Ultra Deep Survey (UDS). The two surveys targeting Galactic fields are: the Galactic Plane Survey (GPS) and the Galactic Clusters Survey (GCS). Scope, layout, and broad science goals of the five components of UKIDSS are described in Lawrence et al. (2007). The instrument used for the survey is the Wide Field Camera (WFCAM) on UKIRT, with pixel scale $0''.4$ and giving a solid angle of 0.21 deg^2 per exposure. The characterization of the YJHK photometric system of the camera, which covers the wavelength range $0.83 - 2.37 \mu\text{m}$, is described in Hewett et al. (2006). In this chapter we used the 9th Data Release of the LAS survey. It covers $\approx 2500 \text{ deg}^2$ from within the footprint of the SDSS, in the four bands Y,J,H,K. The depth of the four bands is respectively (Vega system) 20.3, 19.9, 18.6, 18.2. The survey counts over 69×10^6 of distinct objects detected at least in one band.

The positional errors in right ascension and declination of the sources included

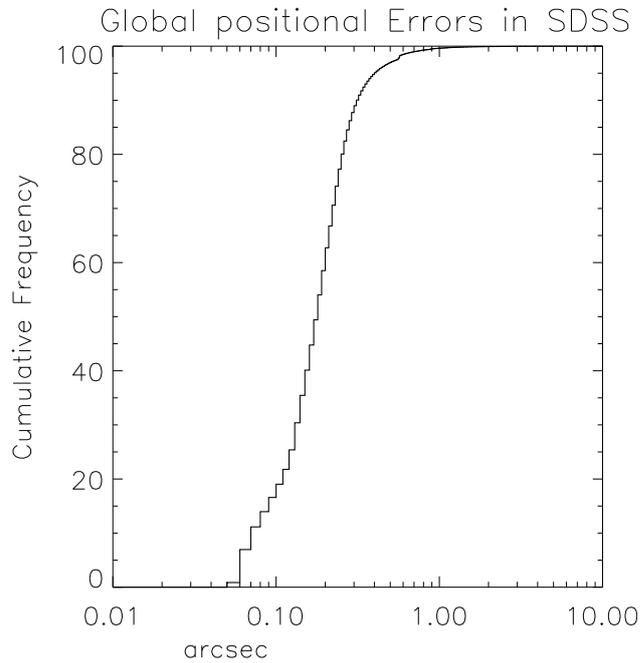


Figure 6.3: *Cumulative histogram of the combined positional error σ_{comb} in SDSS*

in the LAS table, are given in the WFCAM Science Archive (WSA, Hambly et al. 2006). However, these values are computed for small regions of the sky where it is calculated the proper-motion of the sources (Mike Read, private communication). The DR9 was the first release where proper motion, and therefore the positional errors, have been actually calculated.

We estimate the combined positional errors for UKIDSS sources, σ_{comb} , from a sample of sources included in the area $170^\circ < RA < 190^\circ$ and $-0.5^\circ < DEC < 2.5^\circ$ covered also by FIRST and SDSS. This area includes 1,384,818 sources, and for 86,291 positional errors were calculated. We calculate the the $\sigma_{comb,RA} = 0''.017 = \sigma_{comb,DEC} = 0''.017$ as the rms of the individual positional uncertainties. The cumulative histogram of the total positional uncertain is given in Fig. 6.4. Restricting the evaluation to only the 24,335 star-like sources (using the UKIDSS morphological classification) we calculate $\sigma_{comb,RA} = \sigma_{comb,DEC} = 0''.011$.

As further test we repeated the analysis using the estimation of the centroid error on the image of the detection, this quantity is expressed in units of pixels. We plot in Fig. 6.5 these errors for the K-band of the 224,260 star-like sources included in the sky area that we selected. The centroid errors have a median of 0.11 and a σ of 0.088 pixels, that correspond (each pixel size $0''.4$) to a median of $0''.04 \pm 0''.03$, in agreement with the results obtained calculating σ_{comb} .

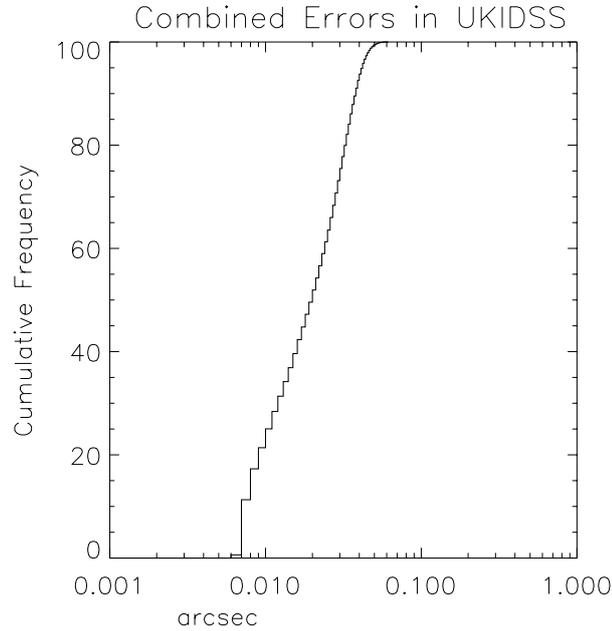


Figure 6.4: *Cumulative histograms of the global positional errors (σ_{comb}) of the sources in the UKIDSS-catalogue*

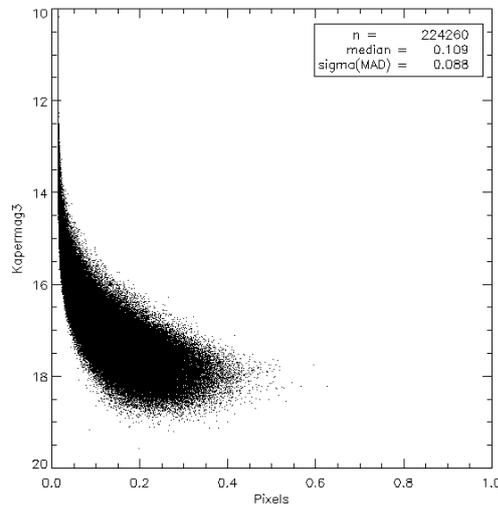


Figure 6.5: *K-magnitude vs the centroid error on the image of the detection for each of the 224,260 star-type source included in the area $170^\circ \leq RA \leq 190^\circ$ and $-0.5^\circ \leq DEC \leq 2.5^\circ$. The pixel errors have a median of 0.11 and a sigma of 0.088 pixels, that correspond (each pixel size $0''.4$) to a median of $0''.04 \pm 0''.03$*

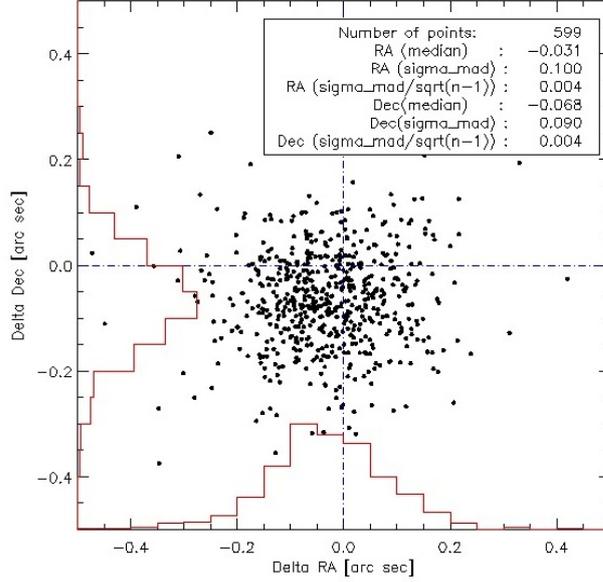


Figure 6.6: *ICRS - UKIDSS LAS DR9 offset for 599 bright sources*

The systematic error is evaluated from the offset between the position of 599 bright UKIDSS sources with the positions of the same sources as given the International Celestial Reference System (ICRS), see Fig. 6.6. From this analysis we calculate systematic positional errors of: $\sigma_{sys,RA} = \sigma_{sys,DEC} = 0''.004$, and random positional errors of: $\sigma_{RA} = 0''.10$, and $\sigma_{DEC} = 0''.09$.

6.3 Preselection

We built the bulk of our catalogue with two subsequent cross-matches. First we cross-matched the FIRST catalogue to UKIDSS (section 6.3.1), and for each radio source we searched for the closest infrared counterpart within $1''.5$ radius. The resulting sample was then cross-matched to SDSS (section 6.3.2), and we selected the subsample of sources without optical counterpart in $10''.0$ radius. The results of this preselection, the choices of the cross-match radius and some details on the SQL used will be discussed in the following subsections.

6.3.1 Cross-match FIRST-UKIDSS

We matched the 816,331 radio-sources of the FIRST catalogue with the DR9 UKIDSS LAS table, searching for the closest match within a $1''.5$ radius. The cross-match radius was chosen to optimize the selection of QSOs and therefore

Table 6.2: Positional errors in the position of the sources given in FIRST, SDSS, UKIDSS. The total σ is given combining systematic and random errors in quadrature

Sample	σ	RA (Systematic)	DEC	RA (Random)	DEC
FIRST ⁽¹⁾	0''.94	0''.030 ^(a)		0''.939 ^(b)	
FIRST ⁽²⁾	0''.58	0''.030 ^(a)		0''.577 ^(b)	
SDSS ⁽³⁾	0''.32	0''.001 ^(c)	0''.001 ^(c)	0''.195 ^(d)	0''.260 ^(d)
SDSS ⁽⁴⁾	0''.12	0''.001 ^(c)	0''.001 ^(c)	0''.082 ^(d)	0''.092 ^(d)
UKIDSS ⁽⁵⁾	0''.02	0''.004 ^(e)	0''.004 ^(e)	0''.017 ^(f)	0''.017 ^(f)
UKIDSS ⁽⁶⁾	0''.01	0''.004 ^(e)	0''.004 ^(e)	0''.011 ^(f)	0''.011 ^(f)

Positional errors on (1) FIRST sources, the whole sample; (2) FIRST star-like radio-sources, i.e. with Major Axis $< 2''$ in the radio; (3) SDSS sources, the whole sample; (4) SDSS star-like sources (SDSS morphological classification) bright, i.e. $i < 21$. These values have been calculated in: (a) White et al. (1997); (b) in this work, see section 6.2.1; (c) in this work from the offset SDSS-ICRS, see Fig. 6.2; (d) in this work, see section 6.2.2; (e) in this work from the offset UKIDSS-ICRS, see Fig. 6.6; (f) in this work, see section 6.2.3.

to have a high completeness for the radio-infrared associations, while minimizing the contamination of false and not QSOs detections.

On the basis of what we discussed in section 6.2, both the combined positional errors of the radio and infrared sources can be approximated to $0''.6$. Therefore doing a cross-match of $1''.5$ radius we include the 98.76% of the true optical-infrared counterparts of the radio sources. From our analysis we conclude that the positional accuracy of the UKIDSS catalogue is even better than in SDSS. In Schneider et al. (2010) they show that the distribution of optical/radio offsets for radio-loud quasars occurs at about $0''.2$, implying a high degree of completeness in our choice of a $1''.5$ cross-match radius. While the use of a larger radius increases the completeness, it also increases the number of spurious sources (stars and galaxies) in our sample. Since we are interested in the selection of radio-detected QSOs, we want to reduce in particular the contamination of stars in our sample. This contamination is critical near the FIRST detection limit, where the astrometric uncertainties are greater, because the association of stars with radio sources is usually not caused by a large population of radio-emitting stars (Kimball et al. 2009), but from the offset from the radio positions. Since the sources of the FIRST catalogue are mutually separated at least by $3''.59$, the use of a smaller radius for the cross-match also avoids the "double-counting problem", i.e. objects that may be the closest counterpart for two different radio sources.

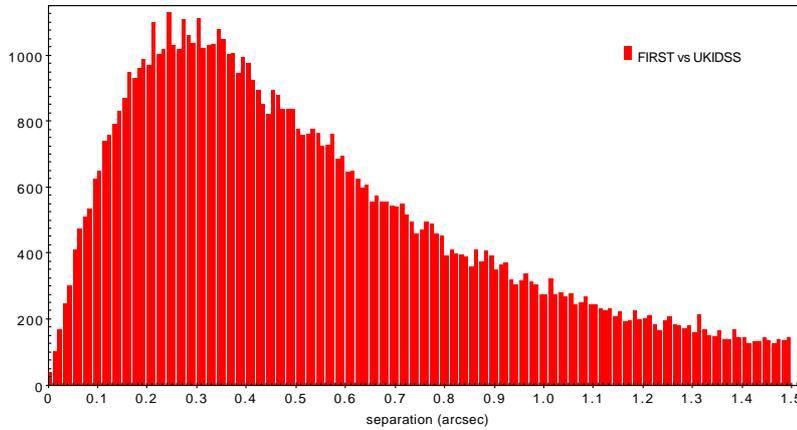


Figure 6.7: *Distribution of the mutual separation between FIRST and UKIDSS sources of our starting sample, composed of 76,541 unique (in both catalogues) sources*

For the cross-match we used the freeform SQL of the WSA-WFCAM Science Archive. Developing an efficient SQL query suited for our problem is not straightforward and we need to take into account a number of possible problems. Taking into account that we use a smaller cross-match radius than the minimum distance between two different radio sources, to a single FIRST source could correspond either: (1) a single unique real UKIDSS source, (2) multiple real UKIDSS sources, (3) a unique UKIDSS source and some or all of their eventually multiple duplicates (e.g. flags overlap duplicates), (4) one or multiple UKIDSS duplicates sources. We want to restrict our analysis to a one-by-one correspondence between the two catalogues, keeping only the closest unique UKIDSS source. In this case the association is likely to be real since the duplicates of the UKIDSS sources can have a positional inaccuracy that affect the reliability of the association. Therefore we developed the SQL to search for the nearest neighbor and exclude all the UKIDSS duplicates. Particular attention must be paid to the order of these two operations, avoiding the exclusion of real associations. In fact a UKIDSS duplicate can be closer to the FIRST source than its correspondent unique source, although both would be within the match radius adopted.

From the cross-match between FIRST and UKIDSS we obtained a starting sample of 76,541 sources; 51,805 of them detected in all the Y,J,H,K bands of UKIDSS. The distribution of the separation distances between radio and infrared counterparts is given in figure n. 6.7. The distance distribution peaks at small distances where associations are likely to be real, then decreases sharply. The width of the peak depends on the astronomical positional accuracy of the surveys, while at bigger distances the distribution does not fall to zero due to the increasing background contamination.

6.3.2 Cross-match [FIRST-UKIDSS] vs SDSS

The sample of 76,541 radio-infrared sources obtained in the previous section was then matched with Sloan Digital Sky Survey (SDSS) data release 7 (DR7). We searched for the nearest nearby primary SDSS object within $10''$ from the UKIDSS object coordinates. The ultimate motivation of our work is to select a sample of radio-infrared sources undetected in the optical, therefore the use of such a large cross-match radius assures that the true optical counterpart of the radio source would not be offset out of the matching radius because of positional uncertainty. Even considering the average positional errors calculated for extended sources of FIRST and SDSS in section 6.2, the combined positional error result to be $\sim 1''$ (see Table 6.2, being $0''.94$ for FIRST and $0''.32$ for SDSS). Using a $10''$ matching radius means to use 10σ combined positional error, i.e. statistically 1 on 6×10^{22} sources are outside of the gaussian area considered.

From this cross-match we obtain that 68,186 sources of our sample have optical-counterpart in the SDSS survey. In Fig. 6.8 we show the plot of the mutual distances between the FIRST and the SDSS positions. The remaining 8,355 are unmatched in SDSS. The next sections will be dedicated exclusively to this sample and to the techniques used in order to clean and classify these sources.

6.4 Sample undetected in SDSS

The sample of our interest is composed of 8,355 sources detected in the radio and in the infrared but unmatched in SDSS-DR7. Their distribution in right ascension and declination is shown in Fig. 6.9. The fact that this sample is undetected in SDSS does not guarantee that all sources are authentically fainter than the optical magnitude threshold of the survey. In fact, the sample can be contaminated by: (1) sources that are unmatched in SDSS for technical reasons related to the survey footprint area and/or detection problems, i.e. "false negative" (section 6.4.1); (2) sources that are "false positive", i.e. that are not likely to be real but are instead false detections in the radio or in the infrared (section 6.4.2). In this section we illustrate the techniques developed and applied to our sample in order to exclude both false positives and false negatives.

6.4.1 False negatives

Some of the sources of our FIRST-UKIDSS sample can result unmatched in SDSS because they are excluded from the PhotoPrimary table. This type of sources are "false negatives" because they are not really obscured in the optical. There are several reasons for which a source that has photometry in SDSS can result

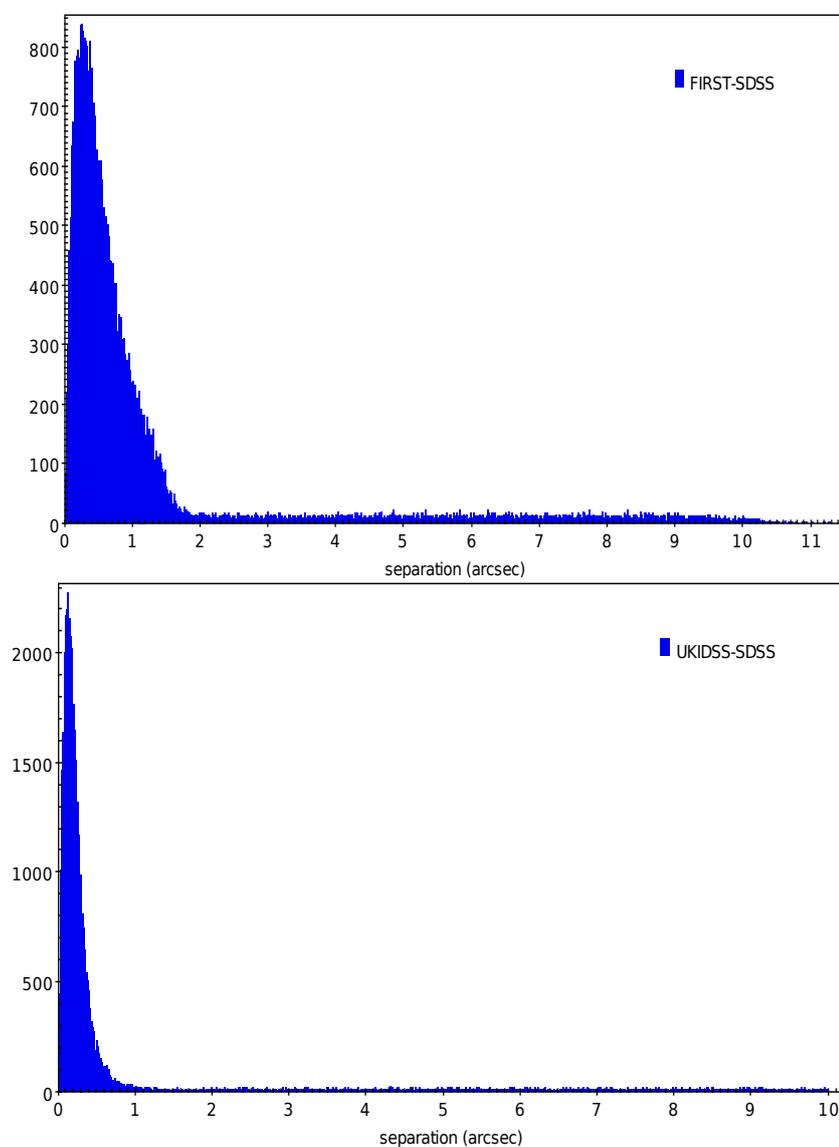


Figure 6.8: *Distribution of the mutual separation between the 68,186 FIRST-UKIDSS sources that have match in SDSS within $10''$. The upper plot shows the mutual distances between the FIRST coordinates and the SDSS coordinates, the bottom plot the mutual distances between the UKIDSS and the SDSS positions.*

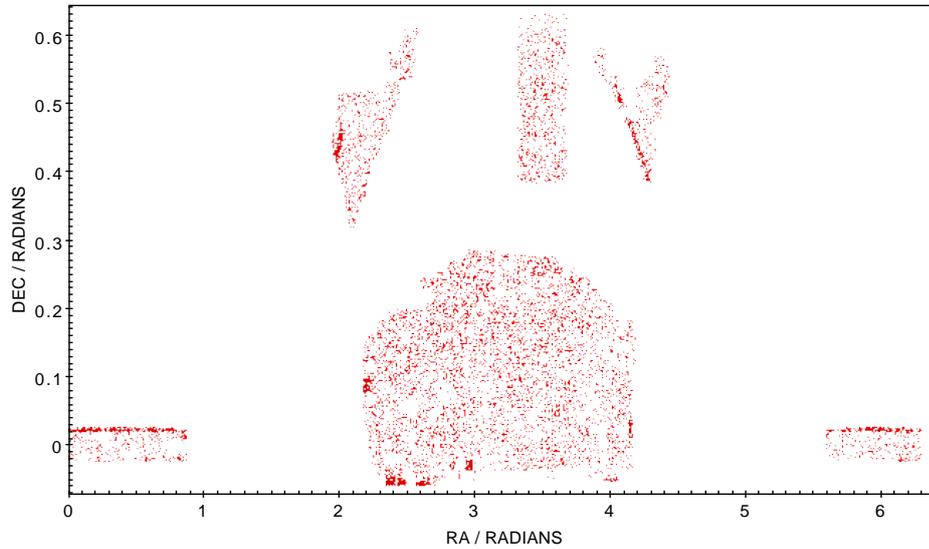


Figure 6.9: *The sample of our interest is composed of 8,355 sources detected in the radio and in the infrared but unmatched in SDSS-DR7*

excluded from the PhotoPrimary table. Here we list all these criteria excluding this type of sources from our sample:

1. Some real objects may be included in surveys areas that have been "masked" because they are potentially problematic for scientific analysis. There are 5 types of imaging masks in SDSS: BLEEDING and BRIGHT_STAR, that define those imaging areas covered by large saturated stars; TRAIL, that define imaging areas covered by trails caused by meteors, satellites, and the like; HOLE, that denote holes in the survey and indicate fields where the data quality is unacceptable; and SEEING that indicate survey regions that fail the seeing cut. We exclude from our sample all the 404 sources that result to be included in one of these imaging-masks. In Fig. 6.10 two examples of masked sources are shown.
2. Some sources fall outside of the SDSS-DR7 footprint. These sources, that we call "out of field", could be either be detected in SDSS or not. Since we search for sources authentically undetected in SDSS, we exclude all the sources "out of field". Once we exclude the "masked" sources, we exclude other 657 because they are "out of field". In Fig. 6.10 we present two examples of sources "out of field".

We also excluded from our sample of interest, all the 823 sources with counterpart in SDSS-DR8 (within $2''$). Of the remaining 6,471 sources "in field", "not

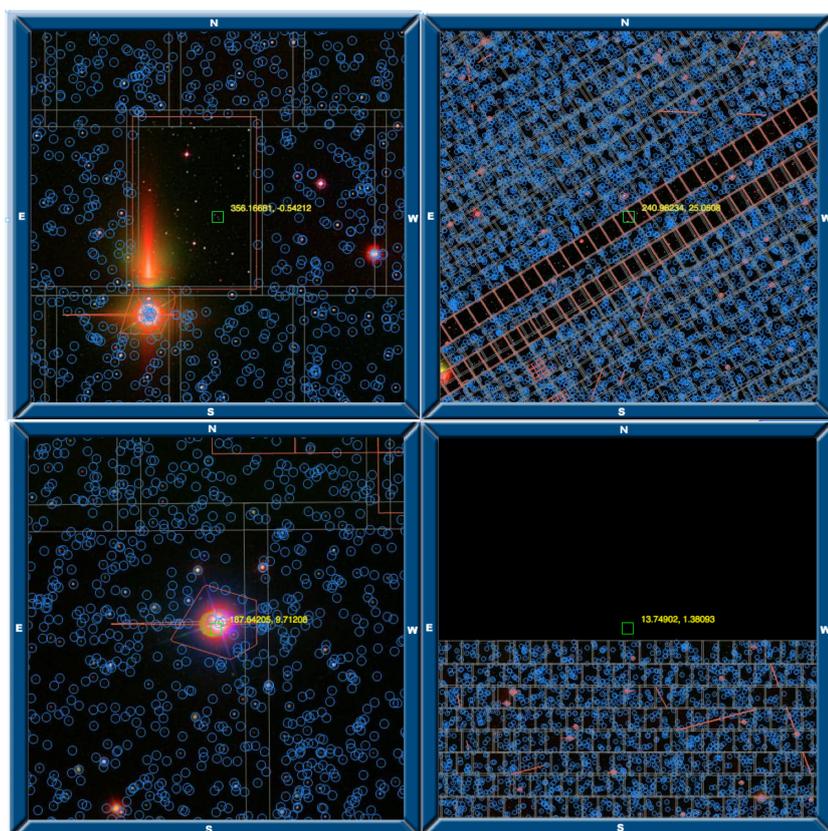


Figure 6.10: Examples of sources excluded from our sample because they were "masked" or outside of the DR7 SDSS footprint area (out of field). The Masks polygons are portrayed in red, the gray squares correspond to the fields and the blue circles are the photometric objects. The two examples on the top are sources masked and out of field: sometimes the mask bounds the entire field. On the bottom, on the left an example of source inside of the footprint area but masked; on the right, a source not masked but out of field.

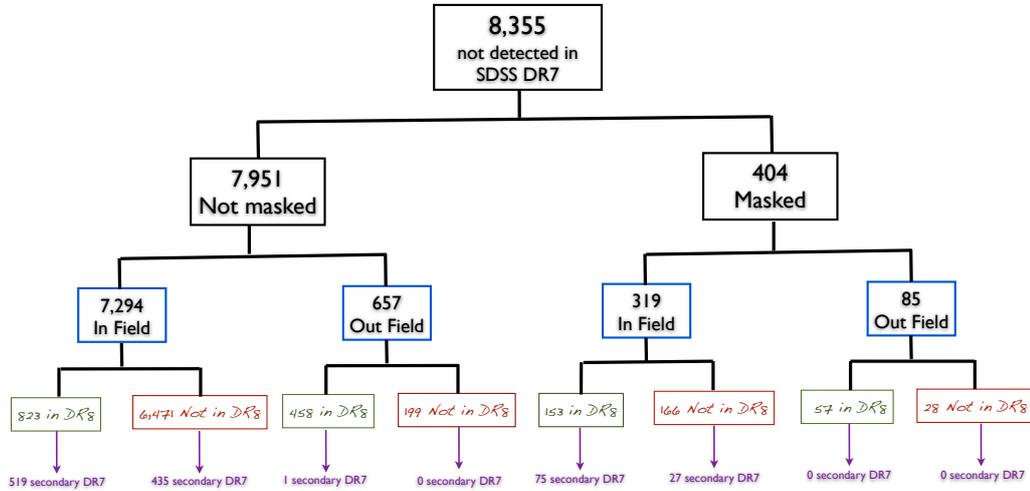


Figure 6.11: Schematic view of the process of elimination of false negatives from the sample of sources detected in FIRST and UKIDSS but undetected in SDSS. See section 6.4.1

masked” and unmatched in DR8, there is still a sample of 435 secondary DR7-SDSS counterpart not having any correspondent primary object (within $10''0$). We do not exclude these sources from our subsequent analysis but we FLAG all of them.

The whole selection-process explained in this subsection is summarized in the schema in Fig. 6.11.

Beside all these criteria applied, eye-ball-check of the SDSS images indicated some problems with the SDSS data that were not flagged. We find:

- objects with ”image” in SDSS but not photometry
- objects that result to be included within a mask in the image, but are not flagged as masked in the data
- objects that result undetected but are very close to extremely bright galaxies
- black field

In order to exclude these sources from the sample of the 6,471 sources ”in-field, not masked and not in DR8”, we downloaded all the images corresponding to the areas close to the positions of our objects. We developed and applied a decision tree method to keep only the sources authentically without image in SDSS. After the application of the decision tree, we also inspected eye-ball the whole sample and we corrected all the false classifications made from the decision tree algorithm. We added an ”eye ball warning flag” to sources with suspicious image.

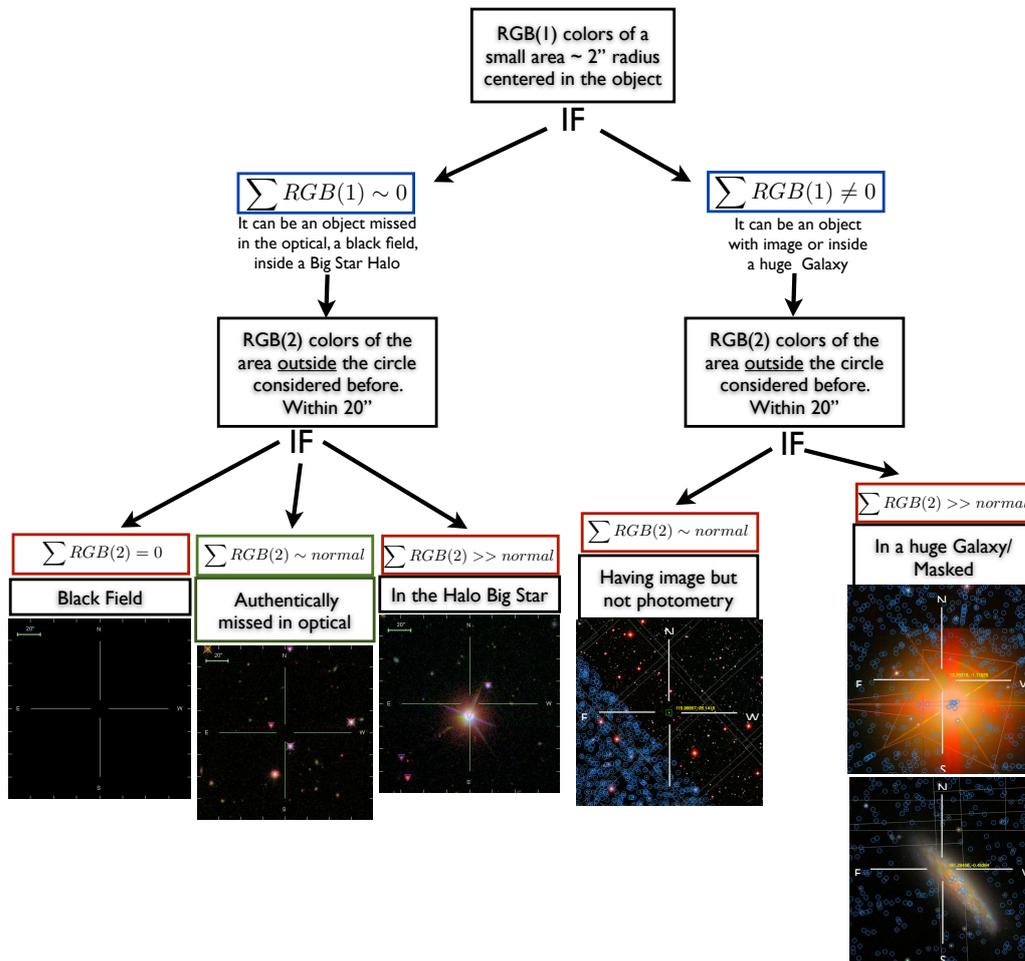


Figure 6.12: Flow chart of the decision tree used to discard false-negative sources.

The decision tree algorithm, classified wrongly less than 0.1% of the sources. In Fig. 6.12 we show the flow chart of the decision tree algorithm. A total of 200 sources have been flagged with this criteria. 45 of these 200 sources are also secondary without correspondent primary object in DR7.

6.4.2 False positives

The subsample of sources undetected in the optical and being "not masked", "in-field" "not detected in DR8" and "not discarded from our decision tree", counts a total of 6,271 objects. Nevertheless our sample can still be contaminated by "false negatives". The false negatives are UKIDSS artifacts, and to remove these objects we used several criteria:

- distribution of the mutual radial separations FIRST-UKIDSS
- use of the UKIDSS flags (see section 6.4.2)
- Visual inspection of the UKIDSS images
- Match of the sample with WISE

In this section we will explore all the criteria that we applied, and that led us to the conclusion that:

- Most of the Y-only sources are spurious
- Most of the K-only sources are likely to be real
- $\sim 80\%$ of the ONE-band only are likely to be real
- $\sim 90\%$ of the TWO-band are likely to be real

UKIDSS detection of our sample

Out of a total of 6,271 objects, 54% of them are detected only in one UKIDSS band, and the 29% in two bands (see Fig. 6.13). For comparison, in Fig. 6.13 we also show the bar chart of the sample detected in the optical (i.e. matched in FIRST-UKIDSS-SDSS) and with FIRST-UKIDSS separation $< 1''.5$. In this comparison sample, the majority of the sources ($\sim 83\%$) are detected in all the 4 UKIDSS band.

Considering the sources detected only in one band, in the sample undetected in the infrared, the majority of them are detected in K. In particular, 45% of the whole sample (2,799 sources) is detected only in K. In Fig. 6.14 we present the bar chart of the sources detected only in one band (145 only in Y, 130 only in J, 321 only in H, 2,799 only in K). This fact can indicate that our sample is detected only in K because it is composed of very red objects. The sample detected in SDSS and only in one UKIDSS band, is in general much fainter than the average sample (the whole sample has average r-magnitude ~ 20.6 , the subsample detected only in one band has average r-magnitude ~ 22.4). When we consider only the brighter sources ($r < 22$) the distribution of the sources detected only in one band is very different from the sources not detected in UKIDSS. The percentage of sources detected only in K increases as the magnitude faints ($r > 22$), as shown in Fig. 6.15

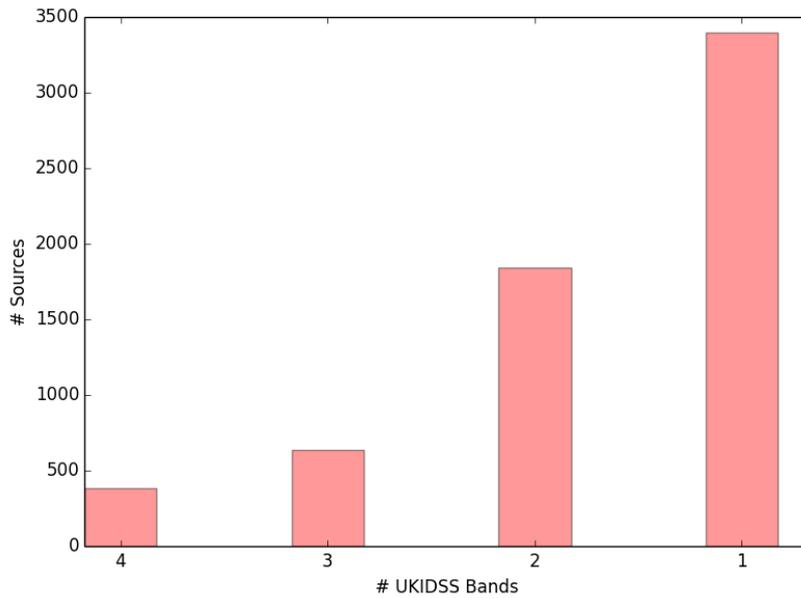


Figure 6.13: Bar chart distribution showing the number of UKIDSS bands in which our sample of interest is detected. We have that $\sim 54\%$ (3395 sources) is detected in 1 band; $\sim 29\%$ (1845 sources) is detected in 2 bands; $\sim 10\%$ (640 sources) is detected in 3 bands; $\sim 6\%$ (381 sources) is detected in all 4 bands.

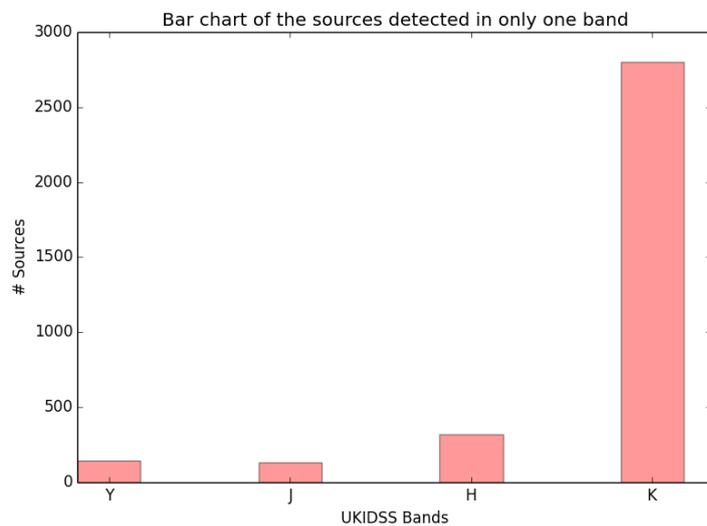


Figure 6.14: Bar chart distribution of the sources undetected in the optical and detected only in one UKIDSS band. On the total of 6,471 sources we have that: 145 are detected only in Y ($\sim 2\%$ whole sample), 130 only in J ($\sim 2\%$ whole sample), 321 only in H ($\sim 5\%$ whole sample), 2,799 only in K ($\sim 45\%$ whole sample)

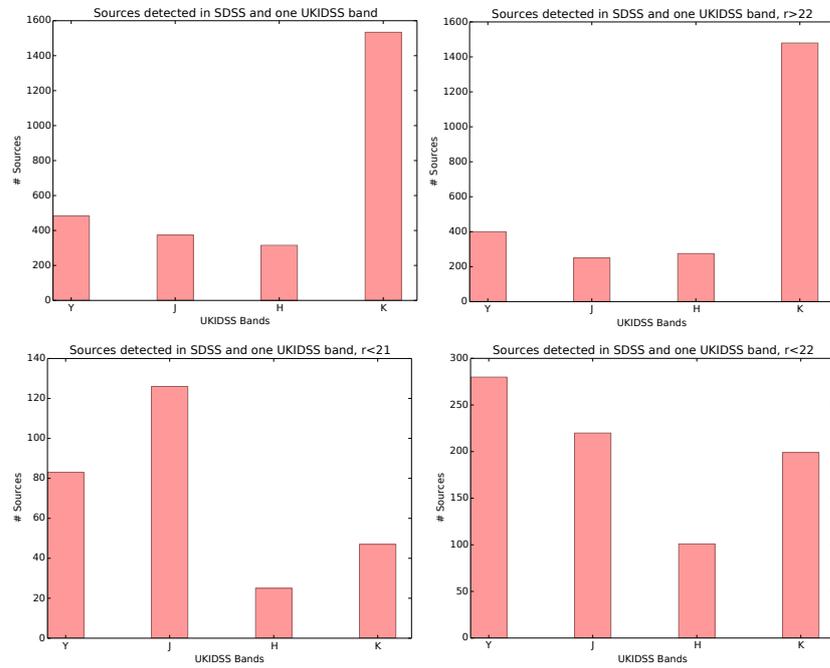


Figure 6.15: Histograms of detections for the sample of FIRST sources with counterpart in SDSS and detected only in one UKIDSS band. From the upper left, clockwise: (1) the whole sample; (2) having $r > 22$; (3) having $r < 22$; (4) having $r < 21$.

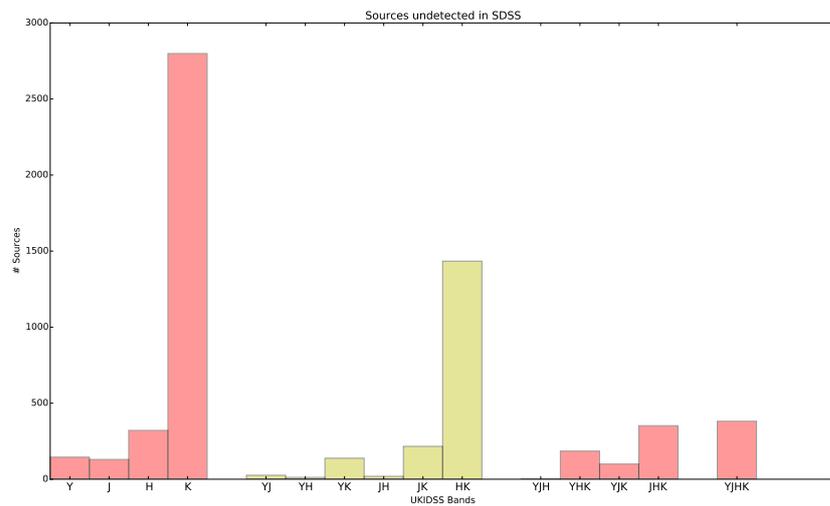


Figure 6.16: Global bar chart distribution of the 6,271 sources undetected in the optical (see Table 6.3).

Table 6.3: *UKIDSS detections for the 6,271 sources undetected in the optical.*

Band	sources
Y	145
J	130
H	321
K	2799
YJ	26
YK	138
JH	19
JK	216
HK	1434
YJH	3
YHK	185
YJK	100
JHK	352
YJHK	381

Mutual radial-separation analysis

The distribution of the radial separation between the FIRST and UKIDSS positions, follows a Rayleigh distribution when the sample includes matches with real sources. In Fig. 6.17 we plot the distributions of the FIRST-UKIDSS separation for the samples undetected in SDSS and detected only in one UKIDSS band (left panel). For comparison we plot the same distributions for the sources detected only in one UKIDSS band but detected in SDSS. The distribution of the sources detected in Y-only and undetected in SDSS, is not in agreement with a Rayleigh distribution. This suggests that the sample is composed by false sources.

Use of the UKIDSS error Flags

We consider two types of UKIDSS warning-flags. The first one is the "Y,J,H,K-ppErrBits". This is a post-processing error quality bit flag assigned in the WSA curation procedure for survey data. In the 4-byte integer attribute, byte 0 corresponds to information on generally innocuous conditions; byte 1 to warnings; byte 2 to important warnings; and byte 3 to severe warnings. The subsample of sources being flagged with at least byte-1, and at least in one band, counts 101 sources. The sources affected by this error are distributed in the UKIDSS bands as shown in Fig. 6.18

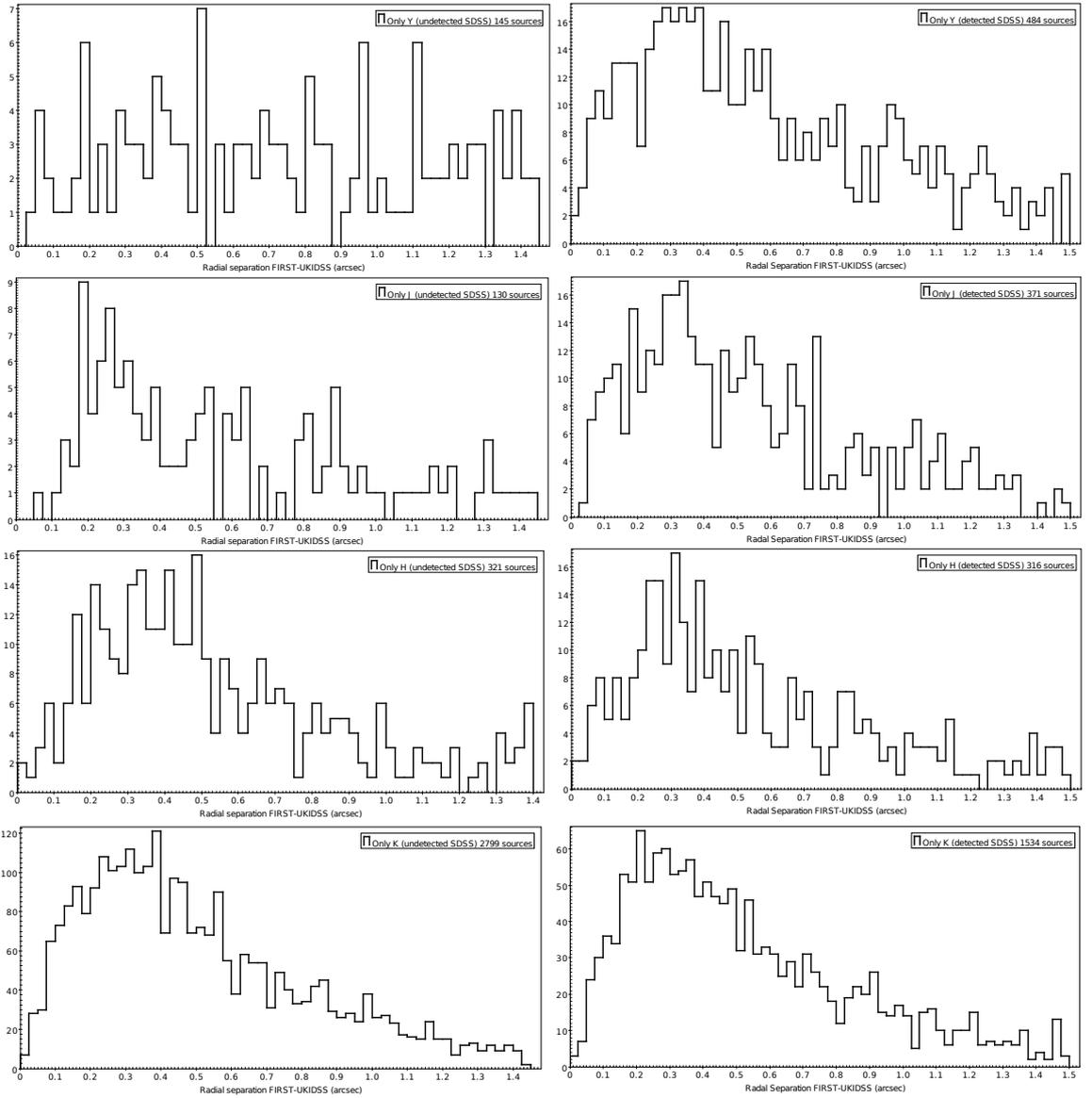


Figure 6.17: *Distribution of the mutual separations FIRST-UKIDSS for the sources detected only in one UKIDSS band. The left panel correspond to the sources undetected in SDSS, the right panel to the samples of sources detected in SDSS.*

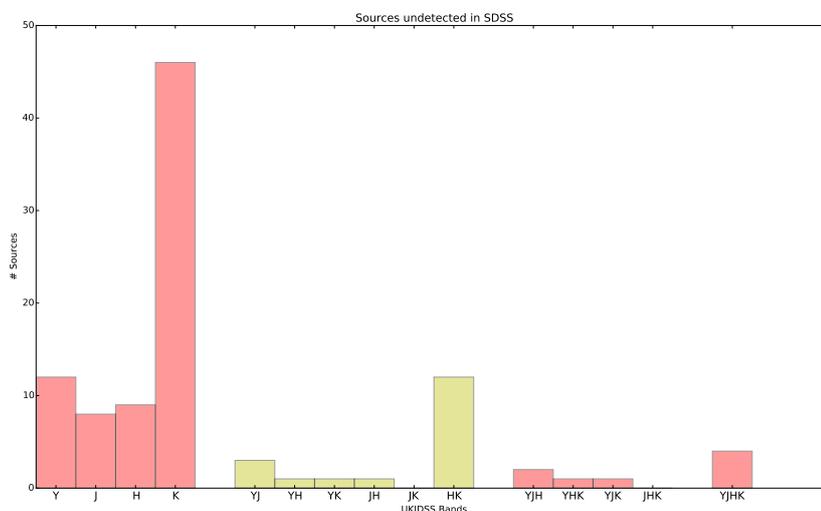


Figure 6.18: Detection chart of the sample undetected in SDSS and flagged by *ppErrBits* UKIDSS warning-Flag, a total of 101 sources.

The second type of UKIDSS warning flag that we consider is the "Y,J,H,K-ErrBits" flags. This one is not actually an error bit flag but a count of the number of zero confidence pixels in the default aperture (2 arcsec diameter). 482 of the sources of our sample have *ErrBits* > 0 at least in one detected-band, but the pixels flagged are ≤ 6 . In Fig. 6.19 it is shown the detection chart of these sources in the UKIDSS bands.

Match with WISE

We matched the 6,261 sources of our sample with the sources of the Wide-Field Infrared Survey Explorer (WISE), using a matching radius of $3''.0$. It results that 5,268 of these sources have counterpart in WISE and therefore are likely to be real. The distance distribution between radio and WISE positions is shown in Fig. 6.21, and it peaks at $\sim 0''.5$. The sources with counterpart in WISE are detected in the UKIDSS bands as shown in Fig. 6.20.

6.5 Eye Ball check of the pre-selection

In order to verify the efficiency of our criteria to discard spurious sources (false positive and false negative), we re-considered the whole sample of 8,355 radio-infrared sources undetected in the optical, and we visually checked the infrared images of the source. We selected the subsample of sources whose images looked unreliable (because they were close to a bright galaxy, inside the halo of a bright

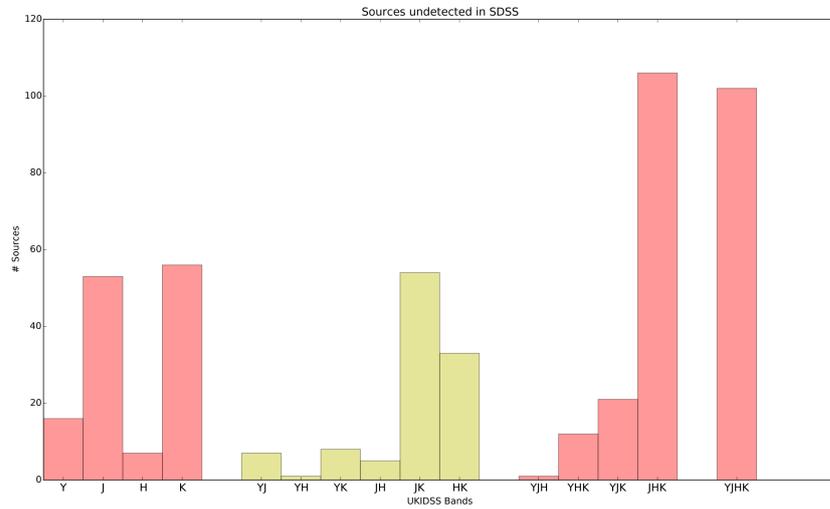


Figure 6.19: Detection chart of the 482 sources undetected in SDSS and with UKIDSS warning flag $ErrBits > 0$, at least in one band.

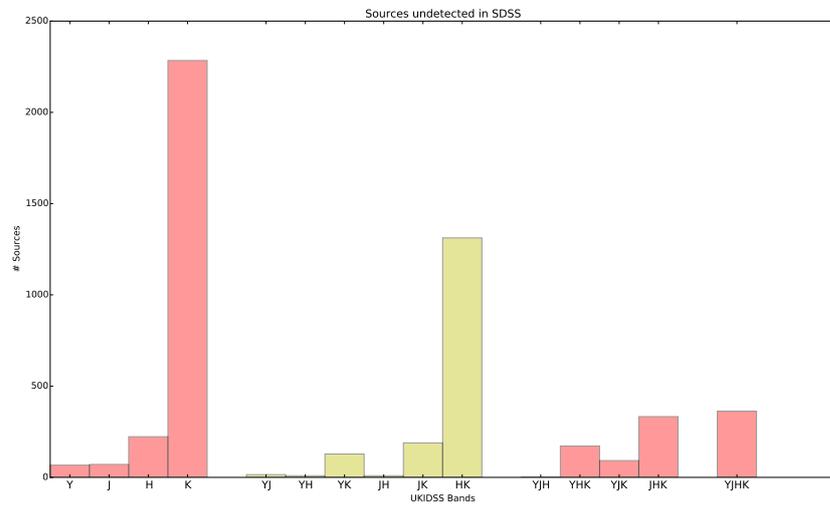


Figure 6.20: Detection chart in the UKIDSS bands for the of 5,268 sources (out of 6,261) detected also in WISE (see section 6.4.2).

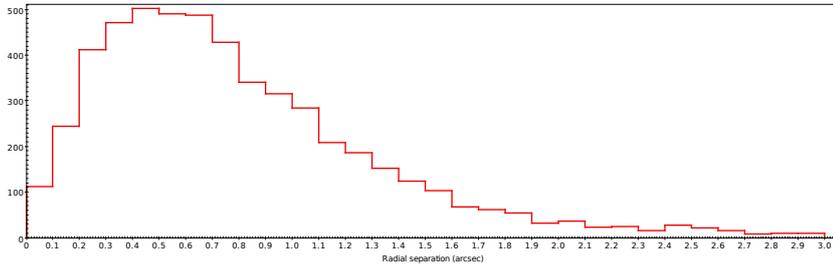


Figure 6.21: *The distance distribution between radio and WISE positions for the 5,268 radio-sources undetected in the optical and detected both in UKIDSS and in WISE (see section 6.4.2).*

star, false detection) and we checked how those sources were classified in the analysis previously discussed in this section. Almost all the sources whose infrared-image looked unreliable, were discarded from our decision tree algorithm or considered unreliable from our analysis, thus proving the effectiveness of our criteria.

6.6 Exploratory Data Analysis

In a number of studies based on the cross-match of radio sources with infrared and optical counterpart, it has been found that radio sources tend to reside in denser clusters (McAlpine et al. 2012). It has been suggested that this effect may result from the tendency for radio-loud AGN to favor denser environments than normal galaxies (Best et al. 2005; Falder et al. 2010), resulting in a large number of close neighbors for these sources. Previously to the analysis of the likelihood ratio, we evaluated through exploratory data analysis the effect of a possible denser environment in radio-loud AGN on the cross-match. In order to do that, we matched a sample of FIRST sources with all the SDSS sources within a fixed radius r from the radio positions. Then the positions of the FIRST sources were offset and the match with SDSS sources was repeated. The second match correspond to substitute the radio-positions with random positions having the same distribution and density of the radio-sources, and it is used to calculate the distribution of the cross-matches obtained for pure statistical chance and not because the same radio source is detected also in the optical. We compared the distributions for the mutual radial separations between central positions (radio or random) and their matched counterparts.

In particular, we consider the FIRST sources included in the area having coordinates in degree $155 \leq \alpha \leq 165$, $22 \leq \delta \leq 32$. This area includes 7,823 radio sources and 2,257,580 optical primary SDSS sources. From the cross-match of the radio-sources with SDSS, we calculate the distribution of the radial separa-

tion between sources and their counterparts. After that we offset the positions of radio-sources in 4 different ways and for each offset we repeat the cross-match. We average the distribution of the radial separation obtained by these 4 offsets, and we finally compare the resulting distribution with the one obtained from the cross-match of SDSS and the real radio positions. As shown in Fig. 6.22, there is a "background excess" and the two distributions became similar for a radius much larger than expected, i.e. around $70''$.

We repeated the cross-match considering only SDSS star and only SDSS galaxy. As shown in the Fig. 6.23 and in Fig. 6.24, the background excess disappears at radius $r < 5''$ in case of star-like sources, but it is enhanced for the case of galaxy-like sources.

The same analysis was repeated using UKIDSS instead of SDSS. With a different FIRST area, i.e. having coordinates in degree $165.5 \leq \alpha \leq 175$, $3 \leq \delta \leq 13$, and including 8,261 sources. The results are shown in Fig. 6.25, in Fig. 6.26 and in Fig. 6.27. We observe a similar behavior than the one used using SDSS counterparts, but the background excess appears more persistent in the case of galaxy-like sources.

6.7 Likelihood ratio

At the time of a cross-match between sources detected in different surveys and different wavelength, the likelihood (LR) ratio permits to establish the probability that a given counterpart is the correct association to a source. Following Sutherland & Saunders (1992), the LH ratio is given by:

$$L = \frac{q(m)f(r)}{n(m)} \quad (6.2)$$

where $q(m)$ and $n(m)$ are the normalized distributions of the magnitude m for the counterpart and background sources, while $f(r)$ is the radial probability distribution function of the offset between the positions of the radio source and its counterpart. The latter given by a gaussian

$$f(r) = \frac{1}{2\pi\sigma_{pos}^2} \exp\left(\frac{-r^2}{2\sigma_{pos}^2}\right) \quad (6.3)$$

where r is the offset between the radio and the associated optical or infrared source. The value σ_{pos} is the combined positional error of the radio and optical (or infrared) sources.

The distribution $q(m)$ is estimated using the method outlined in Ciliegi et al. (2003), which begins by calculating the magnitude distribution of all the possible

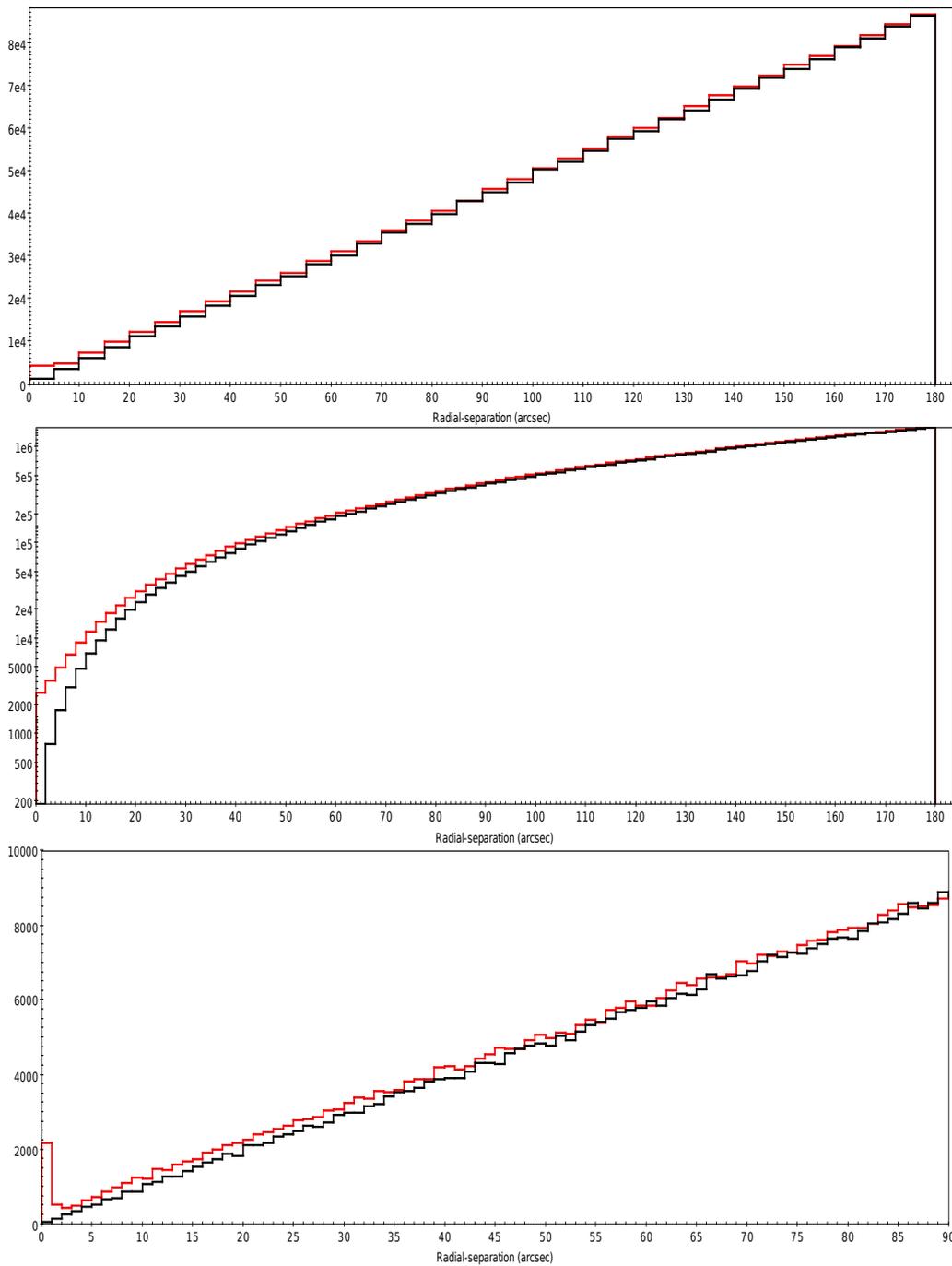


Figure 6.22: *Exploratory Data Analysis for the FIRST-SDSS cross-match (section 6.6), here considered for all kind of sources. The plots showing the mutual separation between the positions of the radio-sources and the positions of their matched optical-counterparts. The red line represent the cross match between the position of the FIRST sources and SDSS, the black line the cross match between random positions (obtained offsetting the positions of the FIRST sources) and SDSS sources. The upper and the middle plot represent the normal and the cumulative distributions up to $180''$; the bottom plot represent the normal distribution up to $90''$.*

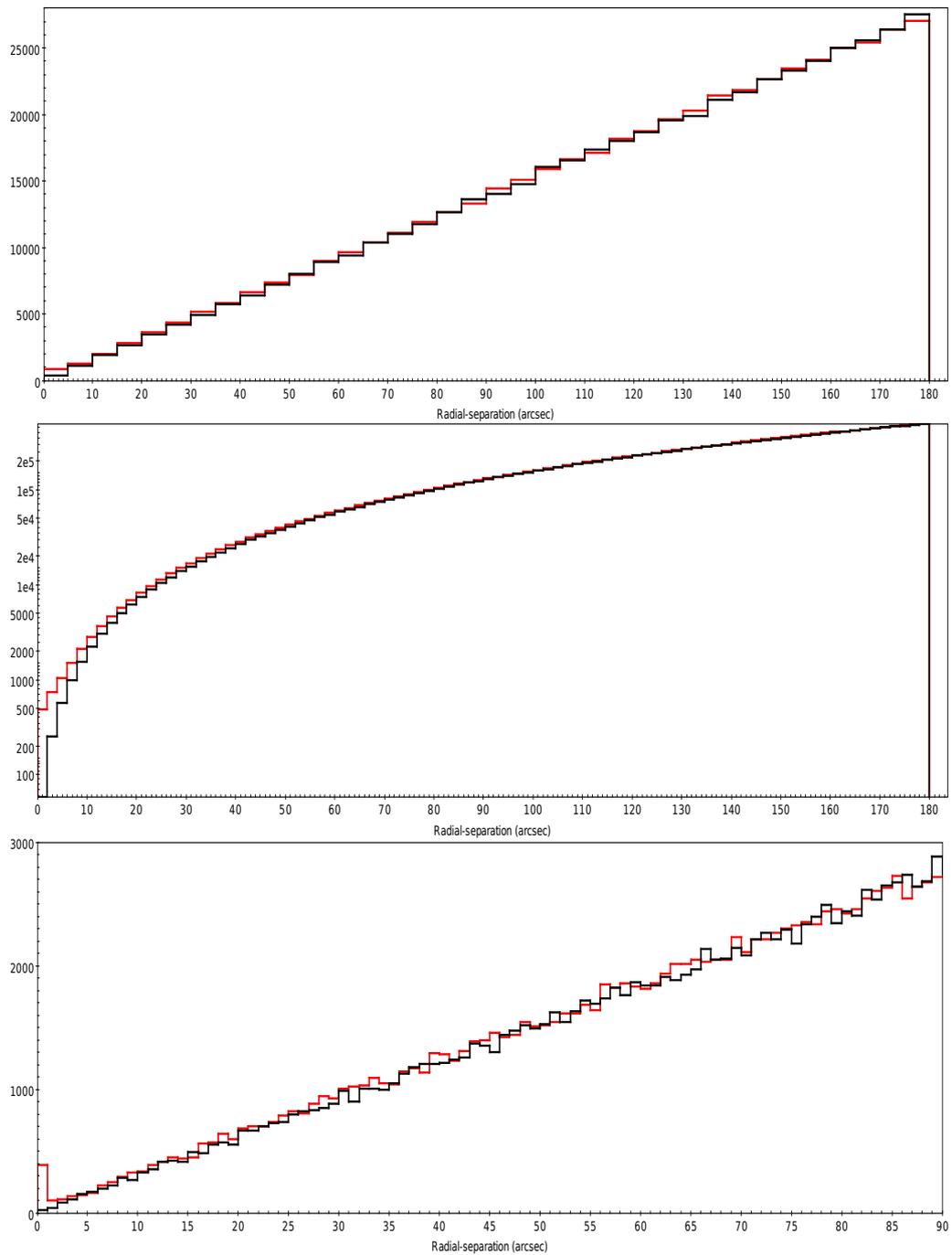


Figure 6.23: *Exploratory Data Analysis for the FIRST-SDSS cross-match (section 6.6), here considering only star-like SDSS-counterparts. The plots showing the mutual separation between the positions of the radio-sources and the positions of their matched optical-counterparts. The red line represent the cross match between the position of the FIRST sources and SDSS, the black line the cross match between random positions (obtained offsetting the positions of the FIRST sources) and SDSS sources. The upper and the middle plot represent the normal and the cumulative distributions up to 180''; the bottom plot represent the normal distribution up to 90''.*

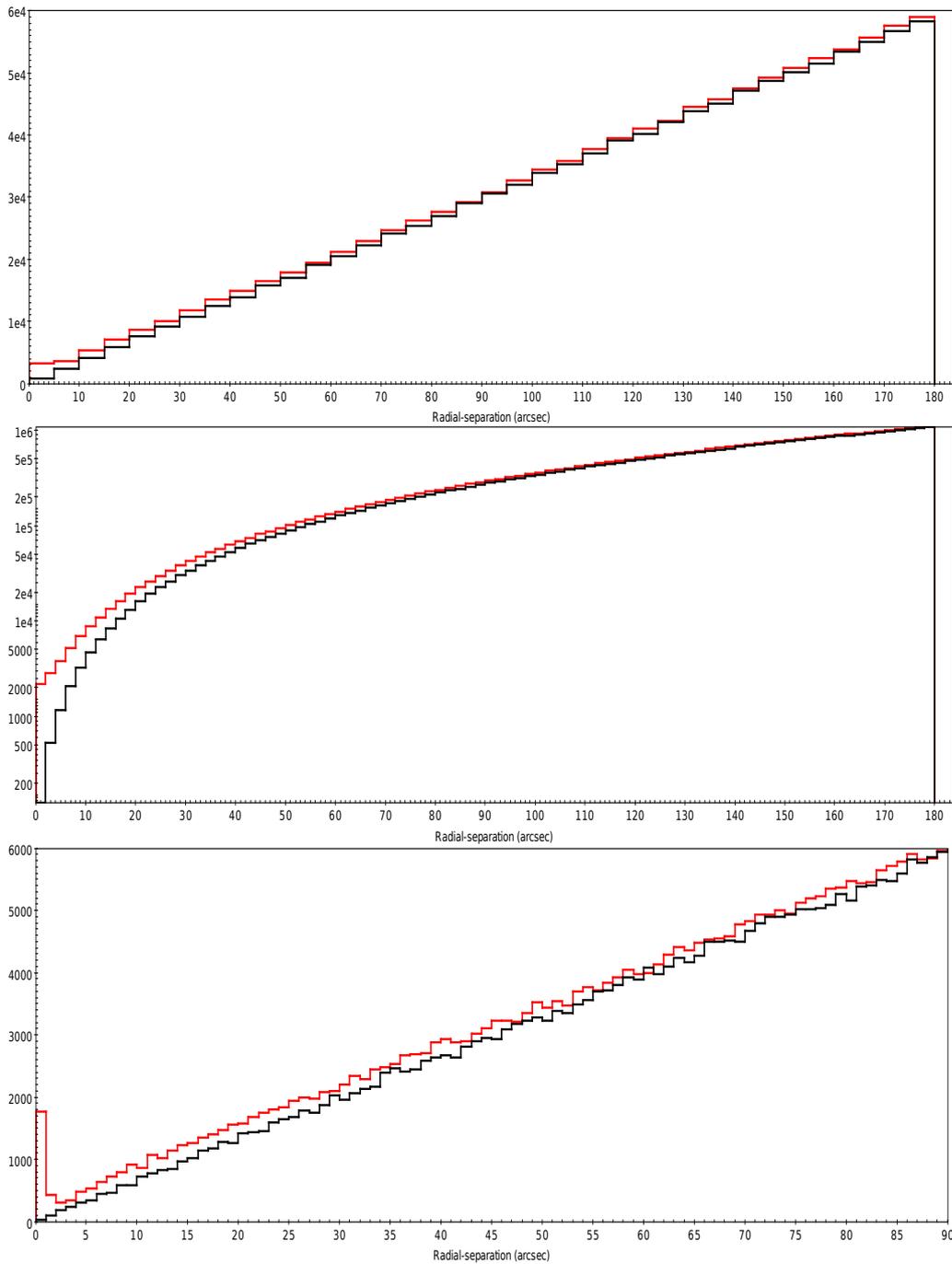


Figure 6.24: *Exploratory Data Analysis for the FIRST-SDSS cross-match (section 6.6), here considering only galaxy-like SDSS-counterparts. The plots showing the mutual separation between the positions of the radio-sources and the positions of their matched optical-counterparts. The red line represent the cross match between the position of the FIRST sources and SDSS, the black line the cross match between random positions (obtained offsetting the positions of the FIRST sources) and SDSS sources. The upper and the middle plot represent the normal and the cumulative distributions up to 180''; the bottom plot represent the normal distribution up to 90''.*

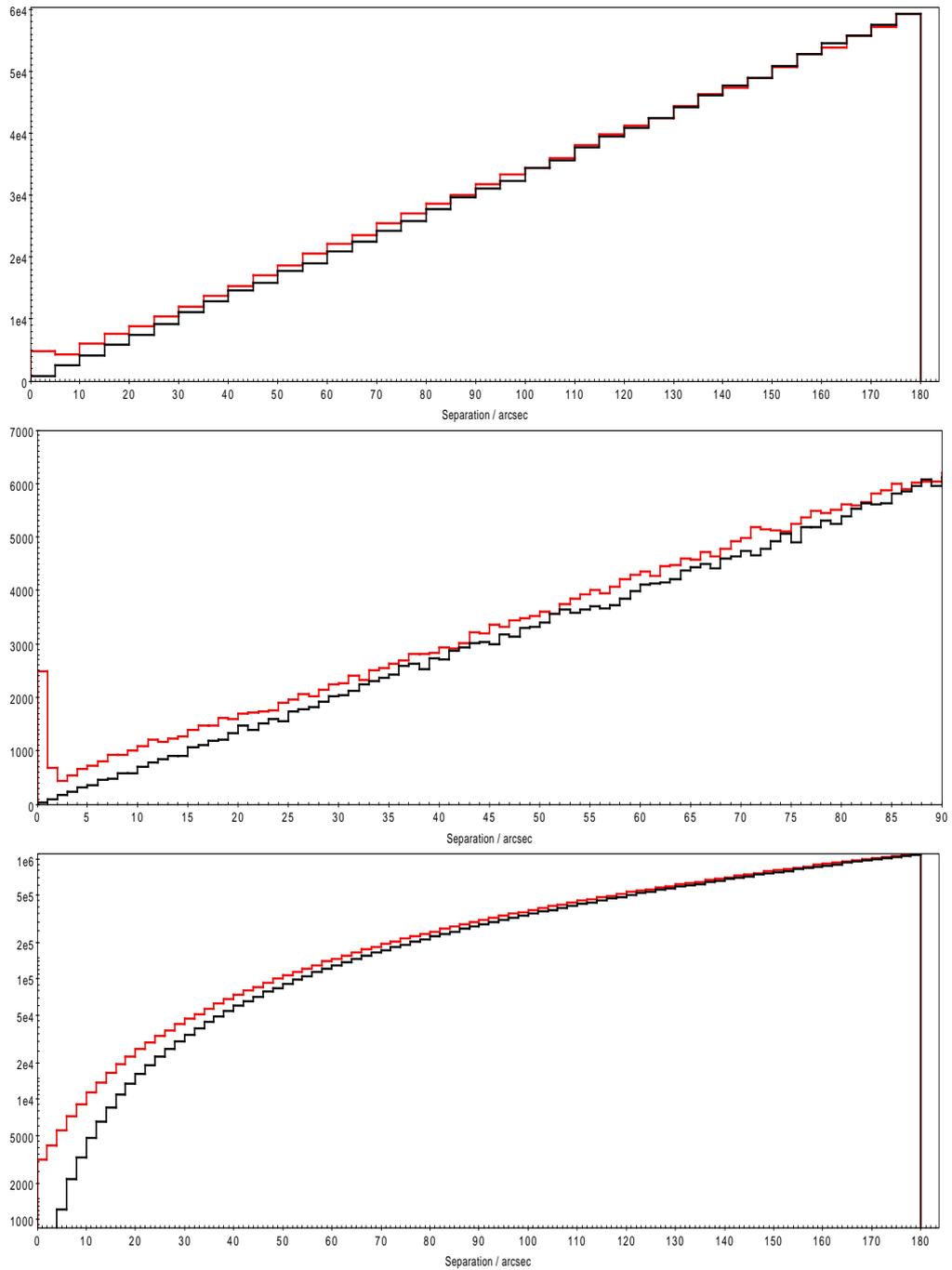


Figure 6.25: *Exploratory Data Analysis for the FIRST-UKIDSS cross-match (section 6.6), here considered for all kind of sources. The plots showing the mutual separation between the positions of the radio-sources and the positions of their matched infrared-counterparts. The red line represent the cross match between the position of the FIRST sources and UKIDSS, the black line the cross match between random positions (obtained offsetting the positions of the FIRST sources) and UKIDSS sources. The upper and the middle plot represent the normal and the cumulative distributions up to 180''; the bottom plot represent the normal distribution up to 90''.*

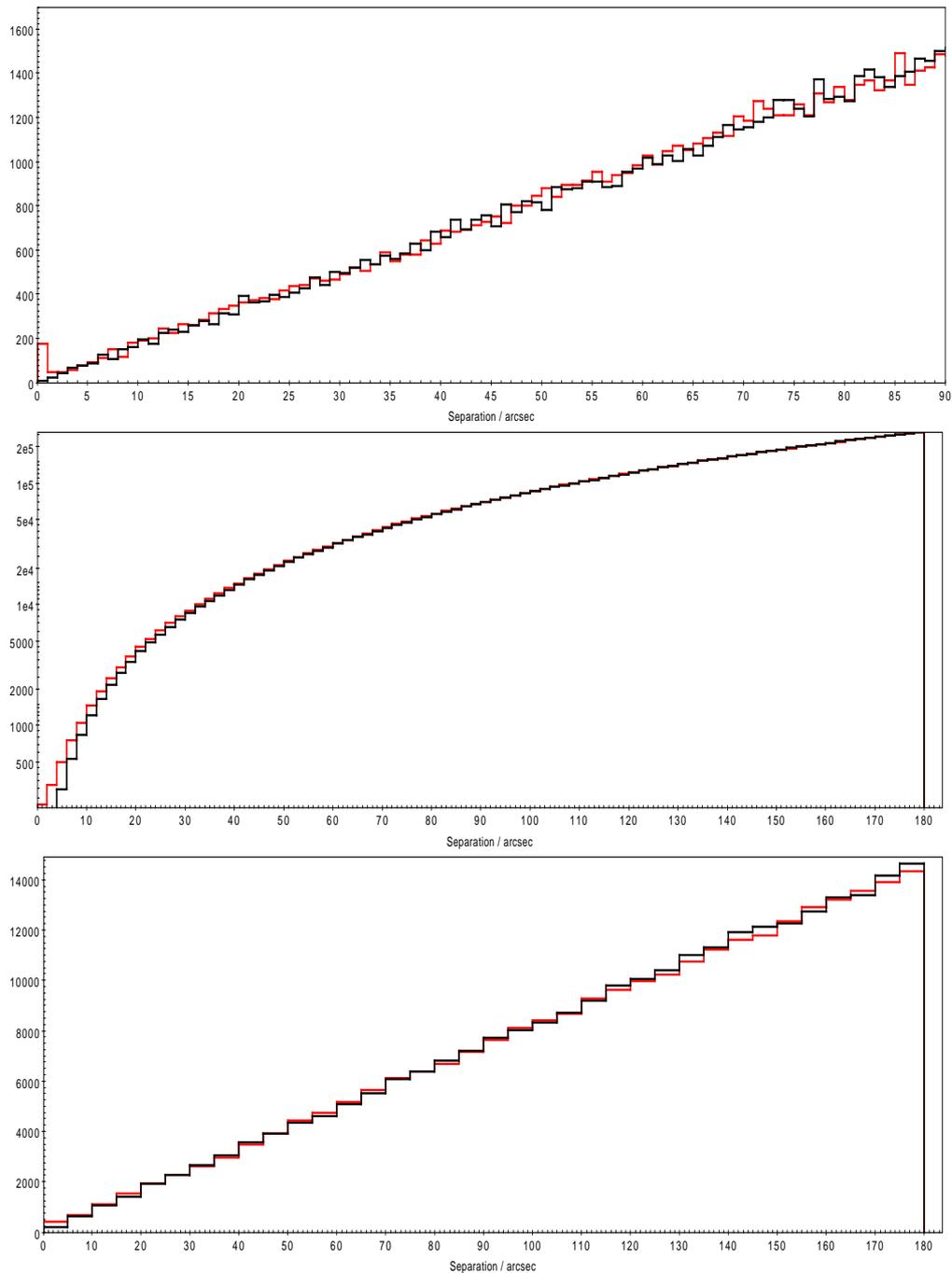


Figure 6.26: Exploratory Data Analysis for the FIRST-UKIDSS cross-match (section 6.6), here considering only star-like UKIDSS-counterparts. The plots showing the mutual separation between the positions of the radio-sources and the positions of their matched infrared-counterparts. The red line represent the cross match between the position of the FIRST sources and UKIDSS, the black line the cross match between random positions (obtained offsetting the positions of the FIRST sources) and UKIDSS sources. The upper and the middle plot represent the normal and the cumulative distributions up to $180''$; the bottom plot represent the normal distribution up to $90''$.

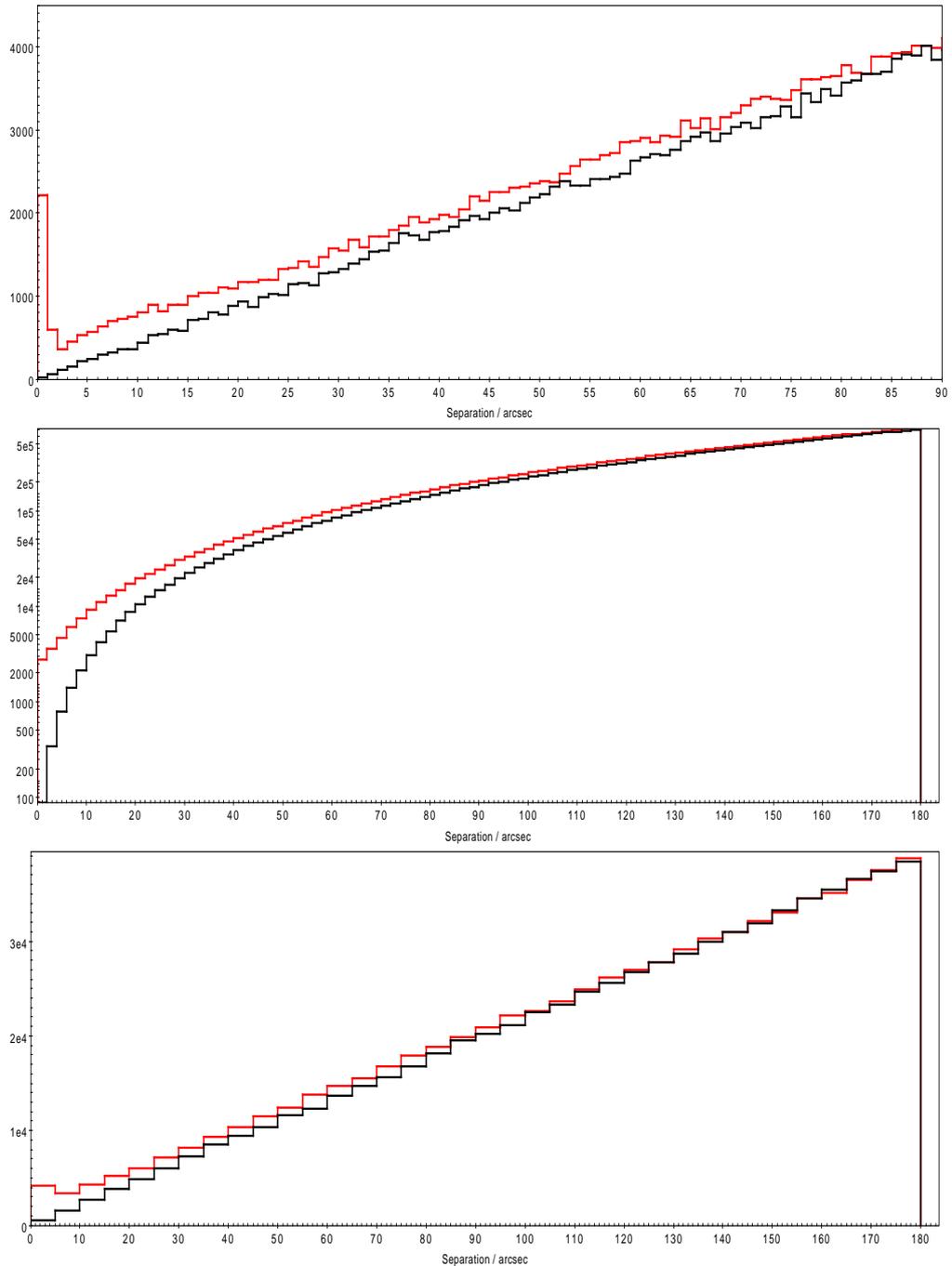


Figure 6.27: *Exploratory Data Analysis for the FIRST-UKIDSS cross-match (section 6.6), here considering only galaxy-like UKIDSS-counterparts. The plots showing the mutual separation between the positions of the radio-sources and the positions of their matched infrared-counterparts. The red line represent the cross match between the position of the FIRST sources and UKIDSS, the black line the cross match between random positions (obtained offsetting the positions of the FIRST sources) and UKIDSS sources. The upper and the middle plot represent the normal and the cumulative distributions up to $180''$; the bottom plot represent the normal distribution up to $90''$.*

counterparts within a fixed search radius r_{max} of the radio positions. This distribution is referred to as $total(m)$. The contribution due to the background source counts is subtracted from $total(m)$ to produce a magnitude distribution of the excess optical sources detected around the radio positions, designated as N_{excess} . This result in an estimation of the number of sources that are in excess to the background, and therefore assumed to be a measure of the true counterpart. Then $q(m)$ is derived from N_{excess} by normalizing and scaling by a factor Q_0 .

$$q(m) = \frac{N_{excess}(m)}{\sum_m N_{excess}(m)} \times Q_0 \quad (6.4)$$

The maximum radius for the match, r_{max} is often chosen to be 5 times the uncertain σ_{pos} . This means that the probability that the true position of the source is outside of the interval $(-5\sigma, 5\sigma)$, is $1/1,744,2278$. In this section we tested several values of r_{max} , in order to evaluate quantitatively the effect of the background excess discussed in section 6.6.

6.7.1 Estimation of the functions

The $n(m)$ function is the probability that a background source has magnitude m . To estimate this, many authors just calculate the distribution of the sources in the matched catalogue and normalize to the total area of the catalogue (for example in Smith et al. 2011; McAlpine et al. 2012). Instead of using that method, to take into account the distribution of the FIRST sources (i.e. not assuming a uniform distribution of sources), we calculate $n(m)$ from the distribution of all matched SDSS sources within an r_{max} distance from the positions of the FIRST sources offset from their original position (as described in the previous section). In practice, being $n'(m)$ the distribution of the matched sources, $n(m)$ is obtained after normalization to the area:

$$n(m) = \frac{n'(m)}{N_{centres} \times \pi r_{max}^2} \quad (6.5)$$

where $N_{centres}$ is the total number of FIRST sources considered for the match.

The probability distribution function $q(m)$ is derived as in equation 6.4 from

$$N_{excess}(m) = total(m) - n'(m) \quad (6.6)$$

The value Q_0 is the fraction of true counterparts which are above the SDSS limit, for definition $Q_0 = \int q(m)dm$, integrated within the magnitude limits of the survey. Practically it is commonly calculated as:

$$Q_0 = \frac{N_{matches} - (\sum_m n(m) \times \pi r_{max}^2 \times N_{centres})}{N_{centres}} \quad (6.7)$$

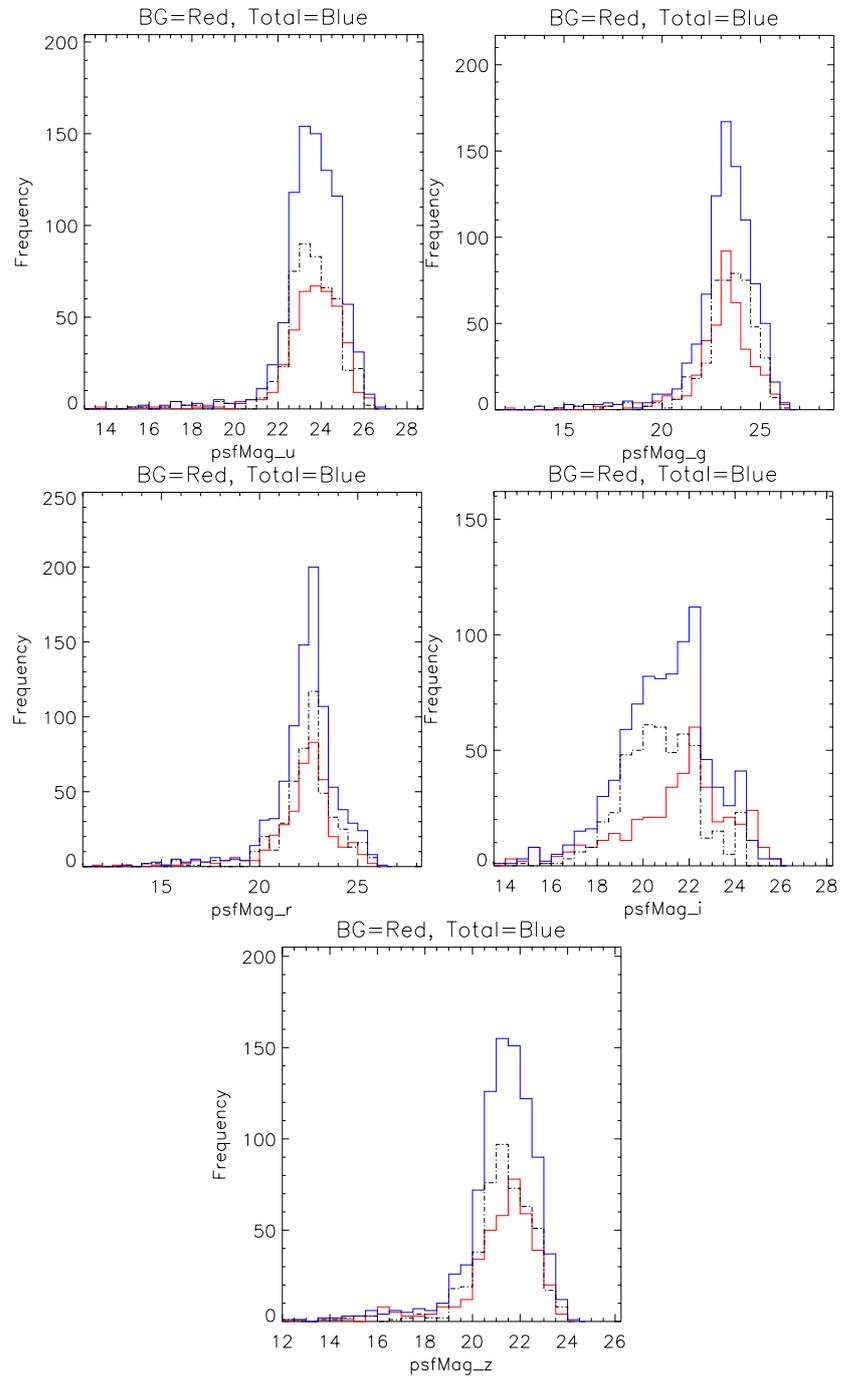


Figure 6.28: Magnitude distribution of SDSS Star-like sources within $5''$ radius around FIRST sources (blue line). We calculated the distribution of background stars (red line) and subtract those to calculate the excess around the FIRST sources (dashed black line)

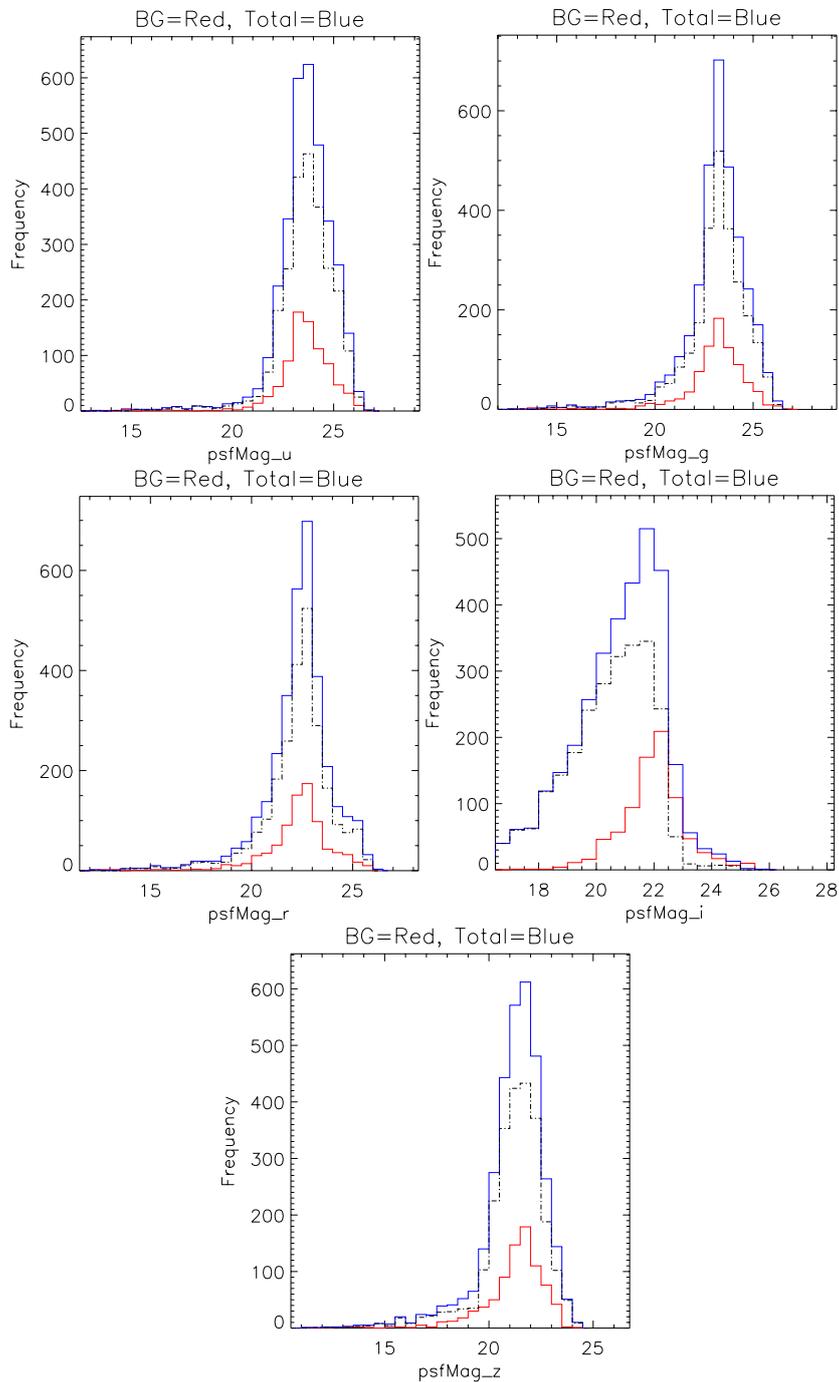


Figure 6.29: Magnitude distribution of SDSS Galaxy-like sources within $5''$ radius around FIRST sources (blue line). We calculate the distribution of background galaxy (red line) and subtract those to calculate the excess around the FIRST sources (dashed black line)

Table 6.4: Dependence of Q_0 from the choice of r_{max} , quantifying the background excess in case of extended and unresolved sources. The SDSS sources are divided into star and galaxies depending on the value of the "type" column in PhotoObjAll. The UKIDSS sources are divided using the value of the MergedClass

r	Q_0 FIRST-SDSS			Q_0 FIRST-UKIDSS	
	Galaxy	Star	Star $i < 21$	Galaxy	Star
1".0	0.22	0.05	0.03	0.26	0.016
2".0	0.26	0.05	0.03	0.32	0.018
3".0	0.28	0.06	0.04	0.34	0.018
4".0	0.29	0.06	0.04	0.37	0.019
5".0	0.30	0.06	0.04	0.39	0.019
6".0	0.32	0.06	0.04	0.42	0.019
8".0	0.37	0.07	0.04	0.47	0.020
10".0	0.43	0.08	0.04	0.53	0.027
15".0	0.58	0.09	0.04	0.66	0.027
20".0	0.73	0.09	0.04	0.79	0.027

where $N_{matches}$ is the total number of sources matched in SDSS within r_{max} from the FIRST sources. Therefore in our method, Q_0 is simply given by:

$$Q_0 = \frac{\sum_m N_{excess}(m)}{N_{centres}} \quad (6.8)$$

As pointed out in McAlpine et al. (2012) in the case of large r_{max} this expression leads to an over-prediction of Q_0 . This effect may result from the tendency for radio-loud AGN to reside in denser environments than normal galaxies. The value of Q_0 depends on the choice of r_{max} but does not depend on the choice of σ_{pos} . Q_0 does not depend on the magnitude in our treatment.

6.8 Reliability of the UKIDSS morphological classification

To estimate the reliability of the morphological classification in UKIDSS, we cross-match the QSOs of the DR7 QSO catalogue with UKIDSS DR10 using a 2".0 cross-match radius. The resulting sample counts a total of 43,441 sources, to 25,052 of them it is correctly attributed a star-like morphology in UKIDSS, while to the remaining 11,879 it is attributed a galaxy-like morphology. The remaining 6,510 are classified as probably star (4,703) and probably galaxies(1,807). The morphological misclassification appears to affect mostly the fainter sources at low

6.8. RELIABILITY OF THE UKIDSS MORPHOLOGICAL CLASSIFICATION 157

Table 6.5: Number of sources with spectroscopic identification, detected in FIRST and in SDSS DR7. The column "type" indicates the morphological classification given to the source in SDSS.

<i>Type</i>	SpecClass	Sources
STAR	Star/qso	8,105
STAR	Galaxy	252
GALAXY	Star/qso	1,222
GALAXY	Galaxy	36,768

redshift, as shown in Fig. 6.30. In Fig. 6.31 we show also the distribution in K magnitude and redshift of the quasars having star-like morphology in UKIDSS.

From this analysis we deduce that about the 30% of the quasars are classified in UKIDSS as galaxy-like sources, but that the majority of these quasars have low redshift. Therefore discarding the sources with galaxy-like morphology it is not likely to discard high- z quasars.

We test the SDSS morphological classification as a function of the radio emission and of the optical colors, considering the sample of 46,347 FIRST sources with spectroscopical counterpart in DR7 SDSS within $1''.5$. In Table 6.5 we compare the morphological (column "type") and the spectroscopical classification (column "specClass") for the whole sample.

We then plot in Fig. 6.32 (left panel) the median SDSS $g-r$ color as a function of the FIRST radio-peak at 1.4GHz, for the 252 sources that are QSOs but have been misclassified as galaxies on the basis of the SDSS morphology type (red line), and for the the QSOs having star-like morphology in SDSS (blue line). The sample of QSOs misclassified as galaxies it is generally redder than the sample correctly classified as QSOs; and it is redder at lower radio-densities. In the right panel of Fig. 6.32, we plot the median i -magnitude of the same samples, again as a function of FIRST radio-peak. We see that the quasars misclassified as galaxies tend to be brighter in the optical for brighter radio-densities.

6.8.1 Efficiency of the morphological classification

In order to have an overall quantitative estimation for the reliability of the star-like morphological classifications in SDSS and in UKIDSS for radio-quasars, we proceed as it follows. From the sample of 68,186 sources detected in FIRST-UKIDSS and SDSS (see section 6.3.1), we select only the 59,093 sources with FIRST-SDSS separation $< 1''.5$. 2,364 of these sources are QSOs of the DR7 QSO catalogue. Looking at the morphological classification of those QSOs, we estimate:

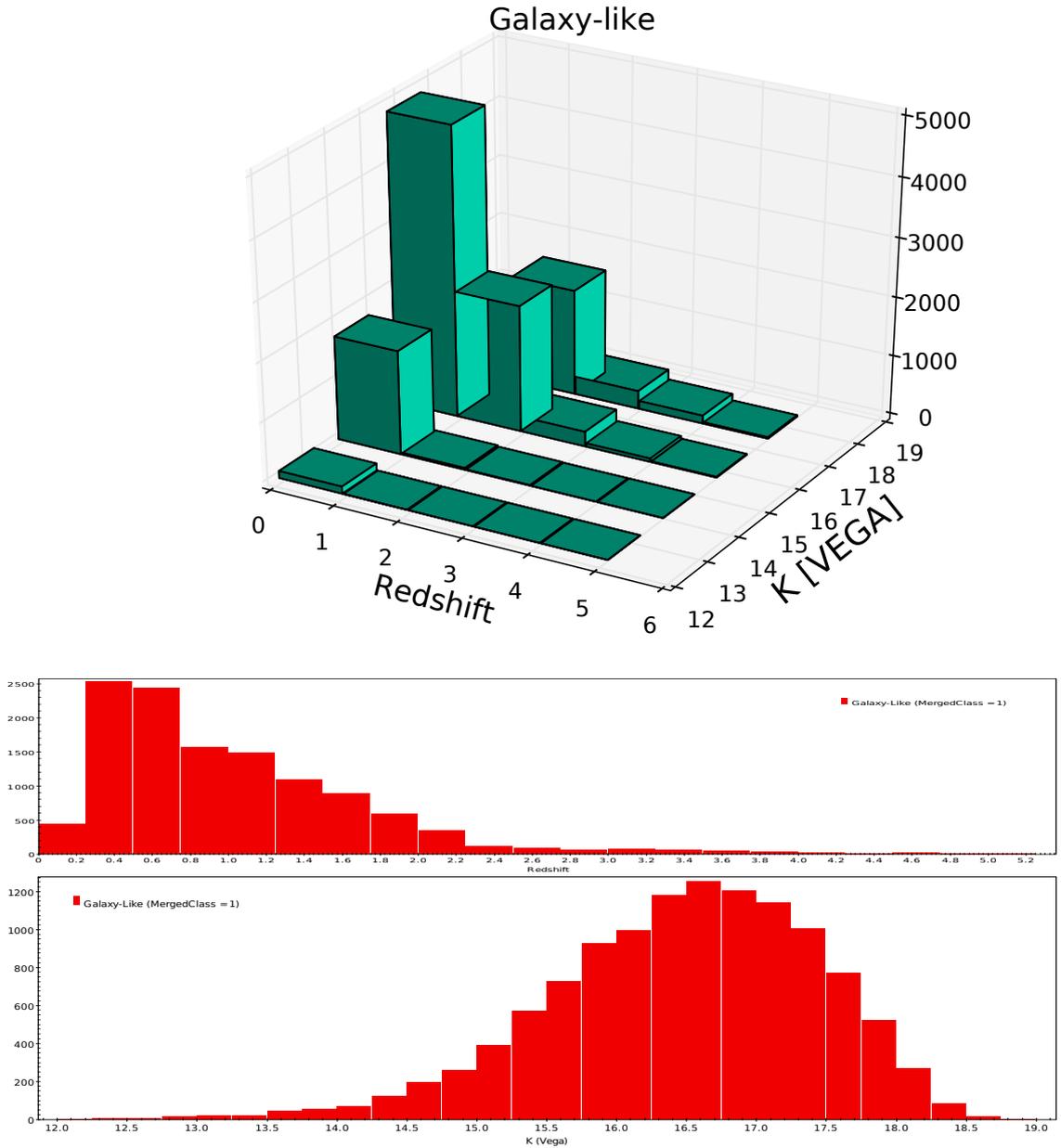


Figure 6.30: To estimate the reliability of the UKIDSS morphological classification for quasars (see section 6.8), we cross-matched the DR7 QSOs catalogue with UKIDSS, obtaining a sample of 43,441 sources. Here we consider the 11,879 quasars that have been misclassified as galaxy-like in UKIDSS on the basis of their morphology. We give the distribution of the sources as function of K(vega) magnitude and redshift.

6.8. RELIABILITY OF THE UKIDSS MORPHOLOGICAL CLASSIFICATION 159

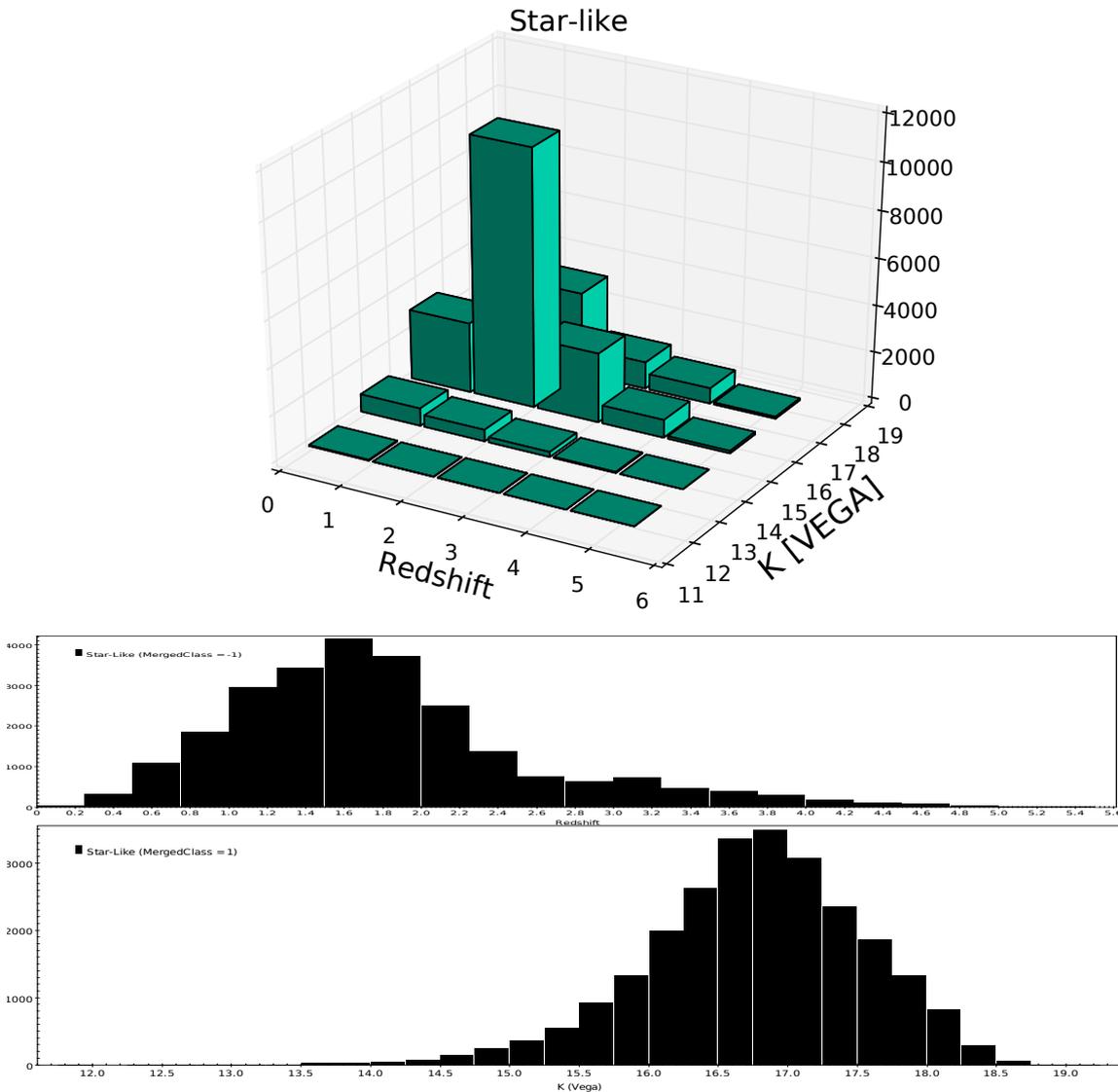


Figure 6.31: To estimate the reliability of the UKIDSS morphological classification for quasars (see section 6.8), we cross-matched the DR7 QSOs catalogue with UKIDSS, obtaining a sample of 43,441 sources. Here we consider the 25,052 quasars that have been classified as star-like in UKIDSS on the basis of their morphology. We give the distribution of the sources as function of K(vega) magnitude and redshift.

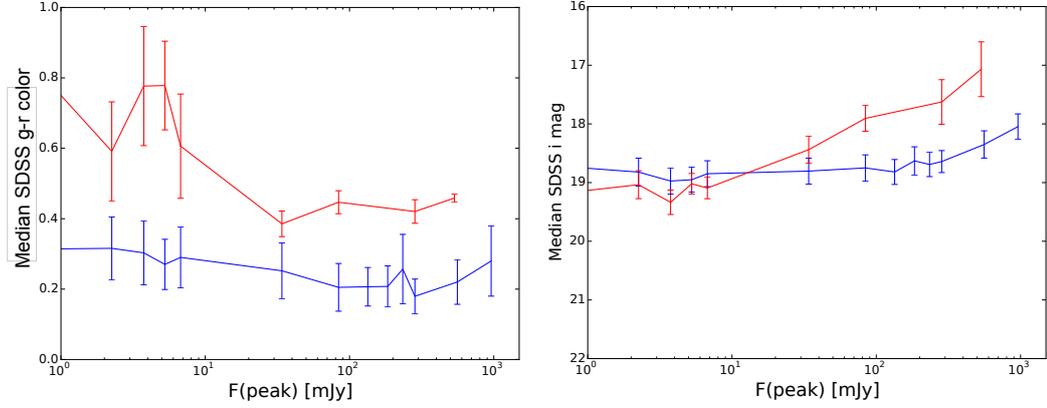


Figure 6.32: Median SDSS g-r color (left panel) and i-magnitude (right panel) for QSOs having star-like (blue lines) or galaxy-like (red line) morphology in SDSS, plotted as function of the FIRST radio-peak at 1.4GHz. The two plots have been obtained averaging magnitude and colors over bins of the radio-peak.

- $P(t6|qso) \sim 2217/2365 \sim 94\%$
- $P(\overline{t6}|qso) \sim (2365 - 2217)/2365 \sim 6\%$
- $P(m4|qso) \sim 1367/2365 \sim 58\%$
- $P(\overline{m4}|qso) \sim (2365 - 1367)/2365 \sim 42\%$

where we have indicated as $P(t6|qso)$ the probability that a quasar is classified as star-like in SDSS; $P(\overline{t6}|qso)$ the probability that it is classified otherwise. Analogously we indicate as $P(m4|qso)$ the probability that a quasar is classified as star-like in UKIDSS and $P(\overline{m4}|qso)$ the probability that it is classified otherwise. In UKIDSS some sources are morphologically classified as "probable star". Considering that 324 of the QSOs of our sample are classified in this way, we have that:

- $P(m4 \cup mp|qso) \sim 1691/2365 \sim 71\%$
- $P(\overline{m4 \cup mp}|qso) \sim (2365 - 1691)/2365 \sim 29\%$

where $P(m4 \cup mp|qso)$ is the probability that a QSO detected in UKIDSS is classified as "star" (m4) or "probable star" (mp). $P(\overline{m4 \cup mp}|qso)$ is the probability that it is classified otherwise.

It is interesting also to estimate the probability that a QSO is classified as a star in UKIDSS when the classification is based on the morphology given in the K-band only (and we indicate it as "m1"). Still considering the subsample of quasars, we have:

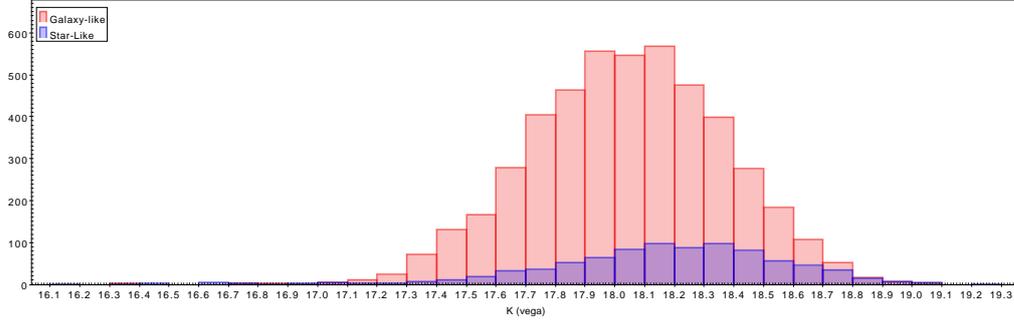


Figure 6.33: K-magnitude distribution of the sources of our catalogue, divided in star-like or galaxy-like, depending on the morphology classification in UKIDSS

- $P(m1|qso) \sim 1275/2365 \sim 54\%$
- $P(\overline{m1}|qso) \sim (2365 - 1275)/2365 \sim 46\%$

resulting that the classification based only on K-band, at least for QSOs, is similarly reliable as the one based on 4 bands.

6.9 Catalogue and Quasar selection

In the previous sections we selected a catalogue of 6,271 radio-infrared sources undetected in the optical, and we demonstrated that the majority of them are likely to be real. Particularly interesting are the brighter-sources detected only in the K-band (see Fig. 6.33), undetected in the optical because they are highly obscured.

We demonstrated that the choice of a tight cross-match radius assures high completeness in the selection of QSOs, and a lower contamination of galaxies. We also demonstrated that the UKIDSS morphological classification is reliable especially for high-z QSOs.

The full catalogue and all the techniques used to select it, are interesting for studies of very red dust-obscured radio-objects. But the catalogue is particularly suited for the selection of obscured high-z QSOs.

The actual selection of QSOs will be part of further careful selection and will be based on WISE colors and on cross matches with other catalogues. Nevertheless in this section we want to demonstrate the potentiality of our catalogue in the selection of QSOs, using the criterium studied by Wright et al. (2010) to select AGNs. The Wright et al. (2010) criterium (see Fig. 6.34 upper plot) is based on WISE colors and it has been demonstrated to be very effective at separating extragalactic sources from stars and brown dwarfs. Therefore we consider the sources of our catalogue that are detected in the $3.4\mu m$, $4.6\mu m$, $12\mu m$, and $22\mu m$ WISE

bands (5,266 sources out of 6,271), and that are classified as star-like or probable-star in UKIDSS. Using the same color-space of the Wright et al. (2010) selection, we verify that 248 of the 424 star-like sources considered, lie within the QSO locus (see Fig. 6.34, middle plot). Considering, then, the 248 sources classified in UKIDSS as "probable star", we see that 194 of them lie in the QSOs locus (see Fig. 6.34, bottom plot). A total of 442 out of 672 sources are therefore candidate red-QSOs.

6.10 Conclusions

- On the basis of the astrometry of UKIDSS and FIRST, our initial choice of the match radius assures that we are 98.76% complete. That means we miss the 1.24% of the authentic UKIDSS counterparts to the FIRST sources. (see section 6.3.1)
- We exclude from our analysis all the sources with SDSS counterpart within $10''$. On the basis of the SDSS astrometry, this assures that $1/(6 \times 10^{22})$ of the sources of our catalogue have a true optical counterpart in SDSS missed for the cross-match radius. (See section 6.3.2). This led the analysis to 8,355 sources.
- Excluding from our sample all the false negative, we end up with 6,271 sources (see section 6.4.1)
- In the analysis of the false positives (see section 6.4.2), we add several check-flags to the sources of our catalogue. In order: 1) We tag the sources detected only in Y with "critical flag" because from the mutual separation analysis we concluded that they are likely false detections.; 2) we tag the sources with UKIDSS post-processing errors, with warning flags numbered from 1 to 3 in agreement with the severity of the error ; 3) we add a "reliable" flag to the sources matched in WISE within $2''$.
- From exploratory data analysis (section 6.6) we find that radio sources tend to reside in denser environments with respect to sources radio-quiet. This effect could affect the reliability of simple cross-matches. Nevertheless empirically we find that this effect is more evident for galaxies.
- We quantify better this effect calculating the likelihood ratio for a sample of FIRST sources matched with UKIDSS and SDSS (section 6.7) and in particular looking at Q_0 , the fraction of true counterparts in a cross-match. We obtain that matching FIRST to SDSS, for galaxies, even at large radius there are counterparts that are likely to be real. The same it is not true for

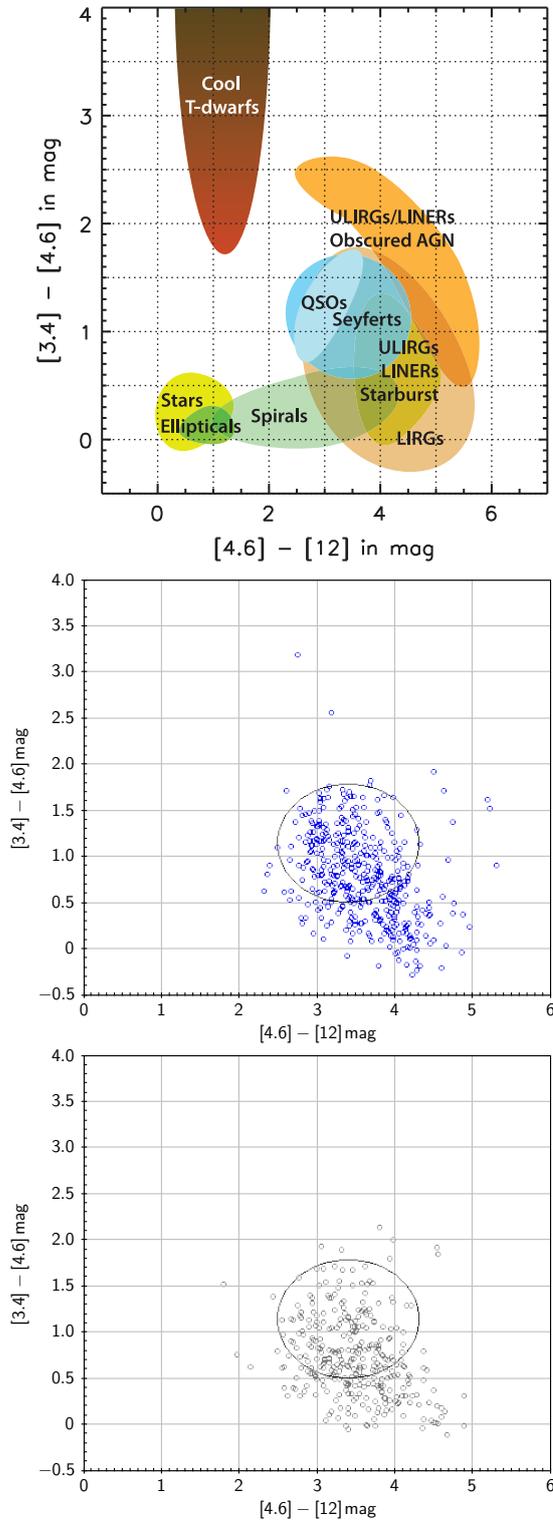


Figure 6.34: Wright *et al.* (2010) (upper plot) used WISE color-space as effective criteria to separate different populations of extragalactic sources from stars and brown dwarfs. We repeat the same color-color plot for the sources of our catalogue classified in UKIDSS as stars (middle plot, the blue open dots) and probable-stars (bottom plot, gray open dots). The majority of our sources lie within the QSO-Seyfert locus)

bright stars, for which the true counterpart lies within $3''.0$. Therefore this analysis justify the need of using tight cross-match radius when selecting radio-QSOs, i.e. the use of larger radius increase the contamination from galaxies. The analysis of Q_0 also suggest that the star-like morphological classification in SDSS it is less reliable for faint sources and that there is a high number of galaxies classified as star-like in UKIDSS.

- The reliability of the UKIDSS morphological classification for QSOs (both RL and RQ) is studied in particular in section 6.8. We conclude that the QSOs misclassified as galaxy-like on the basis of the optical morphology (i.e. in SDSS) are generally redder than QSOs correctly classified as star-like. We suggest that host-galaxy contribution to the colors of the fainter quasars can bias some claims of other authors (White et al. 2007, Helfand et al. 2015) on the fact that radio-detected quasars are redder than radio-quiet quasars. We also verify that the majority of the QSOs misclassified as galaxies on the basis of UKIDSS morphology, have redshift $z < 2$.
- In section 6.9 we stress the interest of the catalogue presented in this Chapter for future research on highly reddered sources. In particular we show that at least 442 sources of our catalogue are AGN candidates. A more careful selection of a list of candidates is the subject of ongoing research from our group.

Chapter 7

Selection of RL QSOs at $z > 4.7$

All of physics is either impossible or trivial. It is impossible until you understand it, and then it becomes trivial.

Ernest Rutherford

7.1 Introduction

As already discussed in Chapter 4, the determination of the QSO Luminosity Function (QLF) is essential to the study active galactic nuclei. In particular the evolution of the shape of the QLF with redshift (changes in the slopes or in the location of the break luminosity) gives one of the fundamental observational constraints to the growth of super-massive black holes (SMBHs) over cosmic ages. A supposed flattening of the bright-end slope for QSOs at $z > 4$ has been widely discussed since it was noticed from the early high-redshift surveys because would give remarkable indications on the cosmological evolution of the Universe. Nevertheless recent works by Shen & Kelly (2012) and McGreer et al. (2013) find no evidence in favour of an evolution of the bright-end slope at high- z . Instead, they find evidence of strong evolution in the break luminosity, as it brightens from $M_{1450}^* \approx -25.4$ at $z = 2.5$ to $M_{1450}^* \approx -27.2$ at $z = 5$.

In the Shen & Kelly (2012) and McGreer et al. (2013) studies, the percentage of radio-quiet quasars overwhelm the radio-quiet population, that is not surprising considering that the radio-loud fractions seems to decrease at high- z (Jiang et al. 2007). In the SDSS 5th Quasar Catalog (Schneider et al. 2010) only 495 QSOs (out of the 105,783) have $z > 4.4$, and of them only 24 are radio-detected in FIRST ($S_{1.4GHz} > 1\text{mJy}$), i.e. $< 5\%$. All of these 24 sources are brighter than $i[AB] = 20.2$ due to the SDSS spectroscopic flux limit.

The study of the Radio Loud Quasar Luminosity Function (RL QLF) at high redshift could give crucial indications of differences if any, between the two populations. In other words, the RL QLF do not need to have the same slope as the entire population of QSOs, and different trends in the evolution of the slope could point to different density evolutions of RL and RQ populations. Nevertheless, the small number of know RL QSOs at $z > 3.5$ renders difficult the determination of unbiased RL QLF, and there are just a few studies in the literature on RL QLF at $z > 3.5$ (McGreer et al. 2009; Tuccillo et al. 2015). Both studies show indications of a flattening of the slope at $3.5 \leq z \leq 4.4$. But there are no other recent studies for the RL QLF at $z > 4.4$. In this range of redshift is therefore needed to obtain a reliable estimation of the QLF, increasing the population of known RL QSOs and extending the selection at fainter optical luminosities. The identification of an unbiased optical-faint sample of high- z RL QSOs, is fundamental to determine the faint-end slope of the optical luminosity function. In this way we will be able to test for the RL population, the results found by McGreer et al. (2013) of an evolving break luminosity.

In this chapter we present a list of 21 candidates to be RL quasars with $4.7 \leq z \leq 5.1$. All these objects are point-like sources in the optical, and they do not have previous spectroscopy. We selected these candidates among a sample of sources detected in the radio (FIRST), in the optical (SDSS DR10) and in the infrared (UKIDSS Large Area Survey DR10, and the UKIRT Hemisphere Survey). The selection methodology is based both on the use of the recent effective color-color optical criteria established in McGreer et al. (2013) that they claim to be 79% efficient, and on additional NIR and MIR criteria to increase the efficiency of the selection. With the identification of these objects we will be able to calculate the RL QLF at $z \sim 5$.

7.2 Preselection of the sample

The 2014-version of the FIRST catalog includes 946,432 detections, from which we excluded the ones having probability $> 35\%$ (i.e. $\sim 9\%$) of being spurious (using the *sidelobe probability* given in the FIRST catalog). We cross-matched the resulting FIRST sources with the nearest UKIDSS DR10 counterparts within $3''.0$ (detected at least in the J-band); when no counterpart was found, we cross-matched the radio sample with the non-public Ukirt Hemisphere Survey (UHS) using the same cross-match radius. This way we obtained an initial radio-infrared sample of 108,513 sources. Subsequently, the sample was cross-matched with the SDSS DR10 catalog, again searching for the nearest counterpart to the radio-position within $3''.0$. This way we selected 87,094 sources.

Since the motivation of this selection was to search for RL QSOs in a range of

Table 7.1: Steps in the pre-selection process discussed in section 7.2

Pre-selection criterion	Selected sources
Starting FIRST-UKIDSS-SDSS sample ($\leq 3''.0$)	87,094
type_i = STAR	13,958
Without spectra in DR10 QSO catalog	11,950
Without spectra in DR7 QSO catalog	10,123
Without spectra in SDSS DR10 SpecObj view	9,207
Without Mag error > 0.2 in all bands	8,748
Exclude sources with 'fatal' error flags	7,776

redshift where they are extremely rare, we used a larger cross-match radius than the one used in Chapter 2, in order to increase the completeness of our selection. The use of a larger cross-match radius reduces the efficiency of the selection (as discussed in Chapter 6), but we will take into account the radio-optical and the radio-infrared distances (see section 7.3) as criteria to establish the reliability of each quasar-candidate. The SDSS morphological type is set as "galaxy" or "star", using the photometrical information in all bands in which the object is detected. The QSOs Ly- α emission line at $z > 4.7$ (i.e. $\lambda > 6931\text{\AA}$) has almost completely left the SDSS r -filter band, and the u and g band-filters are entirely occupied by the Ly- α forest. For this reason we considered only the morphological information given in SDSS for the i -band.

Selecting only the star-like sources in the i -band from our initial sample, we obtain 13,857 sources. We excluded all the sources with unreliable photometry, i.e. all the ones tagged in SDSS with 'fatal' error flags ('BRIGHT', 'SATURATED', 'EDGE' or 'BLENDED') or with *magnitude error* > 0.2 in all bands. Finally we pre-selected sources without available spectroscopical identification in literature. For this final check we searched for spectroscopical identification of the sources of our sample in the DR10- and the DR7-QSOs catalogues, in the SDSS-DR10 and in the data from the NASA Extragalactic Database (NED).

This pre-selection process, summarized in Table 7.1, left us with 7,776 star-like sources without spectroscopic identification. All the sources are detected in the radio (FIRST) in the optical (SDSS) and at least in the J -band of UKIDSS or UHS. In Fig. 7.1 we give the distribution of the distances between radio-infrared (red line) and radio-optical (blue line) positions for our pre-selected sample.

7.3 Candidates selection

In McGreer et al. (2013) they adopted a new and very effective optical-infrared color-color criteria for the selection of QSOs in redshift range $4.7 \leq z \leq 5.1$. The

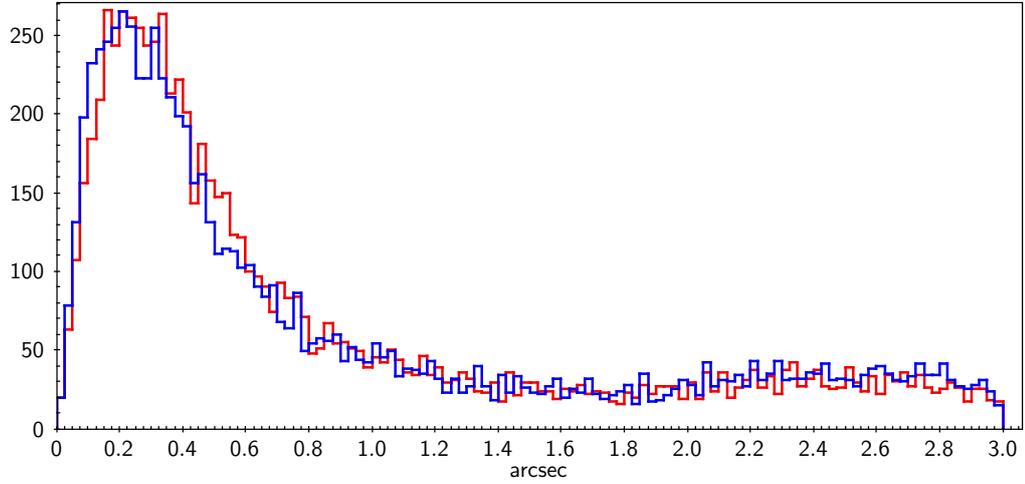


Figure 7.1: *Distribution of the distances between radio-infrared (red line) and radio-optical (blue line) positions for the sample of 7,776 sources pre-selected in section 7.2*

redshift range and the magnitude limit (down to $i[AB] \leq 22$) of their selection were chosen to examine the quasar LF at intermediate redshifts between the precise determination up to $z \sim 4$ (obtained using data from SDSS, Richards et al. 2006), and the results found by Jiang et al. (2008) and Willott et al. (2010) at extremely high- z , i.e. $z \sim 6$. The selection method used by McGreer et al. (2013) is an improved version of the criteria adopted in SDSS to target high- z QSOs candidates, that was found to be extremely efficient. In fact, McGreer et al. (2013) after spectroscopic identification of their 92 candidates, found that 71 of them were high- z QSOs in the desired range of redshift, i.e. an efficiency of $\sim 79\%$.

Since up to the date there is no evidence of dramatic changes in the colors of the RL QSOs population respect to the RQ, we adopted the McGreer et al. (2013) color-color criteria to select RL QSOs $4.7 \leq z \leq 5.1$ candidates from our pre-selected sample of 7,776 sources. This way we obtained a list of 13 candidates (8 detected in UKIDSS and 5 in UHS). In Table 7.2 we list the whole selection-process applied to select these candidates.

We selected a second sample of "lower priority" candidates relaxing the fifth of the conditions listed in 7.2 (this condition is indicated in the Table as criterium "5 loose"), obtaining other 8 candidates (7 detected in UKIDSS and 1 in UHS). Using the sample of high- z QSOs of the DR7 QSO Catalog (see Fig. 7.2) we verified that the "5th loose" criterium is still effective in selecting high- z QSOs but with a higher contamination of $z < 4.4$ QSOs.

Table 7.2: Color-color selection steps for RL quasar candidates in the redshift range $4.7 \leq z \leq 5.1$, see section 7.3.

Selection criterion	Selected sources
Sample pre-selected detected in J	7,776
Crit. 1 : $g - r > 1.8$	638
Crit. 2 : $r - i > 1.2$	342
Crit. 3 : $i - z < 0.55$	153
Crit. 4 : $i - z < 0.625((r - i) - 1)$	55
Crit."5 loose": $i - J_{AB} < ((r - i) - 1) + 0.76$	21
Crit. 5 : $i - J_{AB} < ((r - i) - 1) + 0.56$	13

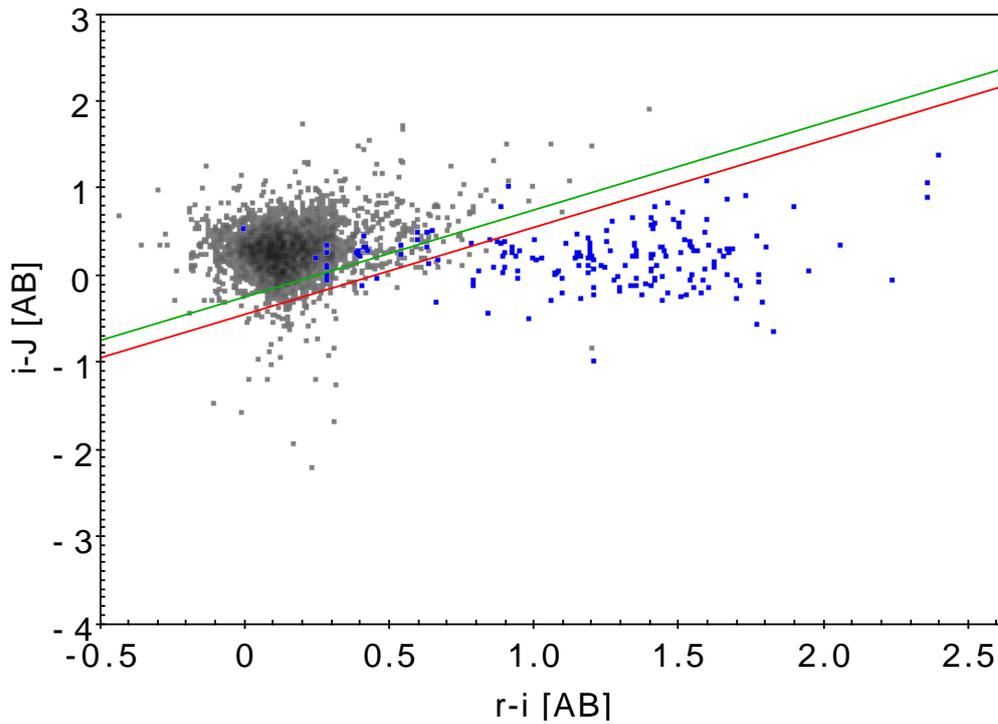


Figure 7.2: High- z QSOs of the DR7 QSO Catalog matched in the DR10 UKIDSS catalogue in color-color optical-infrared space. Gray squares represent the colors of the 2841 QSOs with $3 \leq z \leq 4.4$, red squares refer to the 183 QSOs with $z > 4.4$. The red line refers to the 5th color-color criterium listed in Table 7.2: $i - J_{AB} = ((r - i) - 1) + 0.56$. The green line indicates the "5th loose" criterium: $i - J_{AB} = ((r - i) - 1) + 0.76$.

Table 7.3: RL QSOs candidates at $4.7 \leq z \leq 5.5$ detected in UKIDSS

RA (J2000) (1)	DEC	$(W1 - W2)$ (VEGA) (2)	$\text{Sep}_{(F,S)}$ (arcsec) (3)	$\text{Sep}_{(F,U)}$ (4)	r_{AB} (5)	$S_{1.4GHz}$ (mJy) (6)	NOTE (7)
(Higher priority)							
03:04:34.70	+06:17:42.0	0.22	0".46	0".30	20.57	13.9	GTC
07:41:54.72	+25:20:29.6	0.97	0".08	0".08	20.49	2.1	
08:39:43.29	+05:37:42.7		2".83	2".71	22.53	13.5	
09:25:02.50	+05:28:42.7	0.59	0".79	0".62	22.69	1.2	
11:00:34.13	-02:40:08.0	-0.25	0".59	0".84	22.69	6.6	
11:33:00.26	+12:29:33.1	-0.31	0".34	0".18	22.79	2.9	
12:53:19.78	+23:54:30.0	0.11	0".14	0".13	20.78	1.3	WHT
22:24:55.72	+04:46:18.3	-0.34	0".60	0".59	22.07	1.2	GTC
(Loose criteria)							
00:10:59.58	+05:00:24.6		0".38	0".30	22.54	1.4	GTC
01:45:23.88	+08:39:26.1		2".52	2".01	22.26	1.5	GTC
08:14:37.80	+26:27:13.9	-0.17	0".04	0".15	22.99	20.2	
09:31:51.05	+34:08:42.5	0.29	0".20	0".26	22.45	3.7	
11:01:04.69	+03:01:39.3	-0.08	0".96	1".18	22.52	3.8	
13:13:03.76	+13:09:56.3		1".78	1".68	22.86	8.2	
13:47:08.96	+29:52:17.7	0.19	0".61	0".40	22.83	1.7	

The columns give the following: (1) SDSS J2000 coordinates; (2) $(W1-W2)$ WISE-color; (3) Radial separation FIRST-SDSS; (4) Radial separation FIRST-UKIDSS; (5) SDSS dereddened r magnitude; (6) FIRST peak radio flux density; (7) "WHT" if the sources has been observed at the William Herschel Telescope (WHT) in June 2014, "GTC" if it has been observed at the Gran Telescopio CANARIAS in August 2015

7.3.1 Candidate list and preliminary observational results

The spectroscopic follow-up of the 21 candidates selected in the previous section is still in progress. A total of 3 candidas were observed at the William Herschel Telescope (WHT) in June 2014 and another 8 observed at the Gran Telescopio CANARIAS (GTC) in August 2015. The results of these observations will be discussed in the following sections. In Table 7.3 we give the list of the 15 candidates detected in UKIDSS, the first 8 (higher-priority) satisfy all the criteria of Table 7.4, while the remaining 7 satisfy the "loose" criteria only. In Table 7.4 we list the 6 candidates detected in UHS with the same criteria, 5 of them with "higher-priority".

7.3.2 WHT observations

We obtained long-slit spectroscopy with service time using the ISIS instrument of the William Herschel Telescope (WHT) in one single night in June 2014, for 3 candidates (see columns 7 of Table 7.3 and Table 7.4), all of them having "higher priority". We used 3600 seconds time-exposure for each source and the seeing

Table 7.4: *RL QSOs candidates at $4.7 \leq z \leq 5.1$ detected in UHS*

RA (J2000) (1)	DEC (2)	($W1 - W2$) (VEGA) (2)	$\text{Sep}_{(F,S)}$ (arcsec) (3)	$\text{Sep}_{(F,U)}$ (4)	r_{AB} (5)	$S_{1.4GHz}$ (mJy) (6)	NOTE (7)
(Higher priority)							
15:32:42.83	+42:55:53.1	-0.06	2''75	2''30	22.99	2.22	GTC
16:26:38.87	+08:35:15.4		0''71	0''94	23.08	1.79	GTC
17:06:37.46	+57:50:11.0		0''93	1''11	23.17	1.31	WHT
17:24:13.26	+40:17:00.7	-0.38	0''41	0''66	22.19	7.44	WHT
20:59:57.40	+08:11:40.1		2''22	2''34	22.15	1.66	GTC
(Loose criteria)							
16:09:06.13	+34:01:47.2	-0.131	0''95	1''02	23.55	1.03	GTC

The columns give the following: (1) SDSS J2000 coordinates; (2) ($W1 - W2$) WISE-color ; (3) Radial separation FIRST-SDSS; (4) Radial separation FIRST-UHS; (5) SDSS dereddened r magnitude; (6) FIRST peak radio flux density ; (7) "WHT" if the sources has been observed at the William Herschel Telescope (WHT) in June 2014, "GTC" if it has been observed at the Gran Telescopio CANARIAS in August 2015

was < 1.2 arcsec FWHM.

The data were reduced using standard IRAF routines, and the resulting 3 spectra are shown in Fig. 7.3. For these sources the spectra are too noisy to allow a clear identification, nevertheless no signs of prominent emission are evident, and we consider that these sources are likely to be galaxies or stars.

7.3.3 GTC observations

We obtained B-priority time to use OSIRIS at the GTC to obtain long-slit spectroscopy for the remaining 18 candidates of our selection. In August 2015, 8 of the 18 sources proposed were observed (see columns 7 of Table 7.3 and Table 7.4). Careful reduction of these data is needed and is in progress, although one of the spectra is clearly a bright RL QSOs at $z \sim 4.9$. This spectrum is shown in Fig. 7.4.

7.4 Conclusions

In this Chapter we have presented a selection of RL QSOs in the redshift range of $4.7 \leq z \leq 5.1$. For this selection we adopted the color-color candidate-criteria effectively used in McGreer et al. (2013), in that work assessed to have a 79% efficiency. Although our observational campaign is not complete and the data-reduction is still in progress, the first observations seem to suggest that the same technique that is very effective to select RQ QSOs, is not as effective for RL

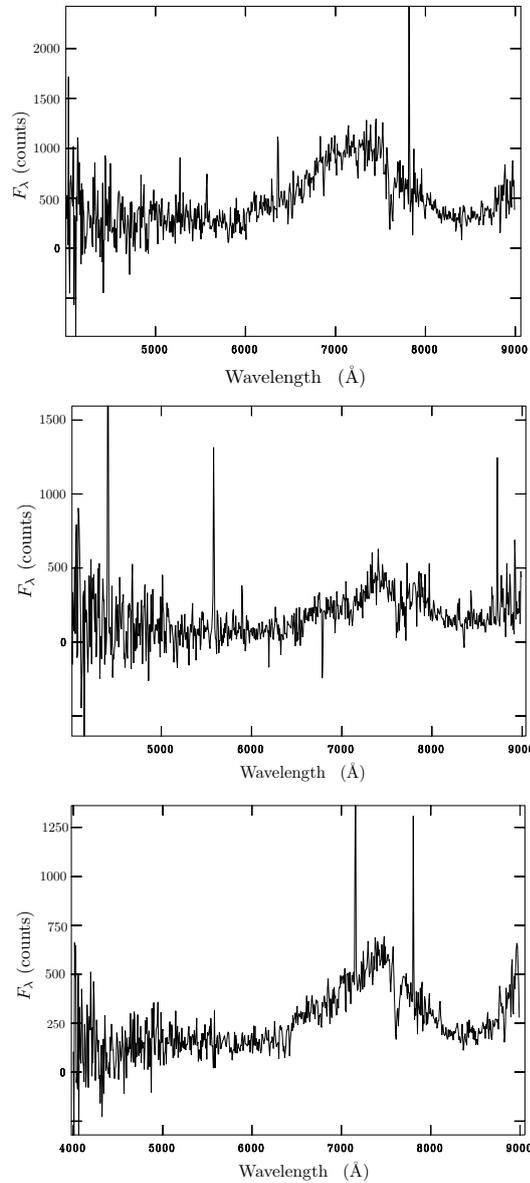


Figure 7.3: Spectra of the 3 RL quasar candidates observed at the WHT in June 2014 and not identified as QSOs. The candidates are listed in Table 7.3 and Table 7.4 and indicated in the NOTE columns as "WHT". The upper source has RA=12:53:19.78, DEC=+23:54:30.0; the middle source has RA=17:06:37.46, DEC=+57:50:11.0; the bottom source has RA=17:24:13.26, DEC= +40:17:00.7

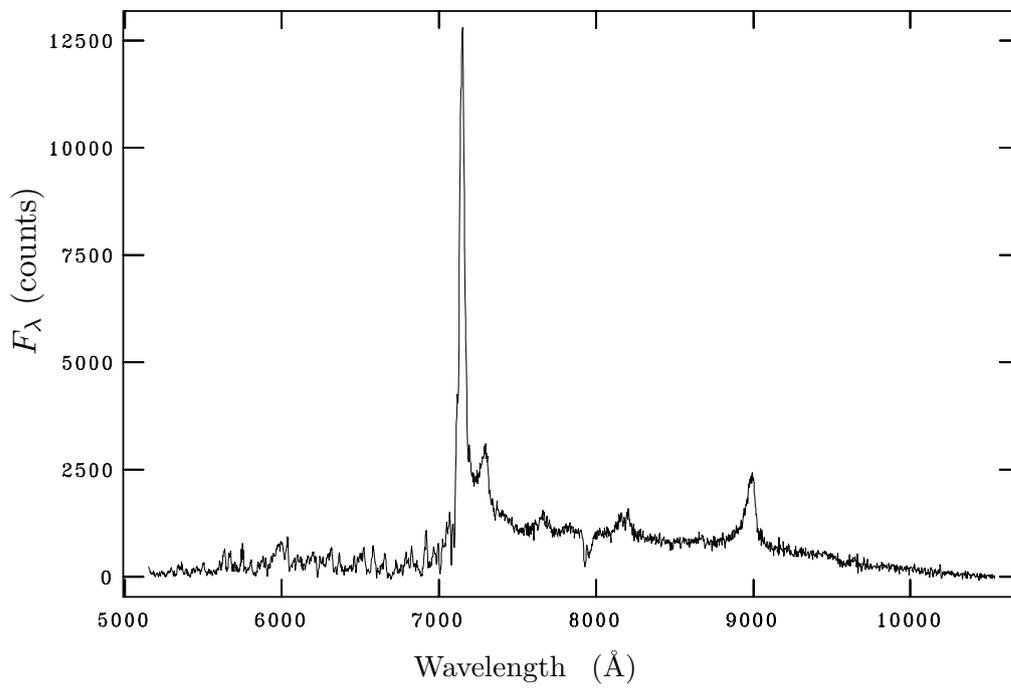


Figure 7.4: *RL QSOs at $z \sim 4.9$ observe at the GTC in August 2015. Part of the quasar candidates in the range $4.7 \leq z \leq 5.1$ listed in Table 7.3. The source has $RA=03:04:34.70$, $DEC=06:17:42.0$*

QSOs. In fact, none of the 3 sources observed at the WHT is a high- z QSO. The data of the 8 sources observed at the GTC still need careful reduction, but at the moment only one is clearly a QSO at $z \sim 4.9$.

In total 8 (3 at the WHT and 5 at the GTC) of the observed sources pass all the McGreer et al. (2013) color-color criteria. Excluding from these 8 sources the 2 that have radio-optical distance $> 1''.5$ (and thus not reliable), we have observed 6 candidates for which the efficiency predicted by the McGreer et al. (2013) criteria is of the 79%. We would expect therefore at least 4 high- z QSOs identifications, that is not what we have obtained so far. The reasons for this selection-difference need further research and they also point to the need to use infrared information for the selection of high- z QSOs. In fact, in the columns 2 of Table 7.3 and Table 7.4, for the sources detected in WISE (within $2''.0$) we give their $W1$ - $W2$ colors. It has been showed that QSOs are likely to have $W1 - W2 > 0.5$ (Wright et al. 2010), and none of the observed sources pass this threshold.

A possible explanation between the difference in the selection-strategy between RL and RQ QSOs could be found in a larger contamination of radio-galaxies that have the same space color of high- z QSOs. The scarcity of radio-loud quasars in our selection could also indicate a dramatic change of color space for the population of radio-loud at extremely high redshift.

Ultimately, if the space density of the radio-loud QSOs population declines at high redshift much more rapidly than that of the radio-quiet QSOs, our results could even indicate a greater intrinsic scarcity of radio-loud quasars at very high redshift. This result would be in agreement with the decline of the radio-loud fraction at high- z found by several authors (Jiang et al. 2007, Krawczyk et al. 2015) and could provide fundamental clues on the causes of the radio-loud phenomenon itself.

Chapter 8

Conclusions and future prospective

The journey of a thousand miles begins with one step.

Lao Tzu

In this thesis we present the results of a research on the selection, identification and study of high-redshift radio-loud quasars. These are rare objects and to find them we need large area surveys of sufficient depth to cover a large enough comoving volume. Modern large-areas surveys of like FIRST, SDSS, UKIDSS, WISE provide efficiently millions of objects from which it is potentially possible to extract a valuable amount of information. However, they require the use of efficient data mining techniques to deal with such amount of data, and sensitive selection algorithms to reduce and accurately estimate the selection biases.

Notwithstanding the great and well known success of the SDSS QSO Catalogue, in chapters 2 and 3 we demonstrate that there is still space for improvement of the selection techniques. We compared our methodology with the SDSS target algorithm and we obtain that, although the efficiencies of the two methodologies are comparable, our neural-network algorithm detects $\sim 97\%$ of the $3.6 \leq z \leq 4.4$ RL QSOs, while SDSS only detect $\sim 85\%$ of them.

The improvement of data-mining techniques on large-surveys can provide major breakthroughs for studies on AGNs, especially when complemented with more detailed data obtained with non-survey telescopes. In this way, after having obtained optical spectra for high- z QSO candidates and therefore assessed the efficiency and completeness of our NN-selection, in Chapter 4 we use this new data and we carry out the most accurate determination of the optical luminosity function for these objects known at the time.

We determine the optical luminosity function for radio-loud QSOs in two redshift bins, $3.6 \leq z < 4.0$ and $4.0 \leq z \leq 4.4$, and measure the total comoving density of QSOs in these two redshift ranges, obtaining a result consistent with

other results and models from literature.

We test some determinations of the radio-loud fraction and use those determinations to estimate the total comoving density of QSOs. We compare the derived density of QSOs with precise determinations for the whole population of quasars from literature. The results we obtained give indications on the correct estimation of the radio-loud fraction at high redshift.

We determine the bright-end slope of the luminosity function in two bins of redshift finding values of the slope consistent with other determinations for the radio-loud and radio-quiet populations. Our results suggest a similar evolution for both radio-loud and radio-quiet populations in this range of redshift. Our results can be also interpreted as suggestive of a flattening of the bright-end slope from $z \sim 2$ to $z \sim 4$, for the radio-loud population only. If confirmed, this would imply an evolution of the density of super-massive black holes associated with radio-loud QSOs, in the sense that they were more abundant at $z \sim 4$. However, to clarify the evolution of the RL population relative to that of the whole population of QSOs, more observational constraints are needed, especially at redshifts above 4.

The identification of a larger sample of RL QSOs, can have fundamental benefits to improve statistical studies on their population, but also for more detailed studies on objects of particular interest. Thus, partially exploiting the spectroscopic data obtained, in Chapter 5 we select a sample of 22 high- z RL broad absorption line (BAL) QSOs and 22 non-BAL QSOs within $3.6 \leq z \leq 4.8$, extending the known number of objects of this class. We observed the brightest objects of this sample in several radio frequencies and with that data we tested the two main scenarios that have been proposed to explain the existence of these objects.

The shapes of the radio spectra allowed us to evaluate statistically the radiative age of the sources obtaining results that do not suggest a particular younger radio phase for BAL QSOs with respect to non-BAL objects, even in this redshift range. Nevertheless, half of both samples can be classified as High Frequency Peakers (HFP), which is interesting because they represent a rapidly transient radio phase and are rarely found. We derived the spectral index for the two samples, finding that a preferred equatorial orientation is present among BAL QSOs, with a more significant discrepancy with respect to previous studies at lower redshifts. This could indicate that, among young radio sources, the outflows responsible for the BAL features can have larger angles with respect to the jet axis than in older radio sources.

Data mining on large-area surveys not only involves the improvement of already existing selection techniques, but also implies the need to face new problems and subsequently contrive new techniques to deal with them. In this way, in Chapter 5, in order to select a sample of elusive and potentially interesting population

of extremely red quasars undetected in the optical, we crossed-matched data from FIRST, SDSS, UKIDSS and WISE.

We developed new algorithms and analysis techniques to discard false matches and to obtain a final catalogue of 6,271 radio-sources, extremely red and that are likely to be real.

From exploratory data analysis we found that radio sources tend to reside in denser environments with respect to sources not radio-active, and that effect could have severe consequences on the reliability of simple cross-matches.

We studied the reliability of the UKIDSS and SDSS morphological classification for QSOs (both RL and RQ) as a function of the radio density, optical and NIR colors. We concluded that host-galaxy contribution to the colors of the fainter quasars could bias some claims of other authors on the fact that radio-loud quasars get redder as the radio-flux density declines.

However, we found clues that something changes in the colors of RL quasars at very high redshift. In fact, in Chapter 7 we present the selection of RL QSOs in the redshift range of $4.7 \leq z \leq 5.1$, where we used highly efficient color-color techniques already used in literature to select radio quiet quasars in the same range of redshift. Although not all the candidates have been observed and the data-reduction for the majority of them is still work in progress, the first results seem to suggest that the same technique that is very effective to select RQ QSOs, is not as effective for RL QSOs.

The reasons for this difference can be explained by the results of Chapter 6: a larger contamination of galaxies in the selection. Alternatively, the scarcity of radio-loud quasars in that selection could be explained in terms of a dramatic change of color for the population of radio-loud at extremely high redshift, not necessarily towards redder colors. Those results could even indicate a greater intrinsic scarcity of radio loud objects at high redshift. In fact, as mentioned, in Chapter 5, we found in both BAL and non-BAL QSOs an unusual high percentage of HPF objects, that are considered to be the youngest radio-sources. Therefore it may be that the radio-loud population declines more rapidly than the radio-quiet to extremely high redshift, also justifying the decline of RLF found by other authors at these redshifts. This could thus provide clues on the causes of the radio-loud phenomenon itself, which could, for example, be connected with episodes of major merging which are rarer in a young universe.

Ultimately, the work presented in this thesis has shed some light on the properties of the RL-QSOs at high- z and on the great potentiality of modern surveys in selecting sample of quasar candidates. The natural follow-up to the research presented in this thesis involve the conclusion and extension of some works here presented. The use of the new generation of large-area surveys to improve the data mining techniques for the selection of RL QSOs, and the use of ground telescopes to further explore the relation between RL and RQ quasars.

In particular we aim to:

- Completing the observation and the analysis of the optical spectra for the list of quasar-candidates given in Chapter 7, will allow us to determine the faint-end slope of the optical luminosity function for RL QSOs at $z \sim 5$. In this way we will be able to determine the break-luminosity and compare it with the one recently found for the RQ population of QSOs.
- Compare the efficiency of the selection methods used for high- z RQ QSOs and the results of our selections, we will also be able to evaluate any possible systematic difference in the brand-band colors of the RL and RQ QSOs population at high- z .
- The new generation of wide surveys in the optical and in the near infrared, as Vista Hemisphere Survey (VHS), the Dark Energy Survey (DES), and (J-PAS) will allow us to improve and reduce the biases in the selection of RL QSOs. Since, at the present, few photometric surveys are able to reach the necessary depths to observe the comoving volume where high- z QSOs are expected to be found. Data from these surveys will finally allow us to place tight constraints on the mass and luminosity functions of RL QSOs and massive galaxies, thus enabling us to obtain some clues on the formation and co-evolution of these objects.
- Even if, after the improvement of the selection strategies and the use of larger area surveys, RL QSOs will be confirmed to appear exceedingly rare at $z \geq 5$ this results may provide substantial support that RL and RQ quasars are tied to different evolutionary paths.
- The nature of the population of extremely red type 1 quasars, studied in Chapter 6, is still a matter of debate. This issue is placed into a wider context of the "black hole -galaxy co-evolution", which incorporates a set of recently found connections between nuclear activity and properties of the host galaxy. It has been shown that, relativistic plasma jet and feedback effect associated to AGN-induced outflows can represent an important process in galaxy evolution. Therefore, understanding the interplay between red RL-QSOs and galaxies up to high-redshift could give some indications to assess the processes of formation and evolution of galaxies and supermassive black holes. The Low-Frequency Array for Radio Astronomy (LOFAR) and the new generation of radio-surveys will allow more detailed studies of the radio emission along with analytic studies on the environment of the radio loud AGNs

References

- Abazajian K. N., et al., 2009, ApJS, 182, 543
- Alam S., et al., 2015, preprint, (arXiv:1501.00963)
- Allen J. T., Hewett P. C., Maddox N., Richards G. T., Belokurov V., 2011, MNRAS, 410, 860
- Ambartsumian V. A., 1971, in O'Connell D. J. K., ed., Study Week on Nuclei of Galaxies. p. 9
- Antonucci R. R. J., Miller J. S., 1985, ApJ, 297, 621
- Avni Y., Bahcall J. N., 1980, ApJ, 235, 694
- Baars J. W. M., Genzel R., Pauliny-Toth I. I. K., Witzel A., 1977, A&A, 61, 99
- Ball N. M., Brunner R. J., 2010, International Journal of Modern Physics D, 19, 1049
- Ball N. M., Brunner R. J., Myers A. D., Tchong D., 2006, ApJ, 650, 497
- Baloković M., Smolčić V., Ivezić Ž., Zamorani G., Schinnerer E., Kelly B. C., 2012, ApJ, 759, 30
- Banerji M., McMahon R. G., Hewett P. C., Alaghband-Zadeh S., Gonzalez-Solares E., Venemans B. P., Hawthorn M. J., 2012, MNRAS, 427, 2275
- Barnes J. E., Hernquist L., 1996, ApJ, 471, 115
- Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
- Becker R. H., Gregg M. D., Hook I. M., McMahon R. G., White R. L., Helfand D. J., 1997, ApJ, 479, L93
- Becker R. H., White R. L., Gregg M. D., Brotherton M. S., Laurent-Muehleisen S. A., Arav N., 2000, ApJ, 538, 72

- Benn C. R., Vigotti M., Pedani M., Holt J., Mack K.-H., Curran R., Sánchez S. F., 2002, *MNRAS*, 329, 221
- Bentz M. C., et al., 2009, *ApJ*, 705, 199
- Bernal J., 1988, National Institute of Standards and Technology Technical Note, 1252
- Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, *MNRAS*, 362, 25
- Bischof O. B., Becker R. H., 1997, *AJ*, 113, 2000
- Blandford R. D., 1992, in Filippenko A. V., ed., *Astronomical Society of the Pacific Conference Series Vol. 31, Relationships Between Active Galactic Nuclei and Starburst Galaxies*. p. 455
- Bolton J. S., Haehnelt M. G., Warren S. J., Hewett P. C., Mortlock D. J., Venemans B. P., McMahon R. G., Simpson C., 2011, *MNRAS*, 416, L70
- Borne K., 2009, preprint, ([arXiv:0911.0505](https://arxiv.org/abs/0911.0505))
- Boroson T. A., 2002, *ApJ*, 565, 78
- Boyle B. J., Shanks T., Croom S. M., Smith R. J., Miller L., Loaring N., Heymans C., 2000, *MNRAS*, 317, 1014
- Bridle A. H., Perley R. A., 1984, *ARA&A*, 22, 319
- Brotherton M. S., van Breugel W., Smith R. J., Boyle B. J., Shanks T., Croom S. M., Miller L., Becker R. H., 1998, *ApJ*, 505, L7
- Brotherton M. S., Tran H. D., Becker R. H., Gregg M. D., Laurent-Muehleisen S. A., White R. L., 2001, *ApJ*, 546, 775
- Bruni G., et al., 2012, *A&A*, 542, A13
- Bruni G., Dallacasa D., Mack K.-H., Montenegro-Montes F. M., González-Serrano J. I., Holt J., Jiménez-Luján F., 2013, *A&A*, 554, A94
- Bruni G., Mack K.-H., Montenegro-Montes F. M., Brienza M., González-Serrano J. I., 2015, preprint ([arXiv:1507.08533](https://arxiv.org/abs/1507.08533))
- Brusa M., et al., 2009, *ApJ*, 693, 8
- Burbidge G. R., Burbidge E. M., Sandage A. R., 1963, *Reviews of Modern Physics*, 35, 947

- Capetti A., Balmaverde B., 2005, *A&A*, 440, 73
- Carballo R., González-Serrano J. I., Montenegro-Montes F. M., Benn C. R., Mack K.-H., Pedani M., Vigotti M., 2006, *MNRAS*, 370, 1034
- Carballo R., González-Serrano J. I., Benn C. R., Jiménez-Luján F., 2008, *MNRAS*, 391, 369
- Ciliegi P., Zamorani G., Hasinger G., Lehmann I., Szokoly G., Wilson G., 2003, *A&A*, 398, 901
- Cirasuolo M., Celotti A., Magliocchetti M., Danese L., 2003, *MNRAS*, 346, 447
- Corbett E. A., et al., 2003, *MNRAS*, 343, 705
- Cristiani S., Vio R., 1990, *A&A*, 227, 385
- Croom S. M., Shanks T., Boyle B. J., Smith R. J., Miller L., Loaring N. S., 1998, *ArXiv Astrophysics e-prints*,
- Croom S. M., Shanks T., Boyle B. J., Smith R. J., Miller L., Loaring N. S., Hoyle F., 2001, *Monthly Notices of the Royal Astronomical Society*, 325, 483
- Croom S. M., Boyle B. J., Loaring N. S., Miller L., Outram P. J., Shanks T., Smith R. J., 2002, *MNRAS*, 335, 459
- Croom S. M., Smith R. J., Boyle B. J., Shanks T., Miller L., Outram P. J., Loaring N. S., 2004, *MNRAS*, 349, 1397
- Croom S. M., et al., 2009, *MNRAS*, 399, 1755
- Dallacasa D., Stanghellini C., Centonza M., Fanti R., 2000, *A&A*, 363, 887
- DiPompeo M. A., Brotherton M. S., Becker R. H., Tran H. D., Gregg M. D., White R. L., Laurent-Muehleisen S. A., 2010, *ApJS*, 189, 83
- DiPompeo M. A., Brotherton M. S., De Breuck C., Laurent-Muehleisen S., 2011, *ApJ*, 743, 71
- DiPompeo M. A., Brotherton M. S., De Breuck C., 2012, *ApJ*, 752, 6
- Dunlop J. S., McLure R. J., Kukula M. J., Baum S. A., O’Dea C. P., Hughes D. H., 2003, *MNRAS*, 340, 1095
- Ellis R. S., Colless M., Broadhurst T., Heyl J., Glazebrook K., 1996, *MNRAS*, 280, 235

- Elvis M., 2000, *ApJ*, 545, 63
- Elvis M., et al., 1994, *ApJS*, 95, 1
- Falder J. T., et al., 2010, *MNRAS*, 405, 347
- Fan X., et al., 2001, *AJ*, 121, 54
- Fanaroff B. L., Riley J. M., 1974, *MNRAS*, 167, 31P
- Fanti R., Fanti C., Schilizzi R. T., Spencer R. E., Nan Rendong Parma P., van Breugel W. J. M., Venturi T., 1990, *A&A*, 231, 333
- Felten J. E., 1976, *ApJ*, 207, 700
- Ferrarese L., Merritt D., 2000, *ApJ*, 539, L9
- Fine S., Jarvis M. J., Mauch T., 2011, *MNRAS*, 412, 213
- Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
- Garofalo D., Evans D. A., Sambruna R. M., 2010, *MNRAS*, 406, 975
- Garrington S. T., Leahy J. P., Conway R. G., Laing R. A., 1988, *Nature*, 331, 147
- Gebhardt K., et al., 2000, *ApJ*, 539, L13
- Ghez A. M., Salim S., Hornstein S. D., Tanner A., Lu J. R., Morris M., Becklin E. E., Duchêne G., 2005, *ApJ*, 620, 744
- Gibson R. R., et al., 2009, *ApJ*, 692, 758
- Glikman E., et al., 2012, *ApJ*, 757, 51
- Goldschmidt P., Kukula M. J., Miller L., Dunlop J. S., 1999, *ApJ*, 511, 612
- Gopal-Krishna Patnaik A. R., Steppe H., 1983, *A&A*, 123, 107
- Greenstein J. L., Schmidt M., 1964, *ApJ*, 140, 1
- Gregg M. D., Becker R. H., White R. L., Helfand D. J., McMahon R. G., Hook I. M., 1996, *AJ*, 112, 407
- Gregg M. D., Lacy M., White R. L., Glikman E., Helfand D., Becker R. H., Brotherton M. S., 2002, *ApJ*, 564, 133
- Gregg M. D., Becker R. H., de Vries W., 2006, *ApJ*, 641, 210

- Gunn J. E., et al., 1998, *AJ*, 116, 3040
- Gunn J. E., et al., 2006, *AJ*, 131, 2332
- Hagan M., Menhaj M.-B., 1994, *Neural Networks, IEEE Transactions on*, 5, 989
- Hall P. B., Osmer P. S., Green R. F., Porter A. C., Warren S. J., 1996, *ApJ*, 471, 1073
- Hall P. B., et al., 2002, *ApJS*, 141, 267
- Hamilton T. S., 2010, *MNRAS*, 407, 2393
- Hasinger G., Miyaji T., Schmidt M., 2005, *A&A*, 441, 417
- Hazard C., Mackey M. B., Shimmins A. J., 1963, *Nature*, 197, 1037
- Helfand D. J., White R. L., Becker R. H., 2015, *ApJ*, 801, 26
- Hewett P. C., Foltz C. B., 2003, *AJ*, 125, 1784
- Hewett P. C., Foltz C. B., Chaffee F. H., 1995, *The Astronomical Journal*, 109, 1498
- Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, *MNRAS*, 367, 454
- Hewitt A., Burbidge G., 1993, *ApJ*, 87, 451
- Ho L. C., Peng C. Y., 2001, *ApJ*, 555, 650
- Hogg D. W., Baldry I. K., Blanton M. R., Eisenstein D. J., 2002, preprint (arXiv:astro-ph/0210394),
- Hooper E. J., Impey C. D., Foltz C. B., Hewett P. C., 1995, *ApJ*, 445, 62
- Hopkins P. F., Elvis M., 2010, *MNRAS*, 401, 7
- Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, *ApJS*, 163, 1
- Hoyle F., Outram P. J., Shanks T., Boyle B. J., Croom S. M., Smith R. J., 2002, *MNRAS*, 332, 311
- Humason M. L., Mayall N. U., Sandage A. R., 1956, *AJ*, 61, 97
- Ivezić Ž., et al., 2002, *AJ*, 124, 2364
- Jiang D. R., Wang T. G., 2003, *A&A*, 397, L13

- Jiang L., Fan X., Ivezić Ž., Richards G. T., Schneider D. P., Strauss M. A., Kelly B. C., 2007, *ApJ*, 656, 680
- Jiang L., et al., 2008, *AJ*, 135, 1057
- Kellermann K. I., Sramek R., Schmidt M., Shaffer D. B., Green R., 1989, *AJ*, 98, 1195
- Khachikian E. Y., Weedman D. W., 1974, *ApJ*, 192, 581
- Kimball A. E., Knapp G. R., Ivezić Ž., West A. A., Bochanski J. J., Plotkin R. M., Gordon M. S., 2009, *ApJ*, 701, 535
- Koo D. C., Kron R. G., 1982, *AAP*, 105, 107
- Korista K. T., Voit G. M., Morris S. L., Weymann R. J., 1993, *ApJS*, 88, 357
- Kormendy J., Gebhardt K., 2001, in Wheeler J. C., Martel H., eds, *American Institute of Physics Conference Series Vol. 586, 20th Texas Symposium on relativistic astrophysics*. pp 363–381 ([arXiv:astro-ph/0105230](https://arxiv.org/abs/astro-ph/0105230)), doi:10.1063/1.1419581
- Kormendy J., Richstone D., 1995, *Ann.Rev.Astron.Astrophys.*, 33, 581
- Kratzer R. M., 2014, PhD thesis, Drexel University
- Krawczyk C. M., Richards G. T., Gallagher S. C., Leighly K. M., Hewett P. C., Ross N. P., Hall P. B., 2015, *AJ*, 149, 203
- Kunert-Bajraszewska M., Marecki A., 2007, *A&A*, 469, 437
- Kunert-Bajraszewska M., Janiuk A., Gawroński M. P., Siemiginowska A., 2010, *ApJ*, 718, 1345
- La Franca F., Gregorini L., Cristiani S., de Ruitter H., Owen F., 1994, *AJ*, 108, 1548
- Laing R. A., 1988, *Nature*, 331, 149
- Lal D. V., Shastri P., Gabuzda D. C., 2011, *ApJ*, 731, 68
- Laor A., 2003, *ApJ*, 590, 86
- Lawrence A., et al., 2007, *MNRAS*, 379, 1599
- Lípari S. L., Terlevich R. J., 2006, *MNRAS*, 368, 1001

- Liu Y., Jiang D. R., Wang T. G., Xie F. G., 2008, *MNRAS*, 391, 246
- Londish D., et al., 2002, *MNRAS*, 334, 941
- Lupton R., Gunn J. E., Ivezić Z., Knapp G. R., Kent S., 2001, in Harnden Jr. F. R., Primini F. A., Payne H. E., eds, *Astronomical Society of the Pacific Conference Series Vol. 238, Astronomical Data Analysis Software and Systems X*. p. 269 (arXiv:astro-ph/0101420)
- Lynden-Bell D., 1969, *Nature*, 223, 690
- Lynds C. R., 1967, *ApJ*, 147, 396
- Maccacaro T., della Ceca R., Gioia I. M., Morris S. L., Stocke J. T., Wolter A., 1991, *ApJ*, 374, 117
- Marecki A., Niezgoda J., Włodarczak J., Kunert M., Spencer R. E., Kus A. J., 2003, *PASA*, 20, 42
- Marshall H. L., Tananbaum H., Avni Y., Zamorani G., 1983, *ApJ*, 269, 35
- McAlpine K., Smith D. J. B., Jarvis M. J., Bonfield D. G., Fleuren S., 2012, *MNRAS*, 423, 132
- McGreer I. D., Helfand D. J., White R. L., 2009, *AJ*, 138, 1925
- McGreer I. D., et al., 2013, *ApJ*, 768, 105
- Miller L., Peacock J. A., Mead A. R. G., 1990, *MNRAS*, 244, 207
- Miller C. J., Nichol R. C., Gómez P. L., Hopkins A. M., Bernardi M., 2003, *ApJ*, 597, 142
- Minkowski R., 1960, *ApJ*, 132, 908
- Miyaji T., Hasinger G., Schmidt M., 2001, *A&A*, 369, 49
- Montenegro-Montes F. M., Mack K.-H., Vigotti M., Benn C. R., Carballo R., González-Serrano J. I., Holt J., Jiménez-Luján F., 2008, *MNRAS*, 388, 1853
- Morris S. L., Weymann R. J., Anderson S. F., Hewett P. C., Francis P. J., Foltz C. B., Chaffee F. H., MacAlpine G. M., 1991, *\aj*, 102, 1627
- Mortlock D. J., et al., 2011, *Nature*, 474, 616
- Murgia M., Fanti C., Fanti R., Gregorini L., Klein U., Mack K.-H., Vigotti M., 1999, *A&A*, 345, 769

- Murray N., Chiang J., Grossman S. A., Voit G. M., 1995, *ApJ*, 451, 498
- Narayanan D., et al., 2010, *MNRAS*, 407, 1701
- O’Dea C. P., Baum S. A., Stanghellini C., 1991, *ApJ*, 380, 66
- Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713
- Orr M. J. L., Browne I. W. A., 1982, *MNRAS*, 200, 1067
- Outram P. J., Hoyle F., Shanks T., Boyle B. J., Croom S. M., Loaring N. S., Miller L., Smith R. J., 2001, *MNRAS*, 328, 174
- Padovani P., 1993, *MNRAS*, 263, 461
- Page M. J., Carrera F. J., 2000, *MNRAS*, 311, 433
- Peacock J. A., Miller L., Longair M. S., 1986, *MNRAS*, 218, 265
- Pei Y. C., 1995, *ApJ*, 438, 623
- Peterson B. M., 1997, *An Introduction to Active Galactic Nuclei* (Cambridge: Cambridge Univ. Press),
- Peterson B. M., 2008, *NewAR*, 52, 240
- Peterson B. M., Horne K., 2004, *Astronomische Nachrichten*, 325, 248
- Pier J. R., Munn J. A., Hindsley R. B., Hennessy G. S., Kent S. M., Lupton R. H., Ivezić Ž., 2003, *AJ*, 125, 1559
- Prandoni I., Gregorini L., Parma P., de Ruiter H. R., Vettolani G., Wieringa M. H., Ekers R. D., 2001, *A&A*, 365, 392
- Proga D., Stone J. M., Kallman T. R., 2000, *ApJ*, 543, 686
- Pudritz R. E., Hardcastle M. J., Gabuzda D. C., 2012, *SSRv*, 169, 27
- Reichard T. A., et al., 2003, *AJ*, 126, 2594
- Richards G. T., et al., 2002, *AJ*, 123, 2945
- Richards G. T., et al., 2005, *MNRAS*, 360, 839
- Richards G. T., et al., 2006, *AJ*, 131, 2766
- Richards G. T., et al., 2009, *ApJS*, 180, 67

- Richstone D., et al., 1998, *Nature*, 395, A14
- Risaliti G., Elvis M., 2004, in Barger A. J., ed., *Astrophysics and Space Science Library* Vol. 308, *Supermassive Black Holes in the Distant Universe*. p. 187 (arXiv:astro-ph/0403618)
- Ross N. P., et al., 2012, *ApJS*, 199, 3
- Saikia D. J., Salter C. J., 1988, *Ann.Rev.Astron.Astrophys.*, 26, 93
- Salpeter E. E., 1964, *ApJ*, 140, 796
- Sandage A., Véron P., Wyndham J. D., 1965, *ApJ*, 142, 1307
- Sanders D. B., Mirabel I. F., 1996, *Ann.Rev.Astron.Astrophys.*, 34, 749
- Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, *ApJ*, 325, 74
- Schlegel D. J., Finkbeiner D. P., Davis M., 1998, *ApJ*, 500, 525
- Schmidt M., 1963, *Nature*, 197, 1040
- Schmidt M., 1968, *ApJ*, 151, 393
- Schmidt M., Green R. F., 1983, *ApJ*, 269, 352
- Schmidt M., Schneider D. P., Gunn J. E., 1995, *AJ*, 110, 68
- Schneider D. P., van Gorkom J. H., Schmidt M., Gunn J. E., 1992, *AJ*, 103, 1451
- Schneider D. P., et al., 2005, *AJ*, 130, 367
- Schneider D. P., et al., 2010, *The Astronomical Journal*, 139, 2360
- Seyfert C. K., 1943, *ApJ*, 97, 28
- Shakura N. I., Sunyaev R. A., 1973, *AAP*, 24, 337
- Shen Y., Kelly B. C., 2012, *ApJ*, 746, 169
- Sikora M., Stawarz Ł., Lasota J.-P., 2007, *ApJ*, 658, 815
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Simcoe R. A., Sullivan P. W., Cooksey K. L., Kao M. M., Matejek M. S., Burgasser A. J., 2012, *Nature*, 492, 79

- Smith H. J., Hoffleit D., 1963, *AJ*, 68, 292
- Smith D. J. B., et al., 2011, *MNRAS*, 416, 857
- Sprayberry D., Foltz C. B., 1992, *ApJ*, 390, 39
- Stern D., Djorgovski S. G., Perley R. A., de Carvalho R. R., Wall J. V., 2000, *AJ*, 119, 1526
- Stern D., et al., 2012, *ApJ*, 753, 30
- Stocke J. T., Morris S. L., Weymann R. J., Foltz C. B., 1992, *ApJ*, 396, 487
- Stoughton C., et al., 2002, *AJ*, 123, 485
- Sulentic J. W., Zamfir S., Marziani P., Bachev R., Calvani M., Dultzin-Hacyan D., 2003, *ApJ*, 597, L17
- Sutherland W., Saunders W., 1992, *MNRAS*, 259, 413
- Tadhunter C., 2008, *NewAR*, 52, 227
- Tanvir N. R., et al., 2009, *Nature*, 461, 1254
- Trump J. R., et al., 2006, *ApJS*, 165, 1
- Tuccillo D., González-Serrano J. I., Benn C. R., 2015, *MNRAS*, 449, 2818
- Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, *ApJ*, 598, 886
- Urrutia T., Becker R. H., White R. L., Glikman E., Lacy M., Hodge J., Gregg M. D., 2009, *ApJ*, 698, 1095
- Urry C. M., Padovani P., 1995, *PASP*, 107, 803
- Vanden Berk D. E., et al., 2001, *aj*, 122, 549
- Vanden Berk D. E., et al., 2005, *AJ*, 129, 2047
- Venemans B. P., et al., 2012, *ApJ*, 751, L25
- Veron-Cetty M.-P., Veron P., 2010, *\aap*, 518, A10
- Vigotti M., Carballo R., Benn C. R., De Zotti G., Fanti R., Gonzalez Serrano J. I., Mack K.-H., Holt J., 2003, *ApJ*, 591, 43
- Visnovsky K. L., Impey C. D., Foltz C. B., Hewett P. C., Weymann R. J., Morris S. L., 1992, *ApJ*, 391, 560

- Warren S. J., Hewett P. C., Osmer P. S., 1991, *ApJS*, 76, 23
- Weymann R. J., Morris S. L., Foltz C. B., Hewett P. C., 1991, *ApJ*, 373, 23
- White R. L., 2000, in Manset N., Veillet C., Crabtree D., eds, *Astronomical Society of the Pacific Conference Series Vol. 216, Astronomical Data Analysis Software and Systems IX*. p. 577
- White R. L., Becker R. H., Helfand D. J., Gregg M. D., 1997, *ApJ*, 475, 479
- White R. L., Helfand D. J., Becker R. H., Gregg M. D., Postman M., Lauer T. R., Oegerle W., 2003, *AJ*, 126, 706
- White R. L., Helfand D. J., Becker R. H., Glikman E., de Vries W., 2007, *ApJ*, 654, 99
- Willott C. J., et al., 2010, *AJ*, 139, 906
- Wilson A. S., Colbert E. J. M., 1995, *ApJ*, 438, 62
- Wisotzki L., 2000, *A&A*, 353, 861
- Wright E. L., et al., 2010, *AJ*, 140, 1868
- Yip C. W., et al., 2004, *AJ*, 128, 2603
- Zel'dovich Y. B., Novikov I. D., 1964, *Soviet Physics Doklady*, 9, 246
- Zhang Y., Zhao Y., 2003, *PASP*, 115, 1006
- de Veny J. B., Osborn W. H., Janes K., 1971, *\pasp*, 83, 611
- de Vries W. H., Becker R. H., White R. L., 2006, *aj*, 131, 666
- van Breugel W., Miley G., Heckman T., 1984, *AJ*, 89, 5