

Dissertation

submitted to the
Instituto de Astrofísica, Facultad de Física
Pontificia Universidad Católica de Chile, Chile
for the degree of
Doctor in Astrophysics

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

Cristina García Vergara
Born in Santiago, Chile

Oral examination: 25 May 2016

Quasar Environments at $z \sim 4$

Cristina García Vergara
Max-Planck-Institut für Astronomie

Referees: Prof. Dr. L. Felipe Barrientos
Dr. Joseph Hennawi

Abstract

In the standard picture of structure formation, the first massive galaxies form in the highest peaks of the density field, which are the cores of massive proto-clusters. Such structures must be exceedingly rare, and thus hard to find. Luminous quasars (QSOs) at $z \sim 4$ are the most strongly clustered population in the Universe and this large auto-correlation demands that they reside in massive dark matter halos, associated with large overdensities of galaxies. This imply a strong QSO-galaxy cross-correlation for luminous QSOs at $z \sim 4$. In order to put the first observational constrain on the QSO-galaxy clustering properties, I present the measurement of the cross-correlation function between QSOs and both Lyman break galaxies (LBGs) and Lyman alpha emitters (LAEs) at $z \sim 4$. I present VLT/FORS1 observations of 6 luminous QSOs fields. Using a novel narrow band filter technique I select LBGs associated with each quasar in a narrow redshift range of $\Delta z \sim 0.2$. I measure the QSO-LBG cross correlation function on scales of $0.1 < R < 10 h^{-1} \text{ Mpc}$ (comoving), which is well fitted by a power law form with a correlation length $r_0 = 9.91^{+3.28}_{-1.79} h^{-1} \text{ Mpc}$ and a slope $\gamma = 2.05^{+0.20}_{-0.46}$. This is in agreement with the theoretical expected clustering computed from the individual QSO and LBGs auto-correlation, assuming a linear bias model. I also measure the auto-correlation of LBGs in the QSO fields, which shows a stronger clustering in comparison with LBGs in random fields. The new technique used for the detection of LBGs is efficient in selecting them in a thin redshift slice, but this shows important shortcomings whereby this result should be carefully considered. Additionally, I present VLT/FORS2 observations of 14 luminous QSO fields, designed to search for LAEs in their environments. I find that QSOs and LAEs are not correlated at $z \sim 4$, and the mean LAEs number density in our fields is consistent with the expected number density in random fields. This could mean either that the QSO auto-correlation length at $z \sim 4$ is overestimated, or that LAEs preferentially avoid QSO environments on $\lesssim 10 \text{ Mpc } h^{-1}$ scales.

Zusammenfassung

Im klassischen Bild der Strukturbildung bilden sich die ersten massiven Galaxien in den dichtesten Regionen des Dichtefeldes, wo sich später die massive Proto-Galaxienhaufen bilden. Diese Strukturen sind sehr selten und deshalb schwer zu finden. Leuchtstarke Quasare (QSOs) bei $z \sim 4$ sind die am dichtesten zusammen geballte Population im Universum und diese grosse Autokorrelation beinhaltet, dass sie sich in schweren Halos dunkler Materie befinden an den dichtesten Regionen im Dichtefeld des Universums. Daraus lässt sich auf eine starke Quasar-Galaxien Kreuzkorrelation für leuchtstarke Quasare bei $z \sim 4$ schließen. Um die Eigenschaften der Gruppierung von QSO-Galaxien zum ersten Mal durch Beobachtungen zu bestimmen, präsentiere ich in dieser Arbeit die ersten Messungen der Kreuzkorrelation Funktion zwischen QSO und Lymanbreak Galaxien (LBGs) sowie Lyman-Alpha-Emittern (LAE) bei $z \sim 4$. In dieser Arbeit zeige ich VLT/FORS1 Beobachtungen von 6 leuchtstarken Quasarfeldern. Wir benutzen eine neue Bandpass Filter Technik, die ich von LBGs ausgesucht habe, die mit jedem Quasar in einem engen Rotverschiebungsintervall von $\Delta z \sim 0.2$ liegen. Ich messe die QSO-LBG cross-correlation Funktion bei $z \sim 4$ in einer Größenordnung von $0.1 < R < 10 h^{-1}$ Mpc (comoving), die man gut mit einem Potenzgesetz mit Längenskala $r_0 = 9.91_{-1.79}^{+3.28} h^{-1}$ Mpc für einen festen Wert für $\gamma = 2.05_{-0.46}^{+0.20}$ beschreiben kann. Dieses Ergebnis stimmt mit theoretischen Erwartungen für Galaxienhaufen überein, die aus einzelne QSO und LBGs Autokorrelation in der Annahme eines linearen Bias-Modells berechnet wurden. Ausserdem bestimme ich die Autokorrelation von LBGs in den Feldern, die sehr viel stärkere Ballung aufweist als LBGs in willkürlich gewählten Feldern. Diese neue Technik, die wir nutzen um LBGs zu entdecken, ist effizient, um Galaxien in einem schmalen Rotverschiebungsintervall zu finden, aber hat viele Nachteile, weshalb diese Methode mit Vorsicht verwendet werden sollte. Zusätzlich zeige ich VLT/FORS2 Beobachtungen von 14 leuchtkräftigen $z \sim 4$ QSO Feldern, die wir ausgewählt haben, um LAEs in ihrer Umgebung zu finden. Meine erste Messung der LAE-QSO Kreuzkorrelation bei $z \sim 4$ stimmt mit keiner Korrelation überein. Ich beobachte, dass die mittelwertige Anzahldichte von LAEs in diesen Feldern mit der, in den willkürlich gewählten Feldern, übereinstimmt. Dieses Ergebnis weist darauf hin, dass entweder die Autokorrelationslänge der QSO bei $z \sim 4$ überschätzt ist, oder dass LAEs vorzugsweise QSO-Umgebungen in Skalen von $\lesssim 10 \text{ Mpc } h^{-1}$ vermeiden.

Resumen

En la visión estándar de la formación de estructura, las primeras galaxias masivas se forman en los *peaks* más altos del campo de densidad, los cuales son los núcleos de proto-cumulos masivos. Tales estructuras debieran ser extremadamente raras, y por lo tanto difíciles de encontrar. Los quásares (QSOs) luminosos a $z \sim 4$ son la población más fuertemente agrupada en el Universo y esta gran auto-correlación exige que ellos residan en halos masivos de materia oscura, asociado con grandes sobredensidades de galaxias. Esto implica una fuerte correlación cruzada de QSOs con galaxias para QSOs luminosos a $z \sim 4$. Con el objetivo de poner la primera restricción observacional sobre las propiedades de *clustering* de QSOs con galaxias, yo presento la medición de la función de correlación cruzada entre los QSOs y las galaxias con salto de Lyman (LBGs) y entre los QSOs y emisores de Lyman alpha (LAEs) a $z \sim 4$. Presento observaciones de VLT/FORS1 de 6 campos de QSOs luminosos. Usando una nueva técnica de filtro de banda delgada, selecciono LBGs asociadas a cada quasar en un delgado rango de redshift de $\Delta z \sim 0.2$. Mido la función de correlación cruzada de QSOs con LBGs en escalas de $0.1 < R < 10 h^{-1} \text{ Mpc}$ (comóvil), la cual es bien ajustada por una ley de potencia con una amplitud de correlación de $r_0 = 9.91^{+3.28}_{-1.79} h^{-1} \text{ Mpc}$ y una pendiente $\gamma = 2.05^{+0.20}_{-0.46}$. Esto está de acuerdo con el *clustering* esperado teóricamente, calculado a partir de la auto-correlación individual de QSOs y LBGs, asumiendo un modelo de bias lineal. Además mido la auto-correlación de LBGs en los campos de QSOs, la cual muestra un *clustering* más fuerte en comparación con LBGs en campos aleatorios. La nueva técnica utilizada para la detección de LBGs es eficiente en la selección de ellas en una ventana delgada de redshift, pero muestra importantes deficiencias por lo que este resultado debe ser considerado cuidadosamente. Además presento observaciones de VLT/FORS2 de 14 campos de QSOs luminosos, diseñadas para buscar LAEs en sus entornos. Encuentro que los QSOs y las LAEs no están correlacionadas a $z \sim 4$, y la densidad numérica media de LAEs en nuestros campos es consistente con la densidad numérica esperada en campos aleatorios. Esto podría significar que la amplitud de la auto-correlación de QSOs a $z \sim 4$ está sobreestimada, o que los LAEs preferentemente evitan entornos de QSO en escalas de $\lesssim 10 \text{ Mpc } h^{-1}$.

Acknowledgements

I would like to thank my supervisors, Felipe Barrientos and Joseph Hennawi, for guiding me and supporting me during the development of this thesis. I feel very fortunate to have been able to learn from two great supervisors, with two different perspectives and ways of conducting research, allowing me to form a broad idea of the scientific career.

I am very grateful to Felipe, who has been present since the beginning of my career, and whose teachings broadened my knowledge about the astronomical and scientific world. I appreciate the guidance and support he gave me during the whole PhD process. I am grateful for the opportunity he gave me to visit different observatories, where I acquired my observational experience and knowledge. I thank him for pushing me to apply for the double degree exchange program with Heidelberg University, which was an enormous opportunity to develop and enhance my scientific skills in the best way possible. I also appreciate the concern he consistently showed about my future and the financial support to finish my PhD.

I deeply thank Joseph for giving me the opportunity to work with him and for welcoming me during the time I was in Heidelberg. His advice, constructive criticism, creativity and immense knowledge provided me with a great learning experience which was critical in the formation of my career. I thank him for his wise advice on how to improve my oral presentations and proposals for observation, which allowed me to get additional data for my PhD thesis. I am so grateful for the time and dedication he gave to my project, both in Heidelberg and upon my return to Chile, which has been essential for the success of this thesis. I also appreciate his interest in my formation as a researcher and his recognition of my progress, beyond just the interest in the success of the thesis. His contagious enthusiasm and passion for science have been motivational to keep myself in the scientific career and it has been a huge privilege to have learned from him.

I want to express my gratitude to the ENIGMA group for the excellent and comfortable learning environment, where I acquired a better understanding of AGNs and high-z Universe and where I learned to develop scientific criticism. Particularly, I would like to thank Fabrizio Arrigoni for his patience and immense help in the process of data reduction presented in this thesis. I am grateful to Christina Eilers and Gabriela Calistro for their help translating the abstract to German.

Several people freely shared their knowledge, perspectives and opinions with me which have had a large impact in the development of this thesis. I would like to thank the following individuals for the important discussions and suggestions: Jorge Gonzalez, Eduardo Bañados, Timo Anguita, Andres Jordan, Alejandra Muñoz, Harold Francke and Ivan Lacerna.

I also would like to acknowledge Gabor Worseck, Yue Shen, Arjen van der Wel and Bram Venemans for kindly providing useful data and material used in this thesis.

I am very grateful to all of my friends for making my life enjoyable, for their constant support, and for their words of encouragement when I felt overwhelmed. A special thanks to my friends Alejandra Rojas, Andrea Corvillón, Fabiola Salinas, Juan Carlos Beamín, Jorge Gonzalez, Alejandra Muñoz, Daniela Carrasco, Catalina Infante, Elena Manjavacas, Gabriela Calistro, Alberto Rorai, Fabrizio Arrigoni, Mauricio Ortiz, Natalia Kudryavtseva, Mauricio Carrasco, Gustavo Morales, Fabian Olivares, Maria Jose Cordero, Jorge Diaz, Romina Ramirez, Cristina Vera, Silvana Lagos and Claudia Bustamante.

I am extremely grateful to my parents, Claudio and Carmen Gloria, who have always supported me unconditionally throughout my career, from the moment I decided to study astronomy. My accomplishments have been possible because of the possibility they gave me to study and for the limitless love they demonstrate each day. I thank my brothers, Jaime and Pablo, my sister-in-law, Kelsey, and my grandparents who all form an important support network.

Above all, I want to thank my beloved partner, Claudio Navarro, who has been the most fundamental pillar in this long journey. He has encouraged me every time I have needed it and he has motivated me to continue when I felt I had no more strength. I thank him for his willingness to follow me when we had to leave our country. I am deeply grateful for his emotional support and love.

This thesis would not have been possible without the financial support from: MECE-SUP program (Programa de Mejoramiento de la Calidad y la Equidad de la Educación Superior), CONICYT-PCHA/Doctorado Nacional 2012-21120442, PUC-HD Graduate Exchange Fellowship by DAAD, and Proyecto Financiamiento BASAL PFB06 de Conicyt.

Contents

| | |
|-------------------------------------------------------------------------|-----------|
| Acknowledgements | ix |
| Contents | xi |
| 1 Introduction | 1 |
| 1.1 Structure Formation in the Universe | 1 |
| 1.2 Galaxy Clusters and Protoclusters | 6 |
| 1.3 AGN as Tracers of Massive Dark Matter Halos | 6 |
| 1.4 Protocluster Search in AGN environments | 7 |
| 1.5 Outline of This Thesis | 10 |
| 2 Basics on Correlation Function and Clustering Measurements | 13 |
| 2.1 Definition of the Correlation Function | 13 |
| 2.2 Estimators of $\xi(r)$ | 15 |
| 2.3 Estimation of r_0 and γ Parameters | 16 |
| 2.3.1 Projected Correlation Function | 17 |
| 2.3.2 Volume Averaged Projected Correlation Function | 18 |
| 2.3.3 Angular Correlation Function | 19 |
| 2.4 The Special Case of Computing Clustering on a Overdensity | 21 |
| 3 Observations, Data Reduction and Photometry | 23 |
| 3.1 A Nobel Method to Detect LBGs | 23 |
| 3.2 QSO Targets | 24 |
| 3.3 Imaging Observations | 25 |
| 3.4 Data Reduction | 26 |
| 3.5 Photometry | 27 |
| 4 Lyman Break Galaxies Selection at $z \sim 4$ | 29 |
| 4.1 LBGs Color Modeling | 30 |
| 4.2 Low Redshift Galaxies Colors | 34 |
| 4.3 Selection Region and LBGs Sample | 38 |
| 4.4 Color Selection Completeness | 43 |
| 5 Clustering Analysis | 49 |
| 5.1 Expected Clustering Signal | 49 |
| 5.2 QSO-LBG Cross-Correlation Measurement | 53 |
| 5.2.1 Fitting r_0 Parameter | 54 |

| | | |
|----------|---------------------------------------------------------------------------------|------------|
| 5.3 | LBG Auto-Correlation Measurement | 58 |
| 5.4 | Contamination effects on the Clustering Measurements | 61 |
| 6 | The First QSO-LAE Cross-Correlation Measurement at $z \sim 4$ | 67 |
| 6.1 | Data Description | 67 |
| 6.1.1 | QSO Targets | 68 |
| 6.1.2 | Observations | 68 |
| 6.1.3 | Data Reduction | 69 |
| 6.1.4 | Photometry | 70 |
| 6.2 | LAEs Selection | 72 |
| 6.2.1 | Selection Region Choice | 73 |
| 6.2.2 | LAEs Sample and Completeness | 76 |
| 6.3 | Clustering Measurements | 79 |
| 7 | Summary and Conclusions | 85 |
| 7.1 | Future Work | 87 |
| | | |
| A | LBGs Sample | 91 |
| | | |
| B | LAEs Sample | 99 |
| | | |
| | Bibliography | 105 |
| | List of Figures | 115 |
| | List of Tables | 119 |

Chapter 1

Introduction

One of the fundamental fields of research in Cosmology is the formation and evolution of large scale structure observed in the Universe. Several simulations and theoretical descriptions of this process are available (e.g. Boylan-Kolchin et al., 2009; De Lucia and Blaizot, 2007; Jenkins et al., 1998; Springel et al., 2005a), but understanding it from an observational point of view is a challenge to make, specially when we focus on the high- z Universe.

This thesis is pointing to that direction, and pretends to observationally understand how structures form in the early Universe and compare this with the theoretical predictions. In particular, the aim is to study the proto-clusters formation at $z \sim 4$.

In sections 1.1, 1.2, and 1.3 I describe the scientific background of this thesis, specifically, I describe the current picture of structure formation and the importance of galaxy clusters to understand it. I also list the reasons to think that QSOs are tracers of dark matter halos in the Universe¹. Then, in section 1.4 I summarize the main results of observational searching for protoclusters at high redshift and I expose the reasons why this thesis focuses on the study of $z \sim 4$ QSOs environments. Finally in section 1.5 I present the outline of this thesis.

1.1 Structure Formation in the Universe

The cosmological principle establishes that the Universe is homogeneous and isotropic on large scales. However, it is very inhomogeneous at small scales, otherwise stars, galaxies and galaxy clusters would not exist. This claim can be seen very clearly in the study

¹Part of the information here provided is based on the books Dodelson (2003); Peacock (1999); Schneider (2015).

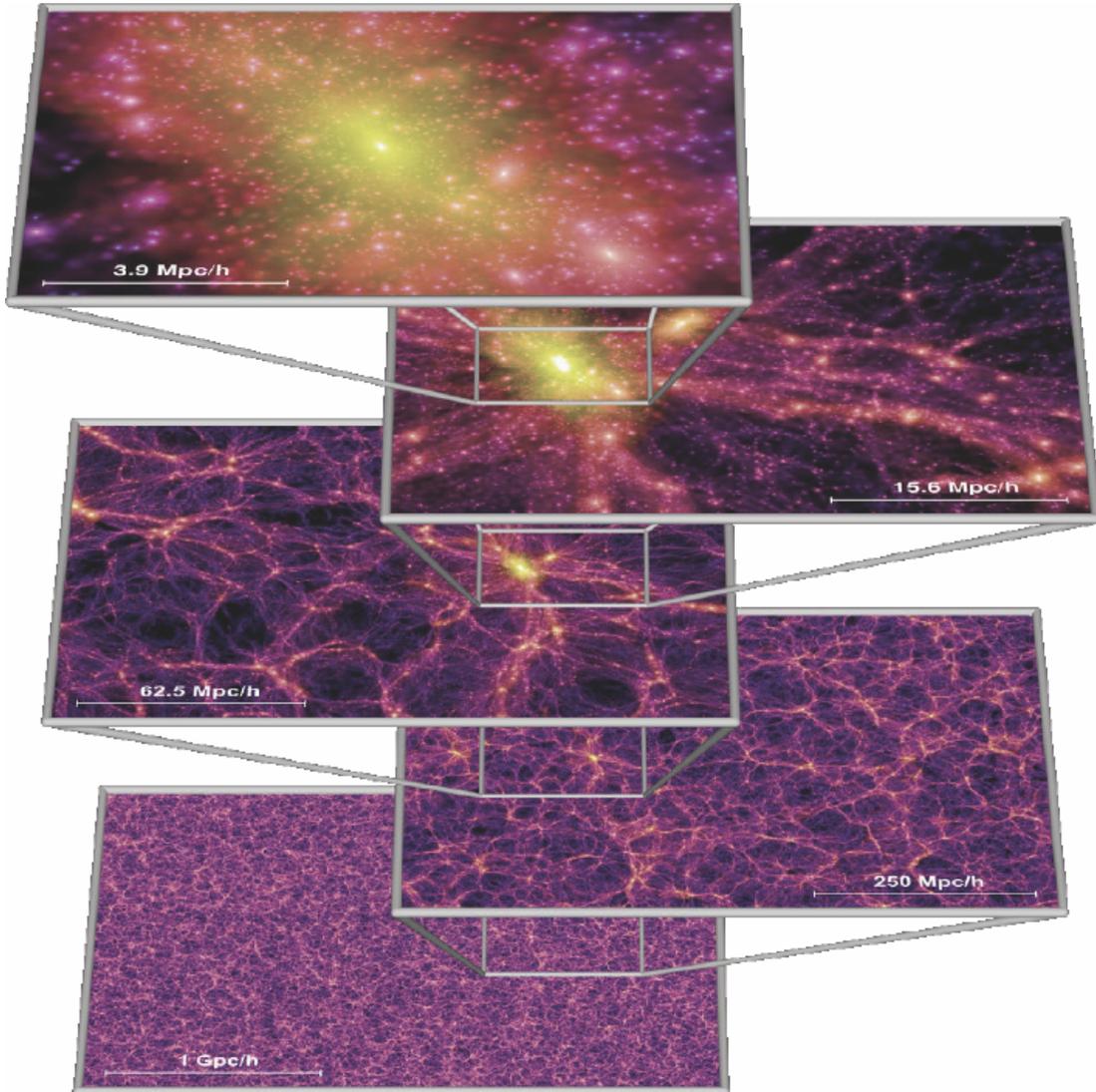


FIGURE 1.1: Simulation showing the dark matter distribution in the Universe at $z = 0$ at different scales. Each slice has a thickness of 15 Mpc h^{-1} . The lower slice shows a homogeneous Universe as the cosmological principle states, but the smaller the scale is, the more distinguishable the structure is, and the inhomogeneity at small scales becomes clear. This figure is taken from Springel et al. (2005b) (The background color of this figure was modified).

presented by Springel et al. (2005b). They performed a simulation of the growth of dark matter structure, where they also follow the formation, evolution and distribution of the visible components as quasars and galaxies. In Fig. 1.1 we show some slices of their simulation at $z = 0$ for different scales. At large scales ($\sim 1 \text{ Gpc h}^{-1}$), the Universe looks very homogeneous, but when they zoom in to a region around a galaxy cluster, some structure becomes distinguishable, and the inhomogeneity is completely clear at the smallest scales ($\sim 3.9 \text{ Mpc h}^{-1}$).

This simulation agrees with observational studies of galaxy distribution in the local

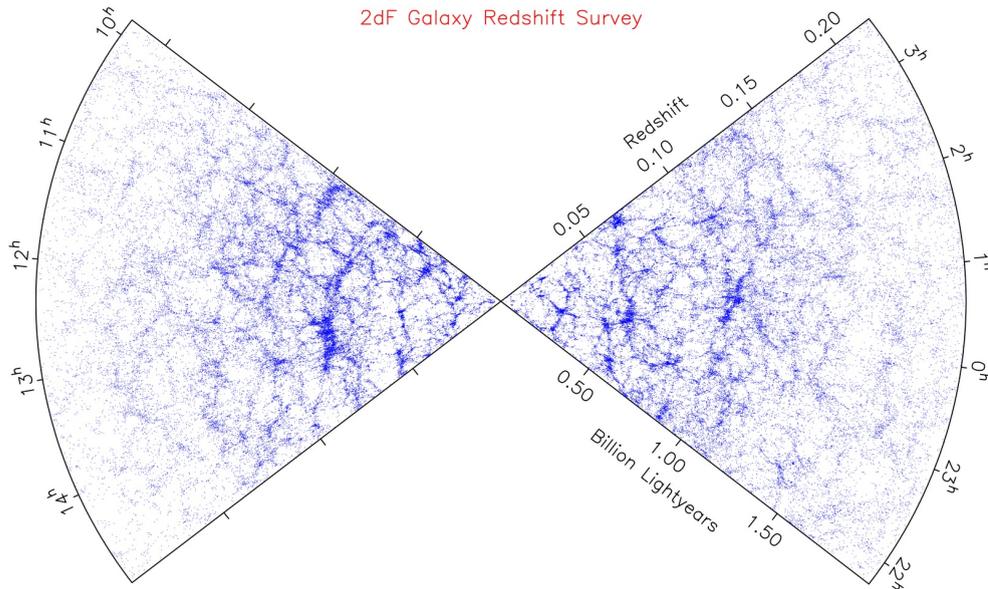


FIGURE 1.2: Distribution of galaxies in the Universe, from the 2dF galaxy redshift survey. The polar angle is indicating the right ascension and the radial axes indicates the redshift. The large scale structure formed by filaments and voids is clearly observed. This figure is taken from the 2dF image gallery.

Universe. The 2dF Galaxy Redshift Survey (Colless et al., 2001) is a spectroscopic survey built by the Anglo-Australian Observatory, which obtained spectra for $\sim 220,000$ galaxies over an area of ~ 1500 square degrees². In the Fig. 1.2 I show the distribution of these galaxies, where we can recognize that galaxies are not randomly distributed in the Universe, but they tend to be clustered, forming a filamentary structure and voids which are regions where apparently there are not visible galaxies. Therefore, it is of general consensus that inhomogeneities exist at small scales in the local Universe.

After this, the natural question that arise is: Did those inhomogeneities exist at earlier times in the Universe? The answer of this question can be taken from the cosmic microwave background (CMB) observations. The CMB is the thermal radiation coming from $z \sim 1100$ when the photons decouple from matter, then this radiation bring us information from the earliest time in the Universe. It was firstly observed by Penzias and Wilson (1965) but some spacecrafts have been launched more recently in order to have more accurate measurements of the CMB (for example COBE³, WMAP⁴, and Planck).

The last measurements show an anisotropy of the CMB, with relative temperatures fluctuations of $\Delta T/T \sim 10^{-5}$, which indicates that the Universe already showed small inhomogeneities at $z \sim 1100$. That means that if we consider the matter density field at

²<http://www.2dfgrs.net/>

³Cosmic Background Explorer

⁴Wilkinson Microwave Anisotropy Probe

this redshift we find some regions where the matter density is a little higher than the mean, and in other regions we find a density a little lower than the mean. Those very small fluctuations had to have grown up over time to give rise to the large fluctuations we observe at $z = 0$ as are for example cluster of galaxies.

In the standard scenario for the structure formation, it is suggested that structures grow hierarchically through gravitational instability (e.g. Dodelson, 2003; Padmanabhan, 2006; Schneider, 2015). In order to understand this statement, we define the relative density contrast as:

$$\delta(r, t) = \frac{\rho(r, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \quad (1.1)$$

where $\rho(r, t)$ is the matter density in a specif time t and position r in the Universe and $\bar{\rho}(t)$ is the mean matter density in the Universe at time t . Here the $\delta(r, t)$ quantity represents how far of the mean density is the density in a certain time and position. Following this notation, in the CMB time, we had very small inhomogenities, then $|\delta(r, t)| \ll 1$, but today we have galaxy clusters where the density is much higher than the mean, then $|\delta(r, t)| \gg 1$.

We can imagine an overdense region in the Universe at $z \sim 1100$, where there is a very small $\delta(r, t)$ value. Even if that fluctuation is small, there will be a little stronger gravitational force there, compared with the force in mean density regions. While the Universe is expanding, this region, will expand a little slower just because their additional auto gravity. As a consequence, the $\delta(r, t)$ value in that region will increase and the gravity in that position will be a little stronger than before, then the expansion will be a little slower there, and so on. This stronger gravity force also attracts the surrounding matter and with time, this region will be increasingly overdense, and the small fluctuation at $z \sim 1100$ will be a large fluctuation at $z = 0$. In the same way, an underdense region in the early Universe will be increasingly underdense over the time.

This gravitational instability process govern the evolution of structures in the Universe and it aims to a hierarchical evolution, where small objects form first and the bigger ones form later. Therefore, stars form firstly, they in turn form galaxies, and galaxies form galaxy clusters to join with other clusters forming the most massive structures in the Universe: the superclusters.

This process can be seen again in the simulation performed by Springel et al. (2005b). In Fig. 1.3 we show some snapshots of their simulation, corresponding at different epochs in the Universe for the same scale, centered on a dark matter halo. Panels show the

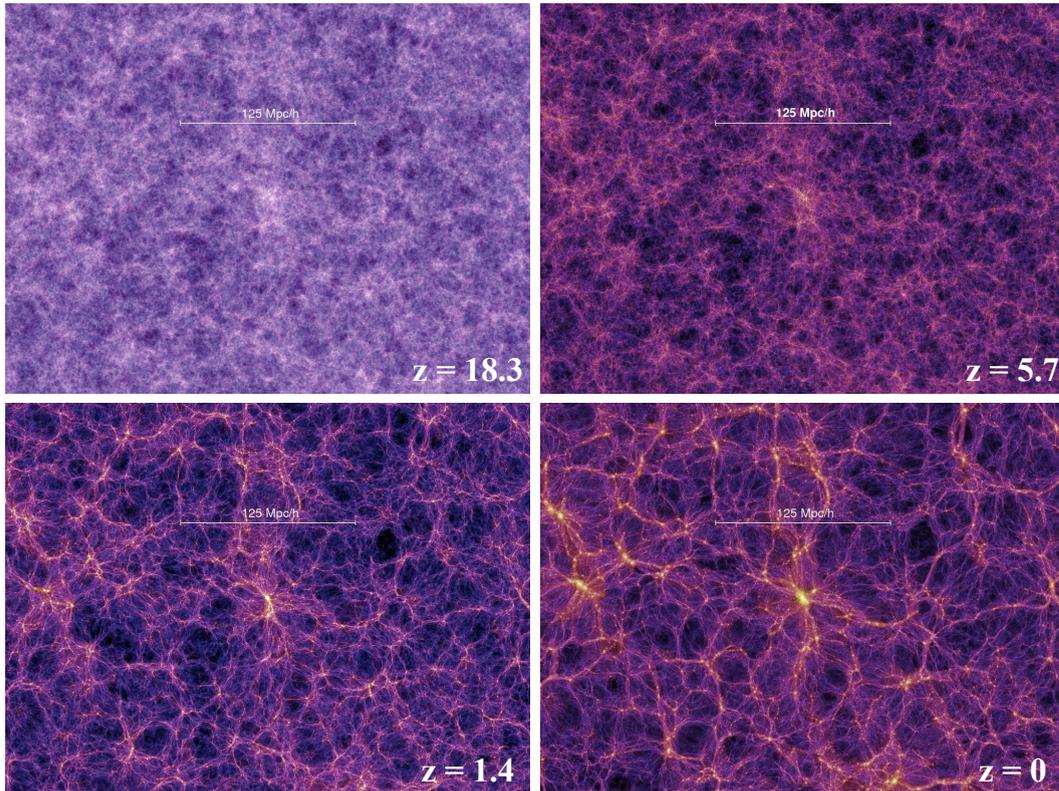


FIGURE 1.3: Simulations of the dark matter distribution in the Universe for four different redshifts and the same scale. At higher redshifts the fluctuations in the density field were smaller and the large scale structure was less well defined in comparison with the lower redshift snapshots. This illustrates the standard structure formation scenario, where structures grow hierarchically through gravitational instability. This figures were made with the images available at [www.mpa-garching.mpg.de /galform/virgo/millennium/](http://www.mpa-garching.mpg.de/galform/virgo/millennium/).

evolution of dark matter distribution from $z = 18.3$ to $z = 0$, where small fluctuations in the density field grow over the time.

Although this theoretical picture of structure formation is very well accepted by astronomers, the observational confirmation is still needed. Because of the big challenge of studying the Universe at high redshift ($z > 1$), there are not conclusive results supporting this scenario.

In this context a big interest in the study of high- z galaxy clusters appears. As they are the most prominent overdensities at $z = 0$, they must have formed in the highest overdensities in the early Universe, tracing the most prominent peaks in the density field.

1.2 Galaxy Clusters and Protoclusters

Clusters of galaxies are the most massive, gravitationally bound structures in the Universe, so they represent the most extreme deviations from the initial conditions in the Universe and therefore they are good evolutionary probes for studying the formation of the large scale structure.

The study of the galaxy clusters has a large impact on the cosmology, since it is known that their number density can be used to put constraints on some cosmological parameters, such as Ω_M and σ_8 (e.g. Bahcall and Cen, 1992; Bahcall and Fan, 1998; Eke et al., 1996; White et al., 1993). Clusters and their properties have been well studied at low redshift ($z < 1$) (Brodwin et al., 2007; Giodini et al., 2013, and references therein), but the detailed process by which clusters of galaxies form and their first evolutionary stages are not yet understood.

If we observe galaxies which will later assemble into a galaxy cluster, we are observing a so-called protocluster (nascent cluster), and then we are situated in high density peaks in the early Universe. In other words, the first massive galaxies form in the highest peaks of the density field, and these are the cores of massive protoclusters, which will evolve into the most massive clusters at $z = 0$.

Because of the extremely small areas of sky surveyed at high- z , compared to the extremely low comoving number density $\sim 10^{-7} \text{ Mpc}^{-3}$ of the clusters (Gioia et al., 2001; Vikhlinin et al., 2009), the evolutionary link between these low- z clusters and high- z galaxies has been challenging to make. The progenitors of clusters are extremely difficult to identify when the density contrast between the forming cluster and the surrounding field is small (as happens at high redshift). For this reason, it is necessary to search for an object that traces these protoclusters.

1.3 AGN as Tracers of Massive Dark Matter Halos

As clusters are related to the highest peaks of the density field, we need to identify where these peaks are. As objects with unique properties might inhabit these high peaks too, we could search for such “tracer objects” to identify these high density peaks. There are several reasons to think that suitable objects for identify high density peaks are active galactic nuclei (AGNs).

Firstly, AGNs are very rare objects at high redshift, specifically the number density of QSOs steadily decrease at $z > 2.5$ (Croom et al., 2009; Hopkins et al., 2007), then, just for a statistical argument, if we find an AGN, it is very probable that we will have a

matter overdensity there. In other words, as AGNs are not abundant, probably they are located in regions where there is major concentration of matter.

Secondly, they are very luminous at optical and near-infrared wavelengths, so finding them is relatively easy, even if they are located at high redshifts.

Thirdly, luminous AGNs host super massive black holes (SMBHs), and the masses of black holes (M_{BH}) are known to tightly correlate with the dark halo mass (M_{halo}) in nearby galaxies (Ferrarese, 2002). Intriguingly, the most luminous QSOs at $z > 3$ have $M_{\text{BH}} \sim 1 - 6 \times 10^9 M_{\odot}$, which is comparable to the most massive known BHs locally. If the present day $M_{\text{BH}} - M_{\text{halo}}$ relation holds at early times, such BHs should reside in exceptionally massive halos.

Fourth, there are some studies which suggest that AGN activity is triggered by processes related to the environment where they are embedded. For example, galaxy mergers could trigger the AGN activity (Bahcall et al., 1997; Wyithe and Loeb, 2002), and galaxy merger occur more probably in dense environments (Lacey and Cole, 1993). This would imply that the existence of an AGN requires a dense environment around it.

Finally, another line of evidence that QSOs trace the rarest environments at high redshift arises from their extremely strong clustering (Shen et al., 2007), which demand that they reside in the most massive DM halos at this epoch.

From a theoretical point of view, lately it has been suggested that luminous QSOs live in massive DM halos, but not necessarily in the most massive (Fanidakis et al., 2013). This result has been also found observationally in some luminous QSO fields at $z \sim 4 - 5$ (e.g. Adams et al., 2015; Husband et al., 2013). However a high signal to noise clustering analysis is necessary to confirm this hypothesis.

All these reasons make AGNs very good candidates to be tracers of protoclusters. Much effort has been put in the testing of this hypothesis, and the majority of the studies have focused in the search of galaxy overdensities in AGN environments at high redshift. In the next section I summarize the main results found on this topic, and I expose the reasons why this thesis focuses on the study of $z \sim 4$ QSOs environments.

1.4 Protocluster Search in AGN environments

Several authors have searched galaxy overdensities around AGNs, traced by Lyman-break galaxies (LBGs) or Lyman alpha emitters (LAEs). The population of active galactic nuclei (AGNs) whose environments have been studied most intensively are the high-redshift radio galaxies (HzRGs) at $z \sim 2 - 4$, which have been shown to often reside

TABLE 1.1: Protoclusters in QSO environments at $z \gtrsim 5$.
 Summary: 21 studied QSO fields: 13 shows galaxy overdensities and 8 do not show overdensities.

| Paper | z | N^a | Galaxy type ^b | Result |
|---------------------------------------|-----------|-------|--------------------------|------------------------------------------------------------|
| Mazzucchelli et al. (2015, submitted) | 5.7 | 1 | LAEs | No Overdensity |
| Simpson et al. (2014) | 7.1 | 1 | LBGs | No Overdensity |
| Morselli et al. (2014) | ~ 6 | 4 | LBGs | Overdensity in 4/4 fields |
| Bañados et al. (2013) | 5.7 | 1 | LAEs | No Overdensity |
| Husband et al. (2013) | ~ 5 | 3 | LBGs | Overdensity in 3/3 fields |
| Utsumi et al. (2010) | 6.43 | 1 | LBGs | Overdensity |
| Kim et al. (2009) | ~ 6 | 5 | LBGs | No Overdensity in 3/5 fields Overdensity in 2/5 fields. |
| Kashikawa et al. (2007) | 4.87 | 1 | LBGs and LAEs | Overdensity |
| Zheng et al. (2006) | 5.8 | 1 | LBGs | Overdensity |
| Stiavelli et al. (2005) | 6.28 | 1 | LBGs | Overdensity |
| Willott et al. (2005) | 6.2 – 6.5 | 3 | LBGs | No Overdensity in 3/3 fields |

^a Number of studied QSO fields.

^b Type of searched galaxies.

in proto–cluster environments (e.g. Intema et al., 2006; Overzier et al., 2008; Venemans et al., 2007).

At higher redshifts the environments of other classes of AGN, such as the optically-selected QSOs, are currently less well constrained. Most of the previous works focused on searching for galaxies around the most distant $z \gtrsim 5$ QSOs, and these results paint a diverse and rather confusing picture: Kashikawa et al. (2007); Stiavelli et al. (2005); Utsumi et al. (2010); Zheng et al. (2006), and Morselli et al. (2014) find a quite strong enhancement of galaxies compared to control fields around $z \sim 5 - 6$ QSOs, whereas Bañados et al. (2013); Willott et al. (2005), and Simpson et al. (2014) find no significant excess of galaxies around QSOs at $z \sim 6 - 7$. Kim et al. (2009) studied five fields of QSOs at $z \sim 6$ and reported a mix of overdensities and underdensities, and Husband et al. (2013) find galaxy overdensities in $z \sim 5$ QSOs environments, but they note that even some randomly chosen patches of sky contain similar galaxy overdensities without AGN as signposts (‘blank fields’) at the same redshift. Indeed, some galaxy overdensities at $z \sim 6$ have been identified in blank fields, found in Lyman break galaxies (LBGs) or Lyman alpha emitters (LAEs) surveys of a few deg² (e.g. Ota et al., 2008; Ouchi et al., 2005; Toshikawa et al., 2012). In Table 1.1 I present a summary of the protocluster search in AGN environments done so far at $z \gtrsim 5$.

These mixed results at $z \gtrsim 5$ have not made it possible to confirm that QSOs inhabit massive dark matter halos, and more studies are still needed to have a complete understanding on this topic.

One complication of these studies is that the majority of them are focused on dropout selection, which selects galaxies over a broad redshift range $\Delta z \sim 1$ (e.g. Ouchi et al.,

2004a), where possibly several galaxies are not physically related to the QSO, which introduces projection effects in the results.

On the other hand, several of these studies focused on the search of galaxies overdensity in individual QSOs fields, which produce poor statistics to draw strong conclusions. Additionally, cosmic variance effect could be important when fields are studied separately, and a correction for this effect could be very hard to do. One way to solve this problem, is to imaging several QSO fields at the same redshift and then stack the galaxy counts, then a QSO-galaxy clustering measurement is possible with higher signal to noise ratio. This method has been used before to measure QSO-galaxy cross-correlation function at the higher redshift so far, done by Trainor and Steidel (2012), who studied 15 luminous QSO fields at $z = 2.7$. The problem of doing this at $z \gtrsim 5$ is that too few QSOs are known, then the observation of a big sample of QSOs at the same redshift is not possible. Additionally the exposure time needed to imaging each QSO field at those redshifts is extremely high, then there is little chance to observe a very big sample.

Even If we could observe a big QSOs sample, an additional complication arises. The lack of a reliable galaxy luminosity function (LF) at these too high redshifts, implies a poor determination of the background number density, which, as we explain in the next chapter, is essential to measure the clustering properties of the sample. This would prevents a suitable analysis and interpretation of results in terms of clustering models.

Given all these challenges, we have chosen to study the environs of luminous QSOs at $z \sim 4$. This redshift election banishes the complication existing at $z \sim 6$ mentioned above as the LF of galaxies at $z \sim 4$ are well known (Ouchi et al., 2008, 2004a), as well as the clustering properties of QSOs and galaxies, which allows us to interpret the results properly.

A line of evidence that QSOs trace the rarest environments at $z \sim 4$ arises from their extremely strong clustering. Indeed, Shen et al. 2007 determined that QSOs at $z > 3.5$ have an auto-correlation length of $r_0 = 24.3 h^{-1} \text{Mpc}$, making them the most strongly clustered population in the Universe, and demanding that they reside in the most massive DM $M_{\text{halo}} \sim 10^{13} M_{\odot}$ halos at this epoch. Thus the generic prediction from hierarchical clustering is that galaxies should be very strongly clustered around QSO at $z \sim 4$. Observationally this should be reflected as a very strong QSO-galaxy cross-correlation function.

The halo masses predicted by QSO clustering at $z \sim 4$ are still very massive, then they represent high-sigma density peaks, and also the ‘characteristic’ L_* of galaxies can be imaged with relatively short integrations, which allows to observe a bigger sample of QSO fields.

So far very few studies of QSO environments at $z \sim 4$ are available. Adams et al. (2015) recently searched for LBGs in nine luminous QSOs environments at $z \sim 4$. They do not measure a correlation function with their data, but only search overdensities of galaxies around QSOs. They find evidence of galaxy overdensity in one field using additional spectroscopy data. However they are not able to confirm or rule out overdensities in the other eight fields only using imaging data, mainly due to the large uncertainties in the observed colors and photometric redshifts of their LBGs candidates. Clearly there is a lack of QSO-galaxy clustering studies at $z \sim 4$.

Finally, an efficient alternative to solve the projection effects, is to use narrow band imaging to search for LBGs or LAEs in a much narrow redshift range ($\Delta z \sim 0.1$) centered on the redshift of the QSO. This minimizes the line-of-sight contamination which is not possible to achieve using the classical broad-band dropout technique (see section 3.1 for an explanation of a method to select LBGs using narrow band filters).

Considering all this arguments, we prefer to focus on $z \sim 4$ QSO fields, using narrow band imaging to study the QSO-galaxy cross correlation function.

1.5 Outline of This Thesis

The main goal of my PhD thesis is to study the QSO-galaxy clustering properties at $z \sim 4$ in order to figure out if QSOs are situated in massive dark matter halos in the early Universe as the theoretical predictions suggest. This is the first time that the QSO-galaxy cross-correlation function at $z \sim 4$ is measured.

To accomplish this objective, I have used data from the ESO archive, corresponding to a program which was proposed with the objective of studying $z \sim 4$ QSO environments. The QSO fields were imaged using a set of narrow band filters in order to select LBGs using a novel technique. From this data we measured the QSO-LBG cross-correlation function at $z \sim 4$.

Additionally, I acquired extra data for this project during my PhD. This corresponds to 14 $z \sim 4$ QSO fields imaging in 3 filters to search for LAEs. This represent the largest sample at this redshift, which allows us to measure QSO-LAE cross-correlation function.

The outline of this thesis is as follows. In chapter 2 I summarize the basics of correlation functions, and I explain how to measure it in the especial case when we are situated in overdense regions of the Universe. In chapter 3 I describe the data from the ESO archive and I present the new technique used to select LBGs. Also, I describe the reduction of the data and the photometry. In chapter 4 I describe a sophisticated simulation used to

study the redshift selection function of our non-standard filters and I present the sample of selected LBGs. I present the results of the clustering analysis of LBGs in QSO environments in chapter 5. In the chapter 6 I present the complete data description, analysis and results for the QSO-LAE cross-correlation function. Finally, I summarize the main results and conclusion of this thesis in chapter 7.

Throughout this thesis, magnitudes are given in the AB system (Fukugita et al., 1995; Oke, 1974) and we adopt a cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.26$ and $\Omega_\Lambda = 0.74$. We have adopted “cMpc” notation to refer to comoving units and “pkpc” to refer to proper units.

Chapter 2

Basics on Correlation Function and Clustering Measurements

In chapter 1, I showed that galaxies are not randomly distributed in the Universe, they tend to be rather grouped and form structures. However, we need a mathematical formalism to quantify the clustering. In this chapter, I introduce the correlation function $\xi(r)$, which is the most used way to quantify the clustering of objects.

In section 2.1 I define the correlation function, then in section 2.2 I show how to measure it, in section 2.3 I explain how to get the physical parameters from this measurements, and finally in section 2.4 I explain how to measure clustering in the particular case where we are located on an overdense region, which is the case developed in this thesis.¹

2.1 Definition of the Correlation Function

In order to understand the concept of correlation function, we first imagine that galaxies are randomly distributed in the Universe with mean number density n_G . The probability of finding a galaxy in a volume element dV will be given by:

$$dP = n_G dV \tag{2.1}$$

If galaxies are randomly distributed, then the probability to find a second galaxy is independent, then the joint probability of finding a galaxy in the volume element dV

¹Part of the information here provided is based on the books Coil (2013); Schneider (2015); Wall and Jenkins (2003)

and in the same time to find another galaxy in the same volume element but at r distance from the first one is just the multiplication of both independent probabilities:

$$dP = n_G^2 dV^2 \quad (2.2)$$

However, as galaxies are not randomly distributed in the Universe, the probability of finding at the same time two galaxies separated by a distance of r is not independent. In other words, if I find a galaxy, it is more probable to find a second one close to this galaxy than at a randomly chosen location in the Universe. That means that there is an excess of probability to find a galaxy close to another one due to their are clustered.

In order to quantify this excess of probability, the two point auto-correlation function of galaxies $\xi_{GG}(r)$ is defined, which has to be introduced in equation (2.2) to take into account that galaxies are not randomly distributed, but rather have some level of correlation. So, the probability in equation (2.2) for a correlated distribution of galaxies is given by:

$$dP = n_G^2 dV^2 [1 + \xi_{GG}(r)] \quad (2.3)$$

Here, the auto-correlation function $\xi_{GG}(r)$ quantifies the excess probability, above a random distribution, of finding two objects, at the same time, at separation r , in a volume element dV . In this way, if $\xi_{GG}(r) = 0$ there is no excess probability, and we recover the random distribution case. On the other hand, if $\xi_{GG}(r) > 0$ then galaxies are clustered.

The auto-correlation function can be also written in terms of the relative density contrast defined in equation (1.1), as:

$$\xi_{GG}(r) = \langle \delta_G(r) \delta_G(r) \rangle \quad (2.4)$$

where $\delta_G(r)$ is the relative density contrast of galaxies. Here the angle brackets indicate that this is an average, then the averaged $\delta_G(r)$ in the Universe has to be used. This is an alternative definition of $\xi_{GG}(r)$ and it is completely equivalent of the one defined in equation (2.3).

If we are interested in how two different type of objects (for example quasars with galaxies) are correlated, the cross-correlation function $\xi_{QG}(r)$ is defined in a similar way as:

$$dP = n_Q n_G dV^2 [1 + \xi_{QG}(r)] \quad (2.5)$$

or in terms of the relative density contrast:

$$\xi_{QG}(r) = \langle \delta_Q(r) \delta_G(r) \rangle \quad (2.6)$$

From the available surveys, the galaxy auto-correlation have been measured (e.g Hawkins et al., 2003; Zehavi et al., 2011). It has been found that the galaxy correlation function is well modeled by a power law as:

$$\xi_{GG}(r) = \left(\frac{r}{r_0^{GG}} \right)^{-\gamma^{GG}} \quad (2.7)$$

where r_0^{GG} is the correlation length, which indicates the amplitude of the clustering, then the higher the r_0^{GG} value, the more clustered the population is.

It has been found that the r_0 value changes with the scale over which it is measured, with the redshift and with galaxy properties as color and luminosity. In particular, it is found that more luminous galaxies are more clustered, and for a fixed luminosity, redder galaxies are more clustered in comparison with the bluer ones (Zehavi et al., 2011).

On the other hand, the clustering of quasars has been also studied, and they show to have similar clustering as local galaxies at least at $z < 2$, with $r_0 \sim 5 \text{Mpc h}^{-1}$ (Porciani and Norberg, 2006). However, a strong increment of the clustering is found at higher redshifts, with $r_0 \sim 16.9 \text{Mpc h}^{-1}$ at $2.9 < z < 3.5$, and $r_0 \sim 24.3 \text{Mpc h}^{-1}$ at $3.5 < z < 5.4$ (Shen et al., 2007).

2.2 Estimators of $\xi(r)$

Several estimators have been proposed to measure the cross and auto correlation functions. In section 2.1 we defined the correlation function as the excess probability, above a random distribution, of finding two objects at separation r . Then the more basic way to measure this for a sample of objects is counting the number of objects pairs with separation r and comparing this quantity with the number of objects pairs with separation r if their were randomly distributed.

Under this basic idea, some estimators have been created. For all of them it is necessary to create a so-called random catalog. This is a catalog of sources randomly distributed

over the same three-dimensional geometric coverage as the data. This means that I have to know both the angular and redshift selection of the data.

Having the data and the random catalog, we can compute the quantities DD , DR and RR which are the number of data-data, data-random and random-random pairs within a bin covering a specific range of radius. The estimators for the correlation function are all based on those quantities. The natural estimator is defined as:

$$\xi(r) = \frac{DD}{RR} \left(\frac{n_R}{n_D} \right)^2 - 1 \quad (2.8)$$

where n_D and n_R are the mean number density of data and random sources. Another very used estimator is one proposed by Davis and Peebles (1983):

$$\xi(r) = \frac{DD}{DR} \frac{n_R}{n_D} - 1 \quad (2.9)$$

(Hamilton, 1993) also introduce an estimator:

$$\xi(r) = \frac{DDRR}{(DR)^2} - 1 \quad (2.10)$$

Finally the mostly used is from Landy and Szalay (1993):

$$\xi(r) = \frac{DD}{RR} \left(\frac{n_R}{n_D} \right)^2 - \frac{2DR}{RR} \left(\frac{n_R}{n_D} \right) + 1 \quad (2.11)$$

For large volumes, all the estimators are equivalent, but at smaller volumes, there are some differences related with statistical properties (robustness, bias, variance, etc). Some studies have compared the different estimators, and as a general consensus, for astrophysical applications, the more recommended estimator is the Landy & Szalay estimator (e.g. Kerscher et al., 2000). This is because it is less sensitive to the size of the random catalogs and handles edge correction, which is important when dealing with a small survey area or with weak clustering.

2.3 Estimation of r_0 and γ Parameters

I have shown how to measure the correlation function from a data set, but now we are interested in how to obtain the physical parameters r_0 and γ , which better describe the clustering of the data.

I mentioned that the correlation function typically has the form of a power law (equation 2.7), where $\xi(r)$ is directly related to r_0 and γ . However, to estimate $\xi(r)$ we need to compute r , which is the three-dimensional distance between the galaxies. Then, in principle, we need both the angular and redshift information of the data, which is not the most common case, especially in high-redshifts studies. Even when we have the redshift information, it is hard to compute an exact value of r , as I explain in 2.3.1.

Here I summarize the different correlation functions that can be measured depending on the available information. I also describe how is the standard procedure to obtain r_0 and γ from the measurements.

2.3.1 Projected Correlation Function

This correlation function is measured for cases in which we have the most complete information: RA, DEC and the spectra, from which the redshift can be computed. It is intuitive to think that here the computation of r is straightforward, but it is not completely true.

Using the angular information, we can easily compute the projected distances between galaxies, R . However the redshift information does not give us an exact estimation of the distance of the galaxy. If we assume a cosmological model, we can compute distances from redshifts. However, objects in the Universe are moving because of the expansion, but they also have peculiar velocities in the line of sight. When the redshift is measured, both movements are included, and this introduce redshift space distortions (Sargent and Turner, 1977). If two galaxies are located at the same redshift, we could measure a small difference between the observed redshift just because the peculiar velocity is different. This makes that the computation of physical distances from redshift is not a good tool for clustering computation.

The good new is that those peculiar velocities only affect to the line of sight distances (π), so the transverse distance is not affected. For this reason, in this case, the projected correlation function $\omega(R)$ is measured, where R is the projected distance between galaxies, which is computed just using the RA and DEC positions. One way to measure $\omega(R)$ is using the estimators defined in 2.2 to compute the redshift-space correlation function $\xi_s(R, \pi)$ for different R and π bins, where π is the line of sight separation between galaxies ($\pi = c\Delta z/H(z)$). The distortions effects on the redshift space can be projected out by integrating $\xi_s(R, \pi)$ along of π to obtain the projected correlation function:

$$\omega(R) = \int_{-\infty}^{\infty} \xi_s(R, \pi) d\pi \quad (2.12)$$

If we integrate the real-space correlation function $\xi(r)$ in the physical space along the line of sight, we will obtain the same result since the radial velocities just move the points inside the volume, then $\omega(R)$ can be written using directly $\xi(r)$ as:

$$\omega(R) = \int_{-\infty}^{\infty} \xi(R, Z) dZ \quad (2.13)$$

where Z is the redshift space distance in physical units (the radial comoving distance) given by:

$$dZ = \frac{c}{H(z)} dz \quad (2.14)$$

with z the redshift. Note that we have written explicitly the dependence of the correlation function as $\xi(R, Z)$ since the real space separation r can be decomposed by $r^2 = R^2 + Z^2$. If a power law form is assumed for $\xi(R, Z)$ as:

$$\xi(R, Z) = \left(\frac{\sqrt{R^2 + Z^2}}{r_0} \right)^{-\gamma} \quad (2.15)$$

then the equation (2.13) can be analytically solved and the parameters r_0 and γ can be directly related with $\omega(R)$ as:

$$\omega(R) = R \left(\frac{r_0}{R} \right)^\gamma \frac{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \quad (2.16)$$

with the $\Gamma(x)$ the Gamma function.

In this way, using the equation (2.12), $\omega(R)$ can be measured, and using the equation (2.16) the measured $\omega(R)$ can be related with the physical parameters r_0 and γ (e.g Coil et al., 2007; Krumpel et al., 2015; Shen et al., 2007)

2.3.2 Volume Averaged Projected Correlation Function

This correlation function is measured for cases in which we have the three dimensional information (RA, DEC, redshift), but is also applicable for cases in which we do not have redshift information, but the technique to select the galaxies ensures a narrow range for the redshift distribution of them. For example, in the cases used in this thesis LBGs and LAEs are selected using narrow bands, which allows us to select galaxies in a narrow redshift range. The approach here described is the one used in this thesis.

The volume average correlation function $\chi(R)$ is defined as the real space correlation function $\xi(r)$ integrated over the volume and normalized by the volume:

$$\chi(R) = \frac{1}{V} \int_{Vol} \xi(R, Z) dV \quad (2.17)$$

This is similar to equation (2.13), but now we are integrating over the entire volume (radial bin and comoving distance components) instead of only the redshift space. Also, $\chi(R)$ is a dimensionless quantity while $\omega(R)$ have distance units. This method arises from the need to take into account the fact that $\omega(R)$ may change over the bins in R when the bins are chosen to be large due to a low number count to measure clustering (e.g. Hennawi et al., 2006).

If a power law form is assumed for $\xi(R, Z)$, as in the equation (2.15), and a cylindrical volume with height ΔZ , and a radial bin defined by R_{min} and R_{max} , then $\chi(R)$ is related to the r_0 and γ parameters by the equation:

$$\begin{aligned} \chi(R) &= \frac{1}{\pi(R_{max}^2 - R_{min}^2)\Delta Z} \int_{Z_{min}}^{Z_{max}} \int_{R_{min}}^{R_{max}} \left(\frac{\sqrt{R^2 + Z^2}}{r_0} \right)^{-\gamma} 2\pi R dR dZ \\ &= \frac{1}{\pi(R_{max}^2 - R_{min}^2)\Delta Z} \frac{2\pi r_0^2}{(2-\gamma)} \left[\int_{Z_{min}}^{Z_{max}} \left(\frac{R_{min}^2}{r_0^2} + \frac{Z^2}{r_0^2} \right)^{\frac{2-\gamma}{2}} dZ \right]_{R_{min}}^{R_{max}} \end{aligned} \quad (2.18)$$

Here, the integral can be solved numerically for Z , and assuming different values for r_0 , γ we can find the model which best fits to our measured $\chi(R)$. The measurement of $\chi(R)$ is done by using one of the estimators described in section 2.2. This is what we actually do for computing the clustering parameters in chapter 5.

2.3.3 Angular Correlation Function

This correlation function is measured for cases in which we just have the two dimensional information of the galaxies: RA and DEC, so this represents probably the most common case. In this case, it is only possible to compute angular distances between them, and then it is just possible to compute the angular correlation function $\omega(\theta)$, where θ is the angular distance between galaxies. Here, as we do not have redshift information we cannot work in physical units as in the case of projected correlation function, then just angular units are used.

When we compute $\omega(\theta)$, we are combining galaxies at very different redshifts, but the three-dimensional correlation of galaxies in space implies their angular positions will be

also correlated (Schneider, 2015). The definition of $\omega(\theta)$ is analogous to the equation (2.3), then the joint probability of finding a galaxy in the solid angle element $d\Omega$ and at the same time to find another galaxy in the same solid angle element but at θ distance from the first one is:

$$dP = n_G^2 d\Omega^2 [1 + \omega_{GG}(\theta)] \quad (2.19)$$

The angular correlation function can be measured directly using an estimator of the section 2.2. Since $\xi(r)$ is typically assumed to be a power law, then $\omega(\theta)$ will also have a power law form, as:

$$\omega(\theta) = A_\omega \theta^{-(\gamma-1)} \quad (2.20)$$

If the redshift distribution of the galaxies is known, then the $\omega(\theta)$ can be related directly with the real space correlation function $\xi(r)$ by using the Limber equation (Limber, 1953), which is computed by integrating $\xi(r)$ along the line of sight. Then, $\omega(\theta)$ is related to the r_0 and γ parameters as:

$$\omega(\theta) = \sqrt{\pi} \frac{\Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \frac{A}{\theta^\beta} \quad (2.21)$$

where Γ is the gamma function, and A is given by:

$$A = \frac{\int_0^\infty r_0^\gamma(z) g(z) \left(\frac{dN}{dz}\right)^2 dz}{\left[\int_0^\infty \left(\frac{dN}{dz}\right) dz\right]^2} \quad (2.22)$$

where dN/dz describes the redshift distribution of galaxies and $g(z)$ is given by:

$$g(z) = \left(\frac{dz}{dr}\right) r^{(1-\gamma)} F(r) \quad (2.23)$$

where $F(r)$ is the curvature factor in the Robertson-Walker metric.

However, it should be noted that the Limber equation is valid in the regime when the maximum R separation in the survey is much lower than the covered ΔZ (e.g. Limber, 1953; Simon, 2007). Then for example for cases where galaxies are selected using narrow band technique, this equation will be probably inadequate. However, for standard LBGs selection, where the covered ΔZ is big, this procedure works properly (e.g. Ouchi et al., 2004b).

2.4 The Special Case of Computing Clustering on a Over-density

In this thesis, we are focusing in the measurement of clustering properties in very special locations in the Universe. We are supposedly located on a matter overdensity, in the most massive dark matter halos, where the matter density is much higher than the mean matter density in the Universe.

Because of this, we cannot follow the same standard procedure as other studies, and we have to measure clustering in a slightly different way. As I mentioned in section 2.1, the galaxy correlation function is computed by comparing the observed galaxy number counts with what is expected for a randomly distributed sample with a density equal to the mean density in the Universe.

Given that typically the clustering measurements of galaxies are done in random locations in the Universe, the galaxy density of the survey is representing in a good way the mean number density of the universe, especially when the survey is covering a big area in the sky. In this case, the galaxy number counts expected for a randomly distributed sample can be computed from the data itself. This means that the number density of galaxies in the survey is computed, and then a mock catalog of sources is created with the same number density. They are randomly distributed on the image, in order to mimic the angular selection function of the survey. Through the use of this catalog the correlation function can be computed.

In our case, we are situated on an overdense region, therefore the mean density of galaxies in our survey is not representing the mean density of galaxies in the Universe, thus, we cannot create the random catalog from the data itself. Instead, we need to theoretically compute how many galaxies have to be created in our images, such that their density be equal to the mean number density of galaxies in the Universe. If we use the luminosity function of galaxies computed in other studies, we can compute this number density, but for that, we need to know perfectly the geometry of the survey. This implies to know both the angular and the redshift selection functions.

The angular selection function is typically easy to trace using masked images, but the redshift selection function is more complicated to know, because it depends on the filter set used to select galaxies, and also in the criteria used to select them. One option is to observe control fields, which are fields observed using the same filter configuration, but in random locations in the Universe (i.e not centered on QSOs). If we had control fields, we could use them to measure the mean number density from the data itself, and we ensure we have the same redshift selection function. Another option is to use

a luminosity function of galaxies, computed using exactly the same filter configuration and selection as ours, then we can ensure that the geometry of the survey is the same.

However, if we have different filter configuration as previous works and we do not have control fields, then we need to perform a simulation to estimate the redshift selection function. If we compute the redshift selection function, then we can estimate the galaxy number counts to be created for the random catalog. After doing this, we randomly distributed them on the images and we measure the correlation function using any of the estimators described in section 2.2.

Note that in the case of overdensities, the DD , DR , and RR factors in those equations do not have to be normalized to the number counts. Since we need to measure how much overdense the region is, we need to use the estimators in units of number of galaxies, instead of normalize them as is typically done.

Chapter 3

Observations, Data Reduction and Photometry

The data set presented in this Chapter was obtained from the ESO Archive (Program ID: 079.A-0644). This program was designed in 2006 in order to search for LBGs in $z \sim 4$ QSOs environments using a novel narrow band filter technique. The aim of this program was to test whether QSOs with the most massive black holes at $z \sim 4$ live in the most massive dark matter halos.

In the first section, I explain the narrow band technique used to select LBGs. In the section 3.2 and 3.3 I summarize the strategy used to select the QSOs to be observed and I give details of the imaging observations. Finally, in the section 3.4 and 3.5 I give a detailed description of the data reduction and photometry.

3.1 A Nobel Method to Detect LBGs

The intergalactic medium (IGM) is the gas existing in the space between galaxies, which is mainly composed by hydrogen and helium. In the local Universe the gas is fully ionized, but at earlier times a fraction of this gas was neutral. It is found that the neutral gas fraction increases with increasing redshift (e.g. Fan et al., 2006).

Considering this scenario, we expect that a large fraction of photons emitted by high redshift sources with $\lambda < 1216\text{\AA}$ be absorbed by the neutral hydrogen of the IGM. This absorption is, as a matter of fact, observed in the spectra of high redshift QSOs or galaxies, where the spectra display a flux break at $\lambda = 1216\text{\AA}$ (e.g. Morganson et al., 2012).

This flux break has been used to detect galaxies at high redshift without the need of taking their spectra. The galaxies detected by this method are called Lyman break galaxies (LBGs) or dropouts. The traditional Lyman break method consists in the use of two broad bands, one located at $\lambda < 1216(1+z)\text{\AA}$, and the other one at $\lambda > 1216(1+z)\text{\AA}$. Given this configuration, it is expected to find a non detection or very faint detection in the band located at $\lambda < 1216(1+z)\text{\AA}$, while a clear detection is expected in the other band, so, a very red color has to be measured. Typically, a third band is added in order to eliminate possible contaminants. (Steidel et al., 1995).

When we study forming structures at high redshift such as protoclusters, we need to select galaxies gravitationally associated to each other. However, the traditional Lyman break method selects efficiently LBGs in a broad redshift slice of $\Delta z \sim 1$ (e.g. Ouchi et al., 2004a), which corresponds to $\sim 520\text{cMpc}$ at $z = 4$. In this broad redshift range, it is likely that an important fraction of galaxies will not be associated, therefore this methodology introduces many false counts in the galaxy number density of a protocluster.

In order to solve that problem, a novel technique we have used, which is quite similar to the standard method, with the difference that the selection of LBGs is performed using two narrow bands (NBs) located very close to each other instead of using broad bands (see for example Fig. 3.1). The advantage of using NBs filters is that they would allow to select LBGs in a very narrow redshift range of $\Delta z \sim 0.2$ ($\sim 104\text{cMpc}$ at $z = 4$) (see section 4.4), which is ~ 5 times smaller than in the standard method. This technique allows to minimize the line-of-sight contributions to the galaxy counts.

This method has never been used before, and because of this the filters used to perform the observations were designed especially for this program, in order to identify LBGs at $z \sim 3.78$.

3.2 QSO Targets

Six QSOs were selected in order to study their environments properties. As we are using a narrow band technique to select galaxies in QSO environments, the QSOs targets have to span a very narrow redshift range, centered in $z = 3.78$, since the filters were designed to select galaxies at that redshift.

Taking advantage of the large sample of SDSS QSOs, it was selected all QSOs in the narrow redshift range of $z = 3.78 \pm 0.04$. In SDSS there were 70 QSOs fulfilling this requirement. Due to the interest of studying the most massive dark matter halos at $z \sim 4$, only the QSOs with the most massive black holes were selected, which are

TABLE 3.1: Targeted Quasars properties.

| Name | RA (J2000) | DEC (J2000) | Redshift | i | $\log(M_{\text{BH}}/M_{\odot})^a$ |
|------------------------|-------------|--------------|----------|-------|-----------------------------------|
| 2207+0043 | 22:07:30.48 | 00:43:29.37 | 3.767 | 19.47 | 9.13 ± 0.16 |
| 0124+0044 | 01:24:03.78 | 00:44:32.67 | 3.834 | 17.99 | 10.15 ± 0.03 |
| 0213-0904 | 02:13:18.98 | -09:04:58.28 | 3.794 | 19.03 | 9.57 ± 0.18 |
| 2003-3300 ^b | 20:03:24.12 | -32:51:45.02 | 3.773 | 17.04 | 9.7 |
| 2311-0844 | 23:11:37.05 | -08:44:09.56 | 3.745 | 20.18 | 9.41 ± 0.24 |
| 2301+0112 | 23:01:11.23 | 01:12:43.34 | 3.788 | 19.44 | 8.55 ± 0.80 |

^a Virial BH masses from Shen et al. (2011). ^b This QSO was not selected from SDSS, but it was targeted because it belongs to the redshift range of interest. The properties shown here are from (McLeod and Bechtold, 2009), who do not report the error for the BH mass measurement.

objects with $M_{\text{BH}} \gtrsim 10^9 M_{\odot}$. As is usual, the M_{BH} was estimated from the emission line widths and continuum luminosities (Vestergaard, 2002). This sample is then formed by bright QSOs with $i < 20.2$ mag.

We checked that none of the QSOs had a reported radio emission counterpart in the the VLA FIRST¹ catalog, since it is known that radio emission could strongly affect the galaxy clustering properties in AGN environments (e.g. Fanidakis et al., 2013; Venemans et al., 2007). A summary of the QSOs properties are shown in table 3.1, where we show a more recent M_{BH} estimation taken from Shen et al. (2011).

3.3 Imaging Observations

Imaging observations were acquired in 3 consecutive nights during 2007 September 9-11², using the FORS1³ instrument on the Very Large Telescope (VLT). The Field of View (FoV) of FORS1 is 6.8×6.8 arcmin² which corresponds to $\sim 3 \times 3$ pMpc² at $z = 3.8$. The instrument scale is 0.251 arcsec/pix for 2×2 binning data.

An important requirement to study the QSO environments is to achieve accurate redshifts of the galaxies in order to select only galaxies that are physically related to the QSO. For this reason, in this study, a novel technique has been used to select LBGs described in section 3.1. Each QSO field was observed in two narrow bands (NB₅₇₁ : $\lambda_{\text{eff}} = 5657\text{\AA}$, FWHM = 187\AA and NB₅₉₆ : $\lambda_{\text{eff}} = 5947\text{\AA}$, FWHM = 116\AA) and one broad band (r_{GUNN} : $\lambda_{\text{eff}} = 6490\text{\AA}$). These narrow bands allow us to identify LBGs at $z \sim 3.78$ associated with the central QSO.

¹Faint Images of the Radio Sky at Twenty-centimeters

²Program ID: 079.A-0644

³FOcal Reducer and low dispersion Spectrograph 1 (Appenzeller and Rupprecht, 1992)

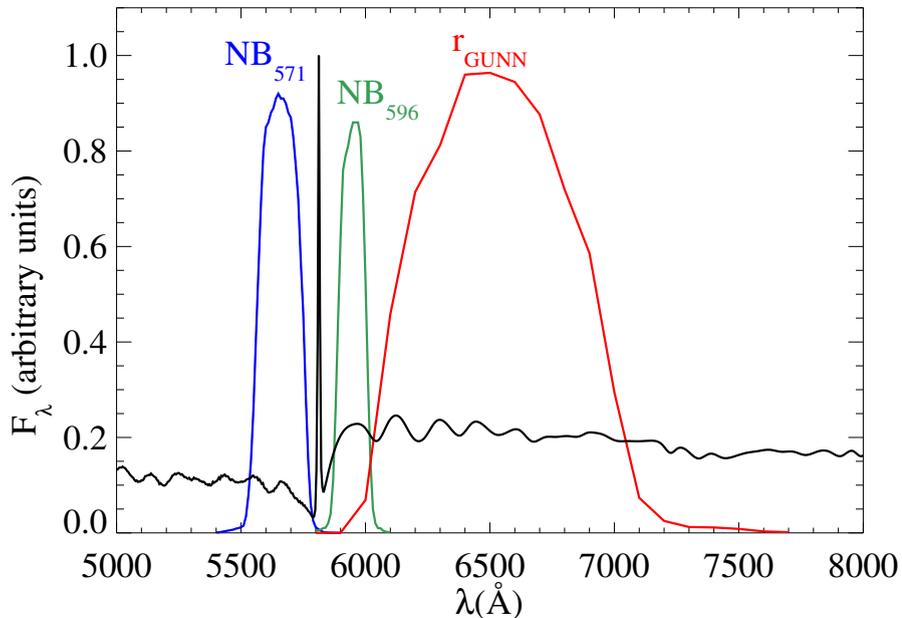


FIGURE 3.1: Filter configuration shown on a LBG simulated spectrum at $z = 3.78$ (see section 4.1 for the simulated spectra details). The narrow bands were designed especially for this program, in order to identify LBGs at $z \sim 3.78$ in a quiet narrow redshift slice of $\Delta z \sim 0.2$.

The total exposure time for the filters was 8000s, 4000s, and 1800s for NB₅₇₁, NB₅₉₆ and r_{GUNN} respectively. Observations were acquired in shorter separated exposures in a dithered mode in order to fill the gap between the CCDs and to facilitate the reduction process (cosmic rays and bad pixel rejection, superflat building, etc). At the beginning of the last 2 nights, spectrophotometric standard stars were observed. The typical seeing during the three nights was 0.6 - 0.8 arcsec.

3.4 Data Reduction

Science images were reduced using standard IRAF⁴ tasks and our own IDL codes. The reduction process included bias subtraction and flat fielding. As our images presented illumination patterns, it was preferred to perform the flat fielding using superflats images, created using the unregistered science frames. For that, we first masked all the objects out and then combined the science frames with average sigma-clipping algorithm.

SExtractor (Bertin and Arnouts, 1996) and SCAMP (Bertin, 2006) were used to detect sources on each individual image and to compute astrometric projection parameters

⁴Image Reduction and Analysis Facility

using SDSS-DR7 r-band catalogs. Finally, the individual images were sky-subtracted, re-sampled and median-combined using SWarp (Bertin et al., 2002). The noisy edges of the combined images were cut and the bright stars were masked in order to avoid wrong objects detection due to star flux contamination.

Because spectrophotometric standard stars for the first night were not observed, we took advantage of two existing SDSS star spectra in one of the fields taken during that night to do the flux calibration. These spectra were convolved with the three filters curve in order to obtain standard magnitudes. These magnitudes were compared with the star instrumental magnitude (obtained using the MAG_AUTO of SExtractor on the combined science images) to obtain the zero-points (ZPs) for each filter. A mean final ZP was calculated from the 2 stars.

The flux calibration for the second and third night was performed using the observed spectrophotometric star SA109-949 (with available spectra from Stone (1996)). As the standard star tabulated spectrum have a poor sampling of 25\AA and we are using two very narrow filters, it is not reliable to estimate the ZPs in these bands using this spectra. So, the spectra was used to convolve only with the broad-band filter curve to obtain the r_{GUNN} ZP. After that, the differential ZPs from the first night were used to determine the narrow bands zero-points for the last two nights. The typical rms for the ZP measurements was ~ 0.13 mag in the case in which a star spectra was available, and ~ 0.21 mag when differential ZP was used.

3.5 Photometry

Object detection and photometry were performed using SExtractor. The r_{GUNN} band was used as a detection image, this means that objects were detected in that image. Subsequently the photometry was done on the found apertures in both narrow band images. The background was calculated in regions of 64 pixels in size and then recomputed locally in an annulus area of 24 pixels of thickness centered around the object. In order to maximize the detection of faint sources, the detection image was smoothed by applying a Gaussian filter of seeing FWHM of 3 pixels and size 5×5 pixels. Every group of at least 5 contiguous pixels having a value above 1.5σ (with σ the background RMS) was considered as an object. After the detection of the objects, the photometry was performed in the unsmoothed image.

In order to ensure an adequate color measurement we need to carry out a photometry in the same object area for the three different filters. Therefore, for each field, we convolved our images with a Gaussian kernel to worsen its PSF to match it with the worst seeing

TABLE 3.2: 5σ limit magnitudes per field measured in a $2''$ diameter aperture.

| Name | NB ₅₇₁ | NB ₅₉₆ | r _{GUNN} |
|-----------|-------------------|-------------------|-------------------|
| 2207+0043 | 25.80 | 25.27 | 25.62 |
| 0124+0044 | 25.94 | 25.47 | 25.68 |
| 0213-0904 | 25.80 | 25.20 | 25.38 |
| 2003-3300 | 25.79 | 25.14 | 25.53 |
| 2311-0844 | 25.78 | 25.36 | 25.60 |
| 2301+0112 | 25.79 | 25.31 | 25.67 |

image for each field ⁵. Then, the objects magnitudes were estimated by the MAG_APER parameter of SExtractor using a fix aperture of $2''$ of diameter. This magnitude is not the total magnitude of the object, but is used to compute the colors of galaxies. With this choice, if galaxies at $z \sim 4$ look as point sources, we are including the flux of 3σ of the object's PSF (for a seeing of 0.8 arcsec). This ensures that we measure the majority of the object's flux, as well as avoid fluxes from other close sources. Magnitudes of objects not detected or detected with a signal to noise ratio lower than 2 either in NB₅₇₁ or NB₅₉₆ were replaced by the corresponding 2σ limiting magnitude.

Magnitudes were corrected by airmass, using the atmospheric extinction curve over Cerro Paranal (Patat et al., 2011), and by galactic extinction calculated using the Schlegel dust maps (Schlegel et al., 1998) and extinction laws of Cardelli et al. (1989) with $R_V = 3.1$. The mean 5σ limiting magnitude of the reduced images was of 25.82 for NB₅₇₁, 25.29 for NB₅₉₆ and 25.58 for r_{GUNN} for a $2''$ diameter aperture. The limit magnitudes per field are showed in Table 3.2.

⁵Specifically, for each field, we computed the seeing of the images in each of the three filters, and determined the bigger one. Then, we computed the standard deviation for that image ($\sigma_{big} = FWHM/2.36$) and then we convolved each of the other two images with a Gaussian function with standard deviation $\sigma = \sqrt{\sigma_{big}^2 - \sigma_{small}^2}$, where σ_{small} is the standard deviation computed for the image to be convolved. In that way, each final image has a seeing given by $2.36\sigma_{big}$

Chapter 4

Lyman Break Galaxies Selection at $z \sim 4$

The selection of LBGs candidates at $z = 3.78$ can be done using the dropout technique. As I explained in section 3.1, this technique is based in the detection of the Lyman break at $\lambda_{\text{Rest-frame}} = 1216$, observed in galaxies at high redshift due to the neutral hydrogen absorption by the IGM. Our two narrow band filters are used for this first selection. We expect that LBGs at $z \sim 4$ have red colors in $\text{NB}_{571} - \text{NB}_{596}$. The thinness and closeness in wavelength of our filters allows us to select LBGs in a narrow redshift range, minimizing line-of-sight contaminants. By using only this color criteria, we could be including some interlopers in the sample. In order to remove them, a third filter is used to have a measurement of the LBGs continuum slope. This can be monitored by the $\text{NB}_{596} - \text{r}_{\text{Gunn}}$ color. In this way the most efficient method to find LBGs is selecting them in the $\text{NB}_{571} - \text{NB}_{596}$, $\text{NB}_{596} - \text{r}_{\text{Gunn}}$ color-color diagram.

In the next sections I detail how the color criteria was chosen in order to select the LBGs sample. Since the filters used in this study are not standard filters, we need to explore how LBGs and low- z galaxies (possible contaminants) populate the color-color diagram in order to perform color cuts which recover the LBGs with the highest completeness and at the same time avoiding low redshift interlopers. Furthermore, as I mention in section 2.4, the modeling of the redshift space selection function of the sample is essential for clustering measurements in overdensities, so we have to perform a very careful completeness assessment. To accomplish this task we created a sophisticated simulation to model the LBGs colors on the color-color diagram. The details of this simulation are presented in section 4.1.

In section 4.2 I show the location of low redshift galaxies in the color-color diagram, which have to be taken into account to choose the optimal selection region. In section

4.3 I show the final color-cuts used, based on both the LBGs color modeling and low- z galaxies location, and I present the final LBGs sample. In section 4.4 I compute the completeness of the sample which also defines the redshift selection function, and I estimate the redshift range (dz) our filters are mapping.

4.1 LBGs Color Modeling

In order to figure out how LBGs populate the color-color diagram, we performed a Monte Carlo simulation where we simulate several LBGs spectra spanning the full range of observed spectral properties.

As a starting point, we considered a public galaxy template spectrum generated from Bruzual and Charlot (2003)¹. Then, we used this template to simulate 1000 new LBGs spectra, modifying its UV continuum slope and Ly α equivalent width (EW) such that our simulated spectra were reproducing all the possible LBGs spectra, based on our knowledge from observed LBGs properties.

The template used initially corresponds to an instantaneous burst model with an age of 70Myr, a Chabrier (2003) IMF and a metallicity of $0.4Z_{\odot}$, as is expected from LBGs at $z \sim 4$ (Jones et al., 2012). The age of the model was chosen to be the same as the used by Ouchi et al. (2004a), who simulated the evolutionary track of a LBG at $z \sim 4$. This allows us to reproduce their track, and then check that our simulation was working properly. Since we just had templates for ages of 25Myr and 100Myr, we performed a linear interpolation to obtain the model with age of 70Myr.

We have assumed a simple LBG spectral model given by an IGM attenuated power law continuum and a Gaussian shape for Ly α line given by:

$$F_{\text{LBG}} = T_z(\lambda)A\lambda^{\alpha} + B \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\lambda - \lambda_{\text{Ly}\alpha}}{\sigma}\right)^2} \quad (4.1)$$

where $T_z(\lambda)$ is the transmission function of the IGM at redshift z and α is the UV continuum slope measured in the wavelength range $1300 - 2000\text{\AA}$.

Each simulated spectrum is created to have an α value taken randomly from a Gaussian distribution with mean $\mu = -1.676$ and $\sigma = 0.39$ as is shown in the left panel of Fig. 4.1. We calculated these values from Bouwens et al. (2009), who presented the UV continuum slopes distribution of LBGs at $z \sim 4$ for samples selected in different magnitude ranges. We took a mean μ and σ value. Since the template used did not

¹Templates obtained from <http://bruzual.org/>

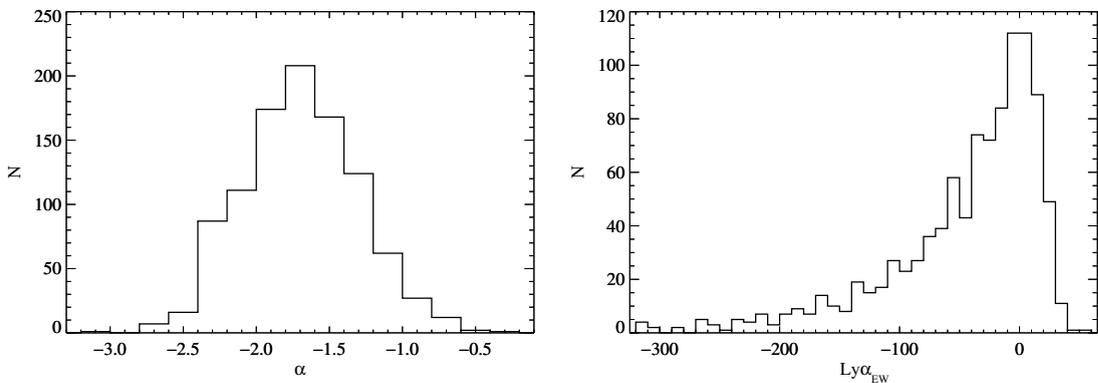


FIGURE 4.1: Distribution of the UV continuum slope α and $EW_{Ly\alpha}$ parameters for the simulated spectra. *Left panel:* UV continuum slope (α) distribution for the 1000 simulated spectra. α was chosen from a Gaussian distribution with mean $\mu = -1.676$ and $\sigma = 0.39$ Bouwens et al. (2009). α was defined as the slope measured in the wavelength range $1300 - 2000\text{\AA}$. *Right panel:* $Ly\alpha$ EW distribution for the simulated spectra. Negative values correspond to emission lines. EW was chosen from a Gaussian distribution with mean $\mu = -25\text{\AA}$ and $\sigma = 40\text{\AA}$ Shapley et al. (2003) plus an exponential tail of high EW values with scale length of $W_0 = -64\text{\AA}$ Ciardullo et al. (2012).

include the $Ly\alpha$ emission line, we added it. We assumed a fix $\sigma_{Ly\alpha}$ value for this line to be 1\AA which fitted with the $\sigma_{Ly\alpha}$ of the composite spectrum of LBGs at $z \sim 4$ (Jones et al., 2012). We chose different B values for the spectral model, such that the $EW_{Ly\alpha}$ was a value taken randomly from the galaxies' EW distribution at $z \sim 4$.

The EW distribution was given by a Gaussian distribution with mean $\mu = -25\text{\AA}$ and $\sigma = 40\text{\AA}$. We chose these EW values from Shapley et al. (2003), who studied the spectra of 1000 LBGs at $z \sim 3$. Here, we have assumed that the EW distribution of LBGs did not evolve significantly from $z \sim 3$ to $z \sim 4$. Also, we have modified this EW distribution in order to consider the tail of high EW values of LAEs given by an exponential function with rest-frame scale length of $W_0 = -64\text{\AA}$ as is presented in Ciardullo et al. (2012) (this is $N \propto \exp(-EW/W_0)$). In this way our models includes both LBGs and LAEs spectra. We calculated the EW as:

$$EW = - \int_{\lambda_i}^{\lambda_f} \frac{F_{Ly\alpha}}{F_{cont}} d\lambda, \quad (4.2)$$

where we define negative values of EW for emission lines and positive for absorption lines. The distribution of EW used in our simulated spectra is shown in the right panel of Fig. 4.1. As $F_{Ly\alpha}$ is defined by a Gaussian with amplitude B, we can determine the B value in such a way that we obtain the desired EW value.

Some studies of high redshift LBGs and LAEs claim a relationship between the $EW_{Ly\alpha}$ and the continuum luminosity. Specifically, a large $EW_{Ly\alpha}$ deficit was found for high

values of the continuum luminosity (Ando et al., 2006; Ouchi et al., 2008; Shapley et al., 2003). However, other studies have found that this correlation is not statistically significant (Ciardullo et al., 2012; Nilsson et al., 2009). For our Monte Carlo simulation we have assumed that EW does not have correlation with the continuum luminosity.

Using the final α and $\text{EW}_{\text{Ly}\alpha}$ distributions, we created a 1000 LBGs spectra in rest-frame. Then, we dust-attenuated each spectrum by:

$$F_{\text{attenuated}} = \frac{F_{\text{intrinsic}}}{10^{0.4E_s(B-V)k(\lambda)}} \quad (4.3)$$

where $k(\lambda)$ is the starburst reddening curve from Calzetti et al. (2000), and $E_s(B-V)$ is the color excess of the stellar continuum given by $E_s(B-V) = (0.44 \pm 0.03)E(B-V)$ (Calzetti, 1997). We have adopted a color excess value of $E(B-V) = 0.16$ according to the values estimated for LBGs at $z \sim 3$ (Shapley et al., 2003) and also used to compute the evolutionary track of LBGs at $z \sim 4$ in Ouchi et al. (2008).

After the dust-attenuation, we multiplied the spectrum at $\lambda \leq 912\text{\AA}$ for the escape fraction, which is the fraction of HI ionizing photons able to escape from the galaxy to the IGM. Although this value is poorly observationally constrained, some studies suggest values from 0.04 to 0.14 (Fernández-Soto et al., 2003; Shapley et al., 2006). For our simulation, we used a fixed value of $f_{\text{esc}} = 0.05$, but we also tested our results using different values. The results of the simulated colors do not change because afterwards this part of the spectra is attenuated by the IGM transmission function.

All this parameters defined our final rest-frame simulated spectrum. Finally, we re-shifted each model spectrum to different redshift values with a step of 0.02 and ranging from 3.2 to 4.4. In the re-shifting process we used the IGM transmission model $T_z(\lambda)$ from Worseck and Prochaska (2011) to attenuate the flux blueward of Ly α line. They built every profile as the average of 1000 Monte Carlo IGM spectra, binned in 1\AA in the observed wavelength. They computed $T_z(\lambda)$ in a redshift coverage of $3.2 < z < 4.4$ in steps of 0.1^2 . We have interpolated between those models in order to obtain $T_z(\lambda)$ in steps of 0.02. This transmission model represent a much more updated model than the traditionally used from Madau (1995) which do not match well with some IGM parameters (as τ_{eff} , the column density distribution or mean free path, for example). Anyway, we also used their models to study the effect of this election in the results of the simulation, and we find no significant changes in the results.

²The models were provided by the author by internal communication.

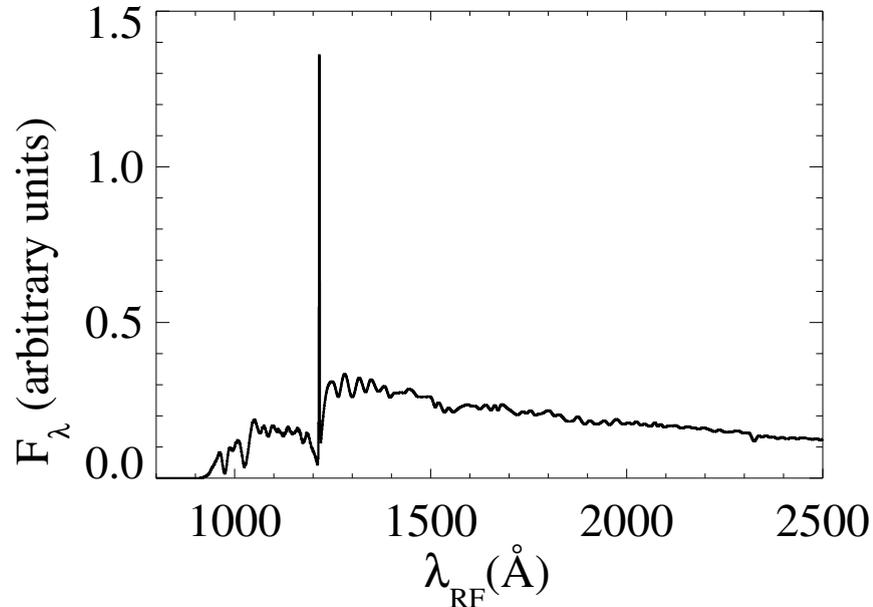


FIGURE 4.2: Mean LBG spectrum at $z = 3.78$. This is computed as the median of 1000 simulated LBGs spectra, which have different continuum slopes and $EW_{Ly\alpha}$.

For each redshift step, we convolved the simulated spectra with our filter transmission curves to obtain the LBG colors³. We also computed the median of our 1000 Monte Carlo spectra, and computed the colors to define the mean LBG evolutionary track. The mean spectrum of a LBG at $z = 3.78$ is shown in the Fig. 4.2 and the colors for the 1000 simulated spectra as a function of redshift are shown in Fig. 4.3. LBGs at $z \sim 3.6 - 3.7$ have blue $NB_{571} - NB_{596}$ colors, because the narrow band NB_{571} is located over the $Ly\alpha$ line, while the narrow band NB_{596} is over the continuum. LBGs at $z \sim 3.8$ show much redder colors because NB_{571} is located bluewards of $Ly\alpha$ lines. The green and yellow points in the upper left region of the plot are LBGs at $z \sim 3.9$ where NB_{596} falls just over the $Ly\alpha$ line, then, when the $EW_{Ly\alpha}$ value is high there is an excess flux in that band, which moves the color to redder $NB_{596} - R_{GUNN}$ values. When the $EW_{Ly\alpha}$ value is low, or is an absorption line, the galaxy become bluer, making that colors of LBGs at $z \sim 3.9$ be distributed in a wide region of this plot. Finally at higher redshifts both NBs are located blueward to $Ly\alpha$, then they have $NB_{571} - NB_{596} \sim 0$.

³The color is defined by

$$color = m_A - m_B = -2.5 \log(f_A/f_B) \quad (4.4)$$

where m_A , m_B , f_A , f_B are the magnitudes and fluxes on the bands A and B. The fluxes are computed convolving the spectrum flux with the transmission filter curve ($T(\nu_0)$) as in Hogg et al. (2002),

$$f_A = \int \frac{d\nu_0}{\nu_0} f_\nu(\nu_0) T^A(\nu_0) \quad (4.5)$$

where ν_0 is the observed frequency.

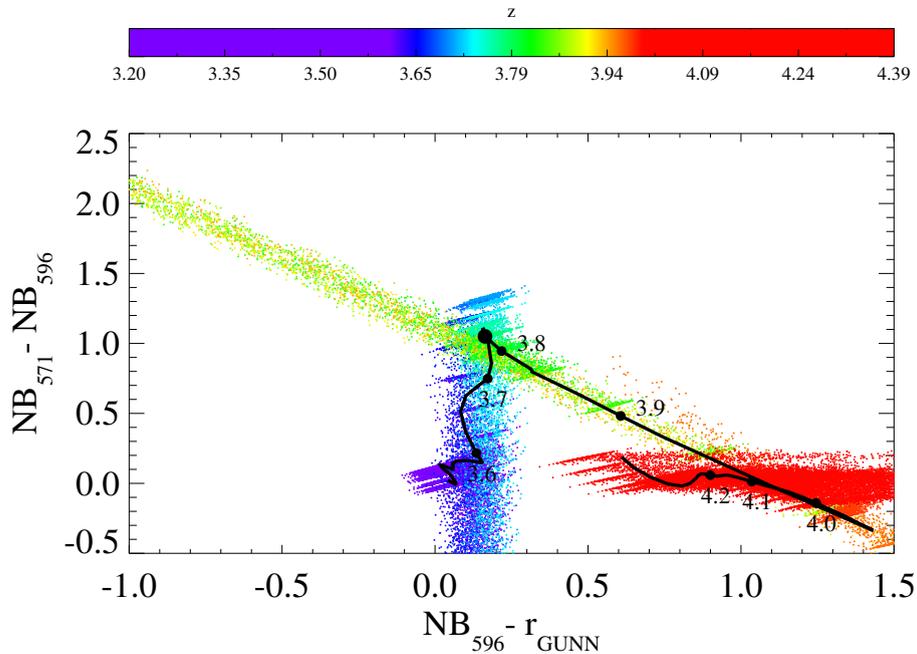


FIGURE 4.3: Color-color diagram showing the simulated colors for 1000 LBGs spectra, plotted as redshift color-coded points according to the color bar. The mean LBG evolutionary track is also plotted as a black curve. The filled points indicate the redshift from 3.6 to 4.2. The larger circle shows the exact position of the mean $z = 3.78$ LBG colors.

From the figure 4.3 we can define a preliminary selection region for LBGs at $z = 3.78$, which has to enclose the majority of the green points in the plot, excluding as much as possible the line-of-sight LBGs. In the next section we analyze how low- z galaxies populate this plot in order to also avoid them in the selection region.

4.2 Low Redshift Galaxies Colors

In the last section we defined the location of LBGs at different redshifts in the color-color diagram. That give us a first guess for the color cuts choice to select LBGs at $z = 3.78$. However, we need also to take into account the location of low- z galaxies in order to avoid contamination in our LBGs sample.

In principle, the most correct way to do that is doing a modeling as we did in the section 4.1 but for galaxies at low redshift. That means, we would have to model all the possible galaxy spectra, for different type of galaxies (elliptical, Sa, Sb, irregulars, etc) at low redshift and figure out how they populate the color-color plot. However, a much less time consuming way, is just to use some mean galaxy spectra with high resolution and compute the mean evolutionary track for these. This give us a basic understanding of

how low- z galaxies populate the color-color plot, but it cannot give us information about the exact contamination level we would have in our LBGs sample.

We used a set of 5 commonly used templates for estimating photometric redshifts, then these span the whole potential range of galaxy SEDs (see Fig. 4.4). The templates are from the code EASY (Brammer et al., 2008) which are distilled from the PEGASE spectral synthesis model. These spectra have a high resolution which is necessary since we are using narrow bands filters with $\text{FWHM} \sim 100 - 200 \text{ \AA}$, which trace the high resolution features of the spectra. We re-shifted the template spectra from $z = 0$ to $z = 3$, and convolved them with our filter transmission curves to generate their evolutionary track.

In the Fig. 4.5 we show those evolutionary tracks (different colors represent different galaxy types) together with the mean LBG evolutionary track that we computed in section 4.1 (showed as red curve). From this plot is possible to realize that the colors of low- z galaxies are very similar to the expected $z = 3.78$ LBG colors. This is a big complication, because this means that our filters cannot efficiently distinguish between them, then we are not able to select LBGs at $z \sim 4$ with high completeness and at the same time high purity.

In order to check that we are computing correctly the LBGs and low- z galaxies evolutionary tracks we computed those using our codes, but now convolving with the transmission curves of B, R, and i filters of the Suprime-cam on the Subaru Telescope in order to reproduce evolutionary tracks and color-cuts used by Ouchi et al. (2004a) to select LBGs at $z = 4 \pm 0.5$. We show our results in the left panel of the Fig. 4.6 where we plot the LBG evolutionary track computed with our code described in the last section as a red curve. We also overplotted the same low- z galaxy spectra used in the Fig. 4.5 and the color cuts used by Ouchi et al. (2004a) as a dashed line. In the right panel of the same figure we show a similar plot presented in their paper. The evolutionary tracks computed with our code for both the LBGs and low- z galaxies are consistent with those presented by Ouchi et al. (2004a) and the color cuts are the suitable choice to selecting correctly the position of LBGs at $z \sim 4$ in the plot.

Comparing the Fig. 4.5 and 4.6 is possible to realize that the evolutionary tracks using broad bands is much more well defined and constrained than using narrow bands. This makes that for narrow bands the low- z galaxies colors be located in a much wider region in the color-color plot. We attribute this to sensitivity of narrow bands to features in the spectrum such as emission or absorption lines. In the case of broad bands, these features are diluted, but in the case of narrow band, this features are traduced in extreme colors which move away on the color-color diagram.

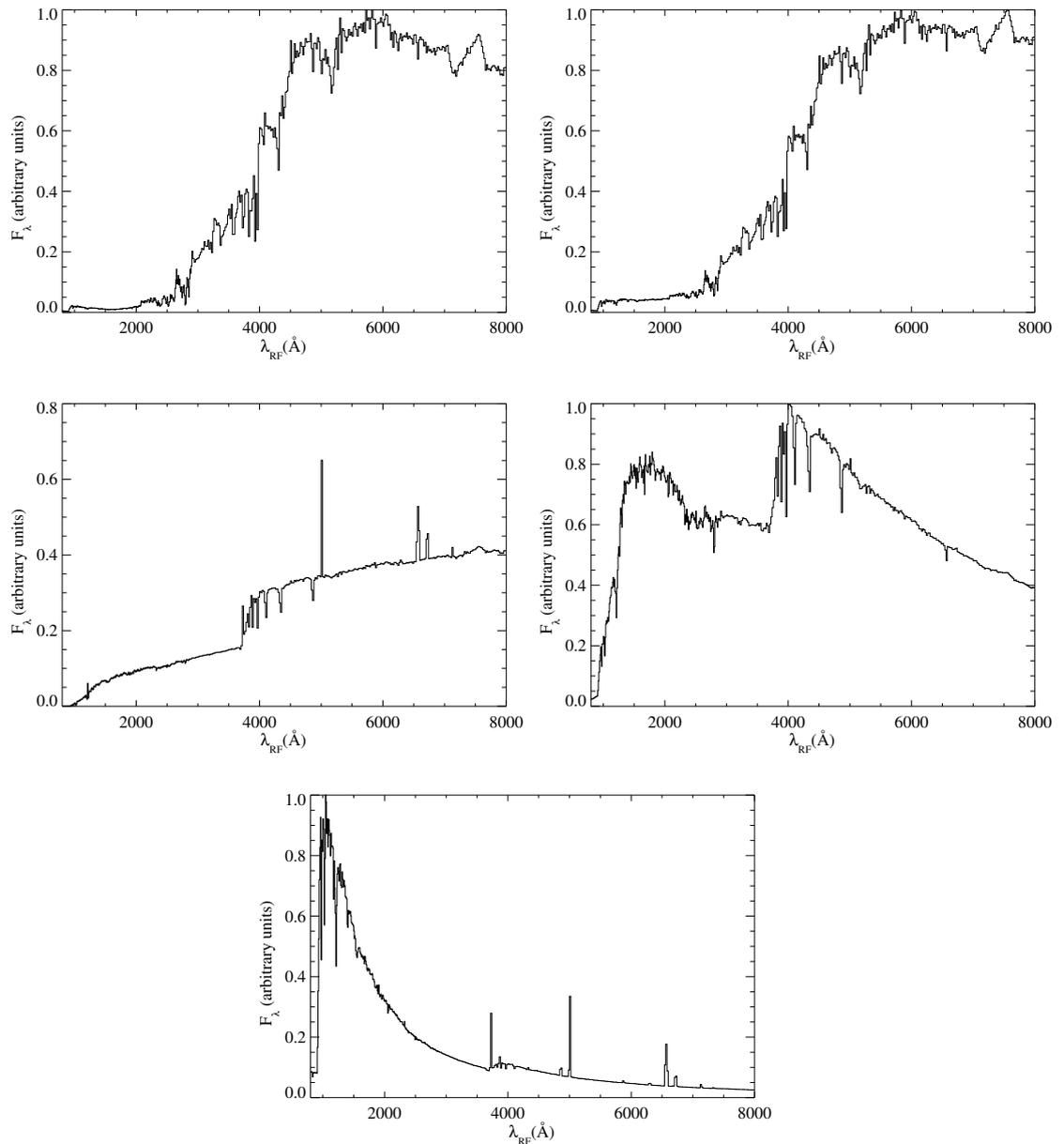


FIGURE 4.4: Spectral templates for 5 different types of galaxies used to figure out how low- z galaxies populate the color-color diagram. The galaxies shown from left to right and from top to bottom are Elliptical, Sa, Sb, Sc, and Irregular respectively.

We have several types of interlopers affecting the LBGs selection. The first type are red galaxies at redshift ~ 0.45 having a large $\sim 4000\text{\AA}$ Balmer break. This break is located just between our two narrow bands, so they present red colors in NB₅₇₁–NB₅₉₆. The continuum just redward to $\lambda = 4000\text{\AA}$ typically increases significantly, then our broad band can not be used to eliminate this interloper (see top left panel of Fig. 4.7). The second type of interlopers are galaxies at $z \sim 1$ with strong MgI and MgII absorption lines at $\lambda = 2852$, and $\lambda = 2799\text{\AA}$ respectively. When the NB₅₇₁ is located over this absorption the filter NB₅₉₆ falls on the continuum, then, red colors are detected again.

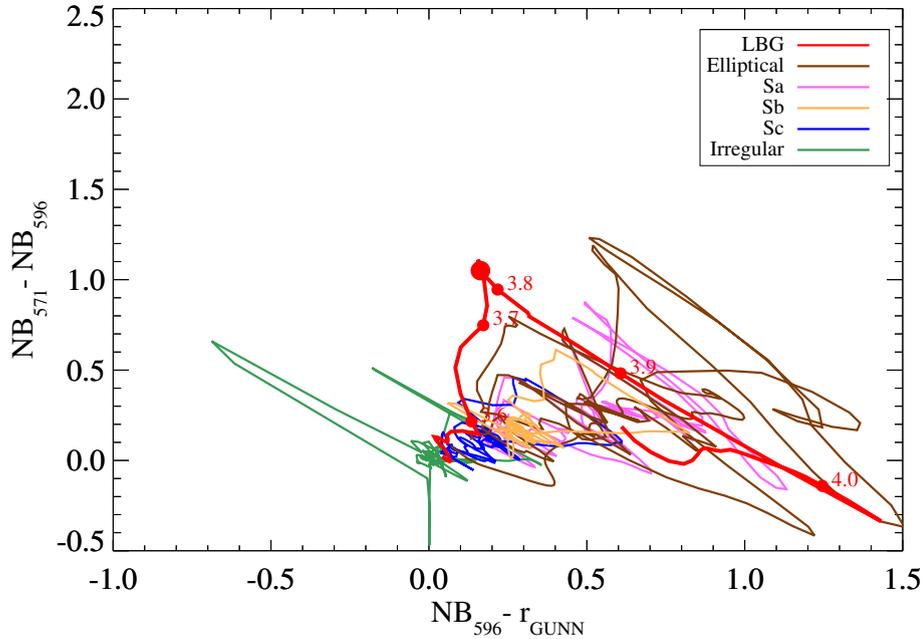


FIGURE 4.5: Low- z galaxies evolutionary tracks redshifted from $z = 0$ to $z = 3$. We plot as brown, magenta, orange, blue, and red curves the evolutionary track of elliptical, Sa, Sb, Sc, and Irregular galaxies, respectively. We overplotted the track of LBGs computed as was explained in section 4.1 as a red curve. Filled circles over the red curve indicate colors of LBGs from redshift 3.6 to 4.0, and the largest red point indicates the exact position of the color of LBGs at $z = 3.78$.

Typically the continuum flux at longer wavelengths increases with a steep slope, then again we have high flux in our broad band (see top right panel of Fig. 4.7). Finally, other interlopers are star forming galaxies at redshift ~ 0.6 with strong [OII] 3727\AA emission line. If the NB_{596} is located just over this line, and NB_{596} over the continuum, we detected red colors and our broad band is not useful to differentiate between this galaxy and a $z \sim 4$ LBG (see bottom panel of Fig. 4.7).

Summarizing, the contaminants are galaxies with extreme features, which means, very intense absorption/emission lines, together with a very bright continuum in the r_{GUNN} band, or galaxies with a so pronounced Balmer break.

This situation, makes us to conclude that this novel method of narrow band selection for LBGs does not work efficiently, because although this selects LBGs in a very narrow redshift range, this also selects a high fraction of low- z contaminants, due to their sensitivity to features in the spectra.

Although these are not encouraging news, we can still choose color cuts avoiding as much as possible the low- z contaminants. However, that will traduce in a selection of only the most extreme LBGs, and the completeness of the sample will be low. We will

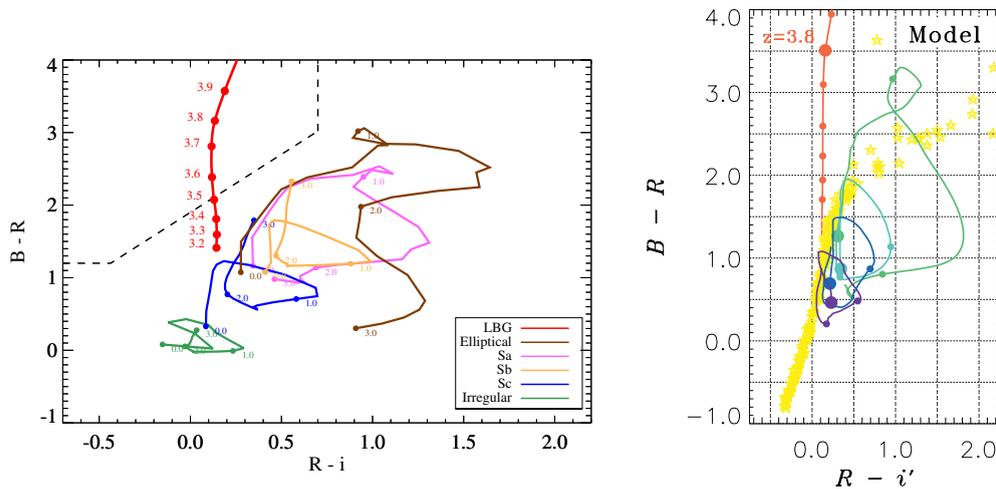


FIGURE 4.6: *Left panel:* Same as Fig. 4.5, but now using the B, R, i filters of the Suprim-cam on the Subaru Telescope. Filled circles on the low- z galaxies evolutionary tracks are indicating the redshifts 0, 1, 2, and 3. The dashed line is showing the color cuts used by Ouchi et al. (2004a) to selects LBGs at $z \sim 4$. *Right panel:* Similar plot but taken directly from Ouchi et al. (2004a). They show as a red line the track of a typical spectrum of star-forming galaxies from $z = 3$ to 4. Filled circles on the red line indicate the redshift from $z = 3.3$ to 3.9 with an interval of $\Delta z = 0.1$. Typical spectra of elliptical, Sbc, Scd, and irregular galaxies are redshifted from $z = 0$ to 3, shown by green, cyan, blue, and violet lines, respectively. Each line is marked by filled circles at $z = 0, 1$, and 2. Yellow stars show 175 Galactic stars.

show in the next Chapter that even in this scenario, we were able to do a clustering measurement using this data.

4.3 Selection Region and LBGs Sample

In general terms, the choice of a suitable color cut to select LBGs has to be done considering mainly three factors: an avoidance of low-redshift interlopers, a maximization of the $z \sim 3.78$ LBGs completeness, and a minimization of line-of-sight galaxy contaminants.

First, we chose a preliminary selection region just based on the Fig. 4.3. We defined two vertical cuts, one to the left of the mean LBG colors at $z = 3.78$ (filled circle in the figure) and one to the right. The first cut is avoiding LBGs located in the upper left region of the diagram, which corresponds to LBGs at $z \sim 3.9$ with strong Lyman alpha emission. The second cut is avoiding LBGs at $z > 3.9$.

A third color cut define a lower limit for $NB_{571} - NB_{596}$, which ensures we are detecting the Lyman break. Additionally, this cut is avoiding LBGs at $z \lesssim 3.7$. Note that we do not defined those three color cuts arbitrarily, but we considered the fraction of LBGs that our selection region is recovering, which is computed from our simulation. In other words, we perform an iterative process computing the completeness of the selection

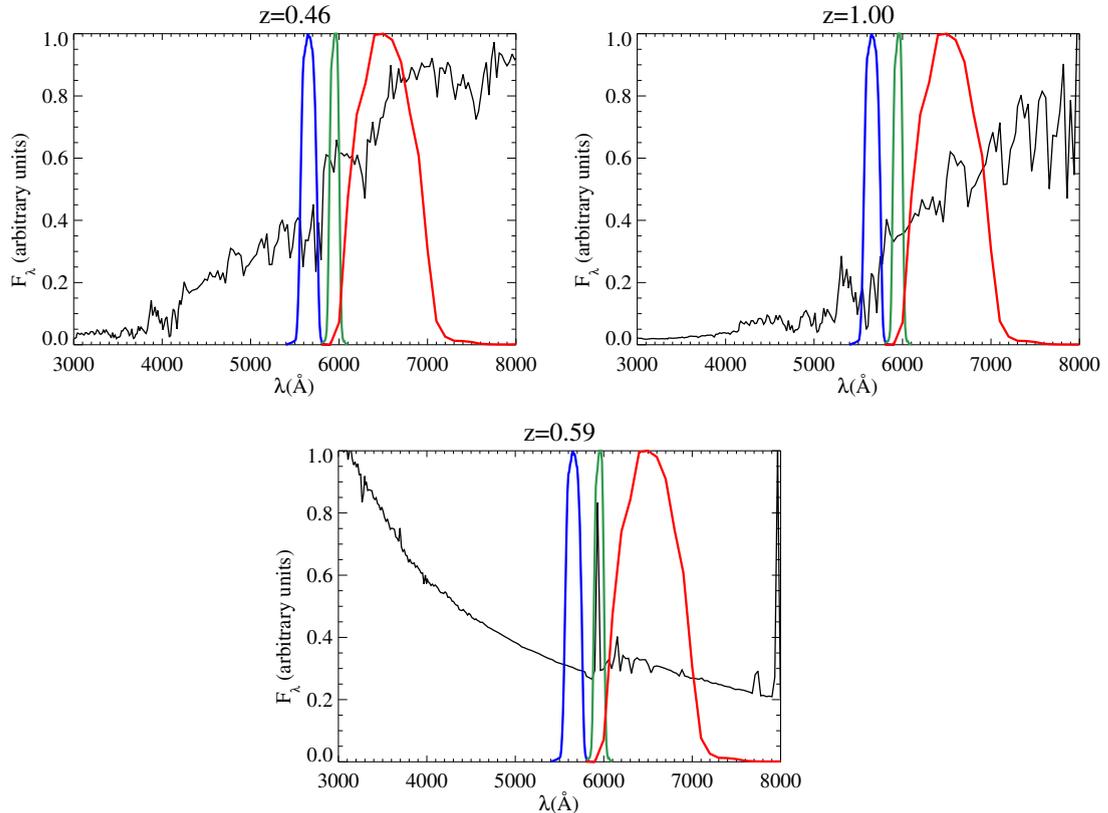


FIGURE 4.7: Examples of some interlopers that could be affecting our LBGs sample. *Top left:* The spectra of a elliptical galaxy at $z = 0.46$, where the Balmer break is located at $\lambda \sim 5840$. *Top right:* The spectra of a galaxy at $z = 1$, where the MgI and MgII absorption are located at $\lambda \sim 5650$. *Bottom:* The spectra of a spiral galaxy at $z = 0.59$, where the OII line is located at $\lambda \sim 5925$.

region for a set of different color cuts (see section 4.4 for the completeness computation details). Then our preliminary selection region is defined by choosing the case in which the LBG line-of-sight contamination is as low as possible while the completeness of LBGs at $z = 3.78$ keep high.

As we discussed in the section 4.2, in our case the low- z contaminants are a large problem, therefore they will have the most incidence in the choice of the color cuts. We reduce the size of this preliminary selection region in order to avoid the region populated by low- z galaxies. The color cut we have to modify for this is the lower limit for $NB_{571} - NB_{596}$, moving it up. If we base just on the LBGs color modeling, we would choose a lower limit for $NB_{571} - NB_{596}$ as an horizontal color cut around ~ 0.8 , but considering the Fig. 4.5 it is natural to modify our lower limit, by using a diagonal color cut, which would avoid the contamination of low- z galaxies in the most efficient way. At the same time we are interested to include, as long as possible, the typical location of LBGs at $z = 3.78$ (marked as a big red circle in the figure 4.5) in order to keep the highest possible LBGs completeness. This defined our new selection region.

Finally a third modification is done, by using the angular clustering measurements of the selected sample. If the sample is highly contaminated, the angular correlation we measure would be consistent with zero, because we would be measuring the correlation of galaxies at very different redshifts, which should be uncorrelated. In the next chapter we explain in detail how we measure the LBGs clustering, however the basic idea I want to explain here is that, if it is highly contaminated, we should be able to detect that in the angular correlation measurements. Based on that idea, we performed an iterative process, where we reduce the size of the selection region doing different color cuts. Then we analyze the angular clustering measurement, in order to ensure this is inconsistent with zero and well described by a power law model. This allowed us to choose a final selection region, which has a low contamination. Note that this do not means that we are building our LBGs sample in order to obtain the desired clustering. For this test we are using the angular correlation function, measured from the data, which is different to the volume average correlation function that we used to measure the real space clustering parameters in our fields. Using this test, we are forcing our data to have an LBG auto-correlation with a slope different to zero, but we are not constraining the amplitude of the correlation. We show this measurement and a deeper explanation in the section 5.4.

Summarizing, the selection regions choice is an iterative process, which takes into account the LBGs color modeling, the low- z galaxies location in the diagram, and the clustering measurement for the selected sample. With this, we are doing everything we could do to avoid a contaminated sample. Note that, even when we ensure that we are not highly contaminated, we are not able to measure exactly how contaminated is our LBGs sample. In other words, we are able to measure the completeness of our sample, but not the purity of this. For a well measured purity of our sample we would have to either perform a color modeling of low- z galaxies, or modeling the influence of contamination on the correlation function (see section 5.4 for details).

The final selection region we choose for selecting $z \sim 3.78$ LBGs is shown in Fig. 4.8 and defined by:

$$NB_{571} - NB_{596} > 1.05 \quad (4.6)$$

$$-0.6 < NB_{596} - r_{\text{GUNN}} < 0.8 \quad (4.7)$$

$$(NB_{571} - NB_{596}) > 0.7(NB_{596} - r_{\text{GUNN}}) + 0.9 \quad (4.8)$$

For LBGs selection, we only consider detected objects with $S/N \geq 4.0$ in both NB_{596} and r_{GUNN} filters. The S/N is defined as the ratio of counts in the $2''$ aperture, given by

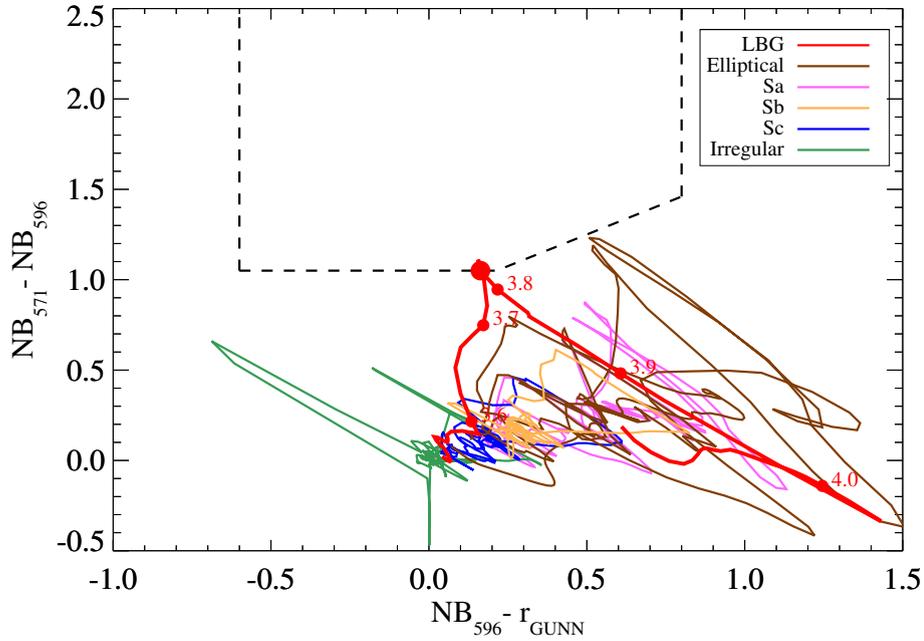


FIGURE 4.8: Same as Fig. 4.5 but we overplotted the selection region as a dashed line.

SExtractor, to the image rms. The image rms is calculated using an IDL procedure which performs a $2''$ aperture photometry in ~ 5000 randomly different positions on the image (avoiding objects position) to compute a robust measurement of the mean for each position. The rms is calculated as the standard deviation of the mean values distribution.

To reduce contamination by false detection, we only consider objects that have FLAGS=0 in SExtractor. That means that we do not consider objects affected by too bright neighbors, blended, saturated, or truncated. We also put a lower limit for the magnitude in order to avoid low-redshift interlopers located in our selection region. For that, we are only selecting object with magnitudes fainter than $r_{\text{GUNN}} = 23.97$, corresponding to LBGs with $L \sim 3.5 L_*$.

We chose this value by computing from the LBGs luminosity function at $z \sim 4$, the luminosity level at which we have lost less than 1% of galaxies. In other words, the 99% of the total number of LBGs have magnitudes between this lower limit and the image limit magnitude $r_{\text{GUNN}} = 25.62$ (limit magnitude at 4σ for a $2''$ diameter aperture) which corresponds to $L = 0.76 L_*$:

$$\frac{\int_{3.5L_*}^{\infty} \phi(L)dL}{\int_{0.76L_*}^{\infty} \phi(L)dL} \sim 0.01 \quad (4.9)$$

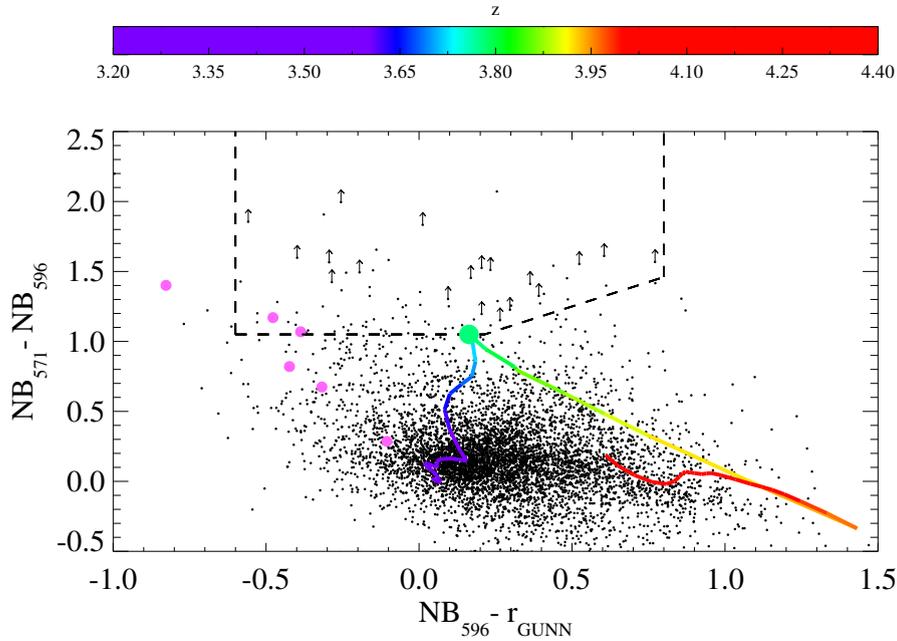


FIGURE 4.9: Color-color diagram for our 6 stacked QSO fields. Here the evolutionary track showed in Fig. 4.8 is plotted as redshift color-coded track according to the color bar. The magenta point indicate the color of the QSO in our filters. Arrows indicate lower limits for $NB_{571} - NB_{596}$ color. These are cases in which the object was not detected in NB_{571} filter at 2σ level and magnitude was replaced by the corresponding limit magnitude.

where $\phi(L)$ is the Schechter function for LBGs at $z \sim 4$, with

$$\phi(L)dL = \phi_* \left(\frac{L}{L_*}\right)^\alpha \exp\left(-\frac{L}{L_*}\right) dL \quad (4.10)$$

Here we have used the Schechter parameters from (Ouchi et al., 2004a): $\phi^* = 2.8 \times 10^{-3} h_{70}^3 \text{Mpc}^{-3}$, $M_{1700}^* = -20.6$ mag and $\alpha = -1.6$. In this way we can safely assume we are excluding only extremely rare bright LBGs.

Given all these selection parameters and the color cuts defined in equations 4.6 to 4.8, we select LBGs in each one of our fields. We show 6 fields stacked color-color plot with all the detected object in Fig. 4.9 where we detected a total of 73 LBGs corresponding to a mean number density of $0.31 \text{ LBGs arcmin}^{-2}$. We also show the individual color-color diagram and the number of LBGs found in each field in Fig. 4.10. Note that the number of LBGs in fields is not directly comparable, because the image area in each field is different (different reduced image size, masked region, etc).

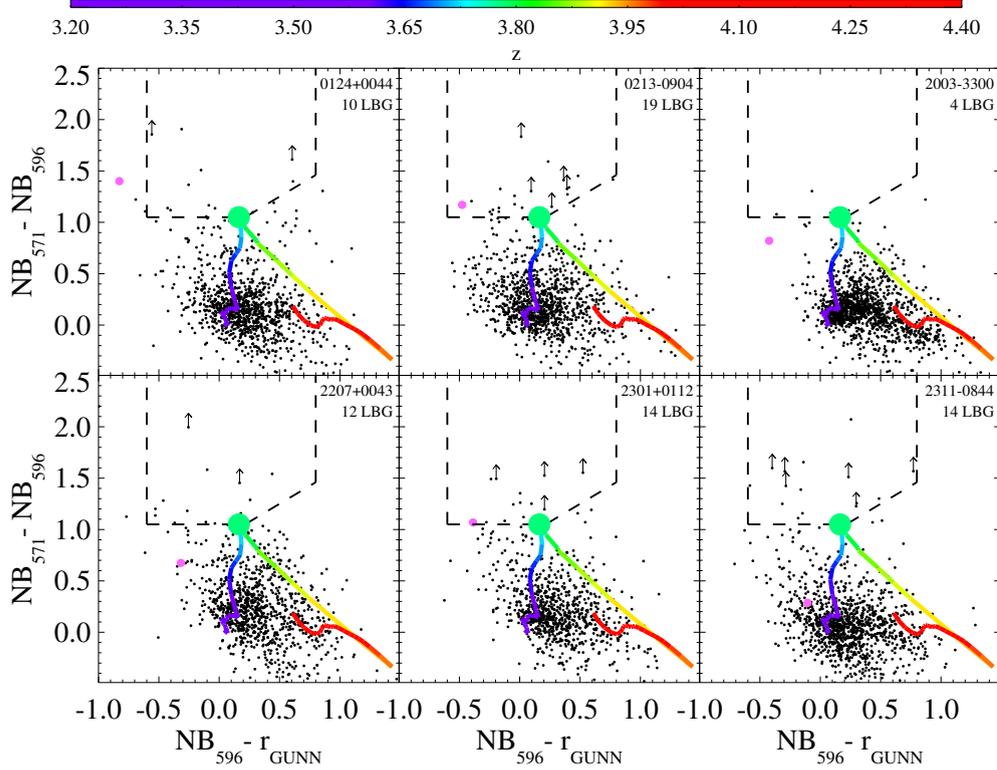


FIGURE 4.10: Color-color diagram for our individual 6 QSO fields. At the top right of each plot the number of LBGs found inside of the selection region is shown.

In Table 4.1 the complete LBGs sample is presented, in Fig. 4.11 we show some examples of dropouts found using this selection criteria, and in Fig. 4.12 we show the distribution of the LBGs for our 6 fields. In appendix A the images of all the LBGs are presented.

4.4 Color Selection Completeness

As I mention in the section 2.4, we need to know the selection function in both angular and redshift space of our sample. The angular selection function can be easily measured by using a mask when we measure the correlation function (see section 5). The redshift space selection function can be computed from our color modeling. We compute the completeness of the color selection region using 1000 simulated spectral models per redshift bin created as we describe in section 4.1. In order to obtain a more realistic completeness function, we added photometric errors in the simulated photometry.

For each spectrum in the simulation, we assigned it an r_{GUNN} magnitude by randomly drawing from the luminosity function integrated in the magnitudes limits of our observations. Here, before of the integration, we have weighted the luminosity function by the completeness in the detection of sources by SExtractor since we are not detecting

TABLE 4.1: LBGs sample. The magnitudes correspond to AB magnitudes measured in a $2''$ diameter aperture for each filter

| ID | RA (J2000) | DEC (J2000) | r_{GUNN} | NB ₅₇₁ | NB ₅₉₆ | ID | RA (J2000) | DEC (J2000) | r_{GUNN} | NB ₅₇₁ | NB ₅₉₆ |
|--------------|---------------|----------------|-------------------|-------------------|-------------------|--------------|---------------|----------------|-------------------|-------------------|-------------------|
| 0124+0044 1 | 21.0628 | 0.6701 | 25.10 | 26.05 | 25.00 | 2207+0043 5 | 331.9200 | 0.7197 | 24.71 | 26.69 | 25.14 |
| 0124+0044 2 | 21.0301 | 0.7776 | 25.13 | 26.50 | 25.41 | 2207+0043 6 | 331.8970 | 0.6925 | 24.03 | 24.70 | 23.50 |
| 0124+0044 3 | 20.9854 | 0.7384 | 24.58 | >26.80 | 25.19 | 2207+0043 7 | 331.9100 | 0.6906 | 25.62 | 26.31 | 25.25 |
| 0124+0044 4 | 21.0644 | 0.7319 | 25.50 | >26.80 | 24.94 | 2207+0043 8 | 331.9300 | 0.6848 | 25.01 | 26.49 | 24.91 |
| 0124+0044 5 | 21.0086 | 0.7312 | 25.24 | 26.35 | 24.98 | 2207+0043 9 | 331.8420 | 0.6847 | 25.59 | 26.51 | 25.33 |
| 0124+0044 6 | 21.0624 | 0.7310 | 25.10 | 26.70 | 24.79 | 2207+0043 10 | 331.8980 | 0.6773 | 24.27 | 25.23 | 24.16 |
| 0124+0044 7 | 21.0159 | 0.7594 | 25.23 | 26.58 | 25.08 | 2207+0043 11 | 331.8850 | 0.6725 | 25.45 | 26.06 | 24.92 |
| 0124+0044 8 | 21.0247 | 0.7534 | 24.95 | 26.06 | 24.92 | 2207+0043 12 | 331.8740 | 0.7401 | 24.71 | 26.05 | 24.89 |
| 0124+0044 9 | 21.0145 | 0.7476 | 24.69 | 26.70 | 25.30 | 2301+0112 1 | 345.3280 | 1.1687 | 25.23 | 26.49 | 25.23 |
| 0124+0044 10 | 21.0110 | 0.7429 | 25.58 | 26.66 | 25.49 | 2301+0112 2 | 345.3140 | 1.1718 | 25.37 | >26.78 | 25.58 |
| 0213-0904 1 | 33.3153 | -9.1322 | 25.09 | >26.94 | 25.10 | 2301+0112 3 | 345.2810 | 1.2176 | 25.05 | >26.78 | 25.25 |
| 0213-0904 2 | 33.2889 | -9.1321 | 25.57 | 26.42 | 25.32 | 2301+0112 4 | 345.3410 | 1.2168 | 25.26 | 26.45 | 25.21 |
| 0213-0904 3 | 33.3341 | -9.1315 | 24.49 | 26.32 | 24.73 | 2301+0112 5 | 345.2880 | 1.2045 | 25.48 | >26.78 | 25.28 |
| 0213-0904 4 | 33.2946 | -9.1311 | 24.50 | 26.43 | 24.98 | 2301+0112 6 | 345.2840 | 1.2028 | 25.43 | 26.22 | 25.15 |
| 0213-0904 5 | 33.3512 | -9.1310 | 25.52 | >26.94 | 25.79 | 2301+0112 7 | 345.3500 | 1.1987 | 25.02 | 26.28 | 25.16 |
| 0213-0904 6 | 33.3429 | -9.1285 | 25.16 | >26.94 | 25.53 | 2301+0112 8 | 345.2830 | 1.1969 | 24.70 | >26.78 | 25.22 |
| 0213-0904 7 | 33.3763 | -9.1275 | 24.60 | 25.41 | 24.20 | 2301+0112 9 | 345.2650 | 1.1874 | 24.72 | 25.87 | 24.75 |
| 0213-0904 8 | 33.2968 | -9.1281 | 25.54 | >26.94 | 25.63 | 2301+0112 10 | 345.2720 | 1.1806 | 25.23 | 26.19 | 25.11 |
| 0213-0904 9 | 33.3222 | -9.1273 | 24.82 | 25.98 | 24.84 | 2301+0112 11 | 345.2870 | 1.1775 | 25.16 | 26.37 | 25.27 |
| 0213-0904 10 | 33.3018 | -9.1276 | 25.20 | 26.33 | 25.16 | 2301+0112 12 | 345.3030 | 1.1735 | 24.49 | 25.74 | 24.25 |
| 0213-0904 11 | 33.2836 | -9.1243 | 25.22 | >26.94 | 25.61 | 2301+0112 13 | 345.3350 | 1.2363 | 24.91 | 26.26 | 25.01 |
| 0213-0904 12 | 33.2767 | -9.1210 | 25.14 | 25.99 | 24.91 | 2301+0112 14 | 345.2770 | 1.2367 | 24.16 | 25.51 | 24.16 |
| 0213-0904 13 | 33.2960 | -9.0492 | 25.39 | 26.32 | 25.22 | 2311-0844 1 | 347.8960 | -8.7096 | 25.69 | 26.63 | 25.38 |
| 0213-0904 14 | 33.3650 | -9.0724 | 24.51 | 25.41 | 24.29 | 2311-0844 2 | 347.9240 | -8.7231 | 24.45 | >26.79 | 25.22 |
| 0213-0904 15 | 33.3310 | -9.0741 | 25.12 | 26.79 | 25.52 | 2311-0844 3 | 347.9170 | -8.7249 | 25.65 | >26.79 | 25.36 |
| 0213-0904 16 | 33.3212 | -9.0772 | 25.33 | 26.30 | 25.16 | 2311-0844 4 | 347.9030 | -8.7347 | 25.05 | >26.79 | 25.28 |
| 0213-0904 17 | 33.2955 | -9.1070 | 25.39 | 26.61 | 25.35 | 2311-0844 5 | 347.9330 | -8.7388 | 24.72 | 26.24 | 24.59 |
| 0213-0904 18 | 33.3111 | -9.0545 | 25.35 | 26.51 | 25.19 | 2311-0844 6 | 347.9220 | -8.7419 | 25.52 | >26.79 | 25.22 |
| 0213-0904 19 | 33.3831 | -9.0543 | 24.88 | 26.62 | 25.28 | 2311-0844 7 | 347.9430 | -8.7429 | 25.40 | 26.33 | 25.22 |
| 2003-33 1 | 300.8650 | -32.8778 | 24.18 | 25.53 | 24.42 | 2311-0844 8 | 347.9430 | -8.7424 | 25.27 | >26.79 | 25.56 |
| 2003-33 2 | 300.8920 | -32.9097 | 24.93 | 25.65 | 24.57 | 2311-0844 9 | 347.9460 | -8.7546 | 24.71 | 26.50 | 25.20 |
| 2003-33 3 | 300.8750 | -32.9159 | 24.63 | 25.80 | 24.71 | 2311-0844 10 | 347.9310 | -8.7650 | 25.59 | >26.79 | 25.19 |
| 2003-33 4 | 300.8540 | -32.8583 | 25.06 | 26.29 | 24.99 | 2311-0844 11 | 347.8910 | -8.7236 | 25.58 | 26.60 | 25.44 |
| 2207+0043 1 | 331.8270 | 0.6683 | 25.93 | 26.75 | 25.47 | 2311-0844 12 | 347.9150 | -8.7248 | 24.32 | 26.65 | 24.58 |
| 2207+0043 2 | 331.9070 | 0.6693 | 25.16 | >26.78 | 25.33 | 2311-0844 13 | 347.8990 | -8.7257 | 24.30 | 25.86 | 24.60 |
| 2207+0043 3 | 331.9320 | 0.7323 | 25.24 | 26.26 | 25.20 | 2311-0844 14 | 347.9180 | -8.7342 | 25.37 | 26.60 | 25.21 |
| 2207+0043 4 | 331.9030 | 0.7306 | 25.04 | >26.78 | 24.79 | | | | | | |

the 100% of the sources at the fainter magnitudes. We included this, so the simulated r_{GUNN} distribution is a good reproduction of the real data. Then, we re-scaled the flux of our model spectra and compute the magnitudes in the other two filters NB₅₇₁ and NB₅₉₆. This defined our model photometry.

From the LBGs data sample we computed a median relation between magnitudes and their signal to noise (S/N) ratio for each filter. We use this median relation to assign a S/N to the modeled magnitudes. Then we added noise to the photometry and this defines the final color of the simulated spectra.

In this way, we computed the color of each model spectrum and for each redshift bin. Then we computed the completeness function by calculating the fraction of the simulated spectra that fulfilled our color cut criteria and limit magnitudes for NB₅₉₆ and r_{GUNN} . This completeness define the redshift selection function of the sample. The error inclusion obliges to enlarge the selection region in order to recover also the fainter LBGs, which have greater photometric errors (this was considered for the choice of our

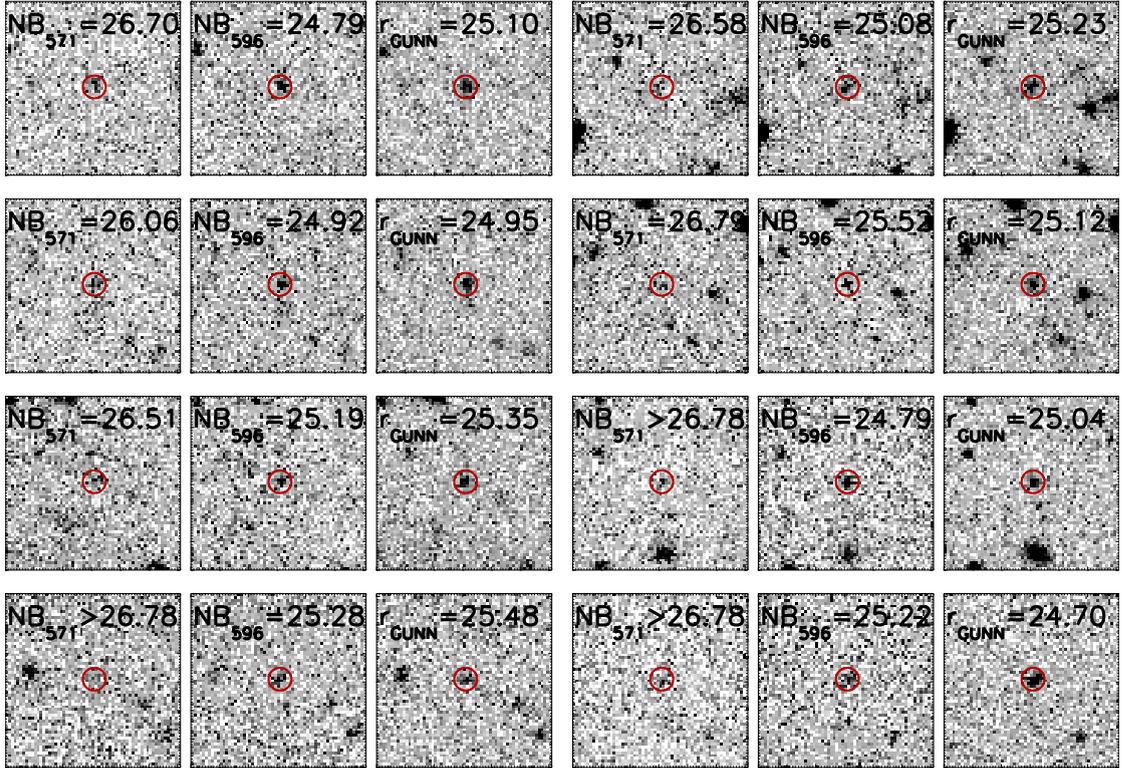


FIGURE 4.11: Images of some selected LBGs. From left to right we show the NB_{571} , NB_{596} and r_{GUNN} images. Each panel is $15''$ on a side. The red circle show the position of the detected object, and the size correspond to the region in which the photometry was done ($2''$ in diameter). The magnitudes are indicated in each panel.

selection region in the section 4.3). We show the colors of the simulated spectra, including photometric errors in Fig. 4.13, where it is possible to observe the enlargement of the location of LBGs in comparison with Fig. 4.3, where photometric errors were not included.

Note that as the limit magnitude of our fields are different, the r_{GUNN} distribution for each one is also different. We computed the completeness for each individual field, which varied from each other by small quantities. The median color selection completeness of our fields is shown in the Fig. 4.14.

Despite we had to go away from the main bulk of LBGs in the color-color diagram, we are recovering $\sim 35\%$ of LBGs at $z \sim 3.78$ in a $\Delta z \sim 0.04$. Our window selects LBGs from a wider redshift, ranging from $z \sim 3.60$ to $z \sim 3.95$ but with much lower rate. The narrow band technique effectively selects LBGs in much narrower redshift range, in comparison with broad band, but unfortunately, as contaminants have similar colors of the LBGs in our filters, we could not get a high completeness in the sample. This completeness measurement defines the redshift selection function that is used for the clustering measurements in the next chapter.

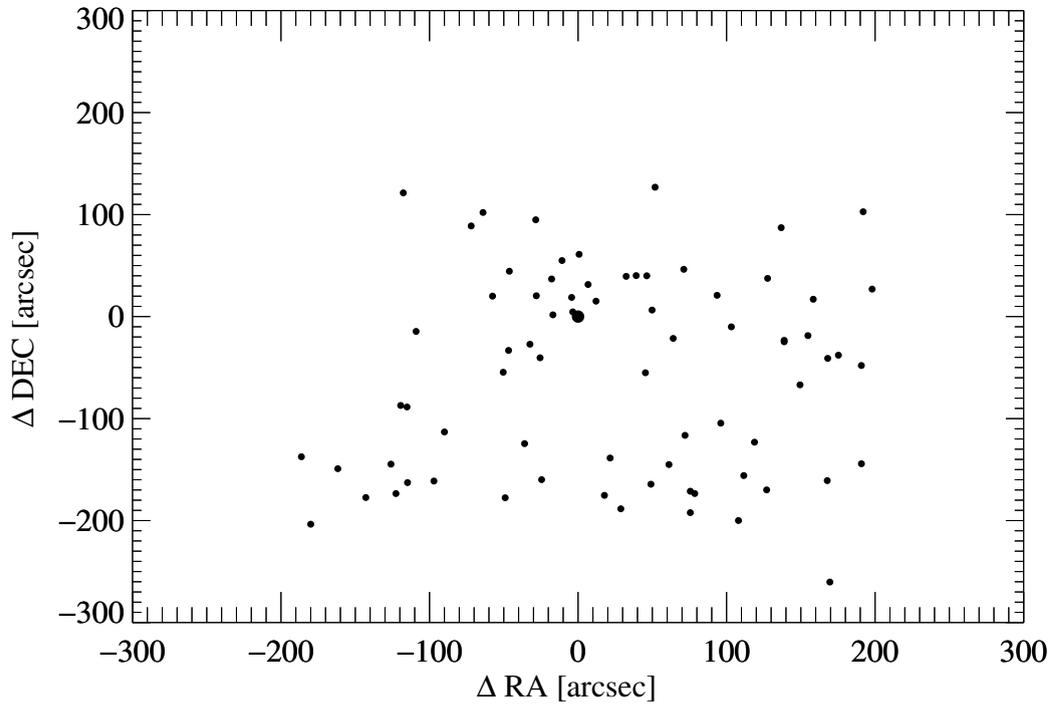


FIGURE 4.12: The distribution of LBGs around the quasar in the plane of the sky for the stacked 6 fields. The central QSOs is located in 0.0 and is plotted by a large black circle.

We calculate an effective coverage in redshift as $\Delta z = \sum_i \delta z C(z_i)$, with $C(z_i)$ the fraction of LBGs recovered by the selection region at each redshift (the completeness) and δz the bin size in the completeness computation ($\delta z = 0.02$). We obtained a mean $\Delta z = 0.076$, which corresponds to $\sim 4.760 \text{ km s}^{-1}$ at $z = 3.78$.

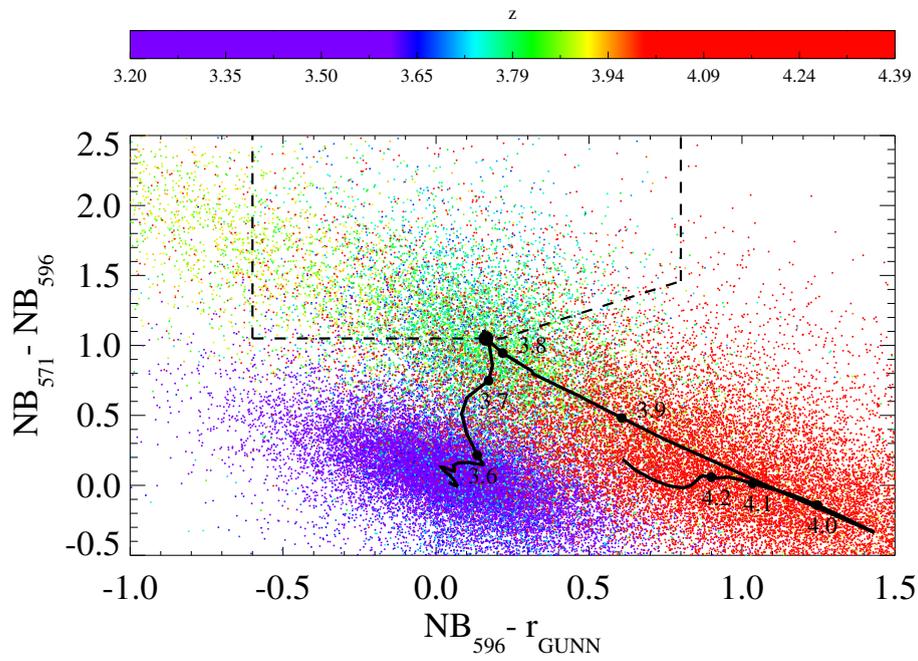


FIGURE 4.13: Color-color diagram showing the simulated colors for 1000 LBGs spectra including photometric errors, plotted as redshift color-coded points according to the color bar. The mean LBG evolutionary track is plotted as a black curve. The filled points indicate the redshift from 3.6 to 4.2. The bigger circle shows the exact position of the mean $z = 3.78$ LBG colors. We overlotted the selection region as a dashed line.

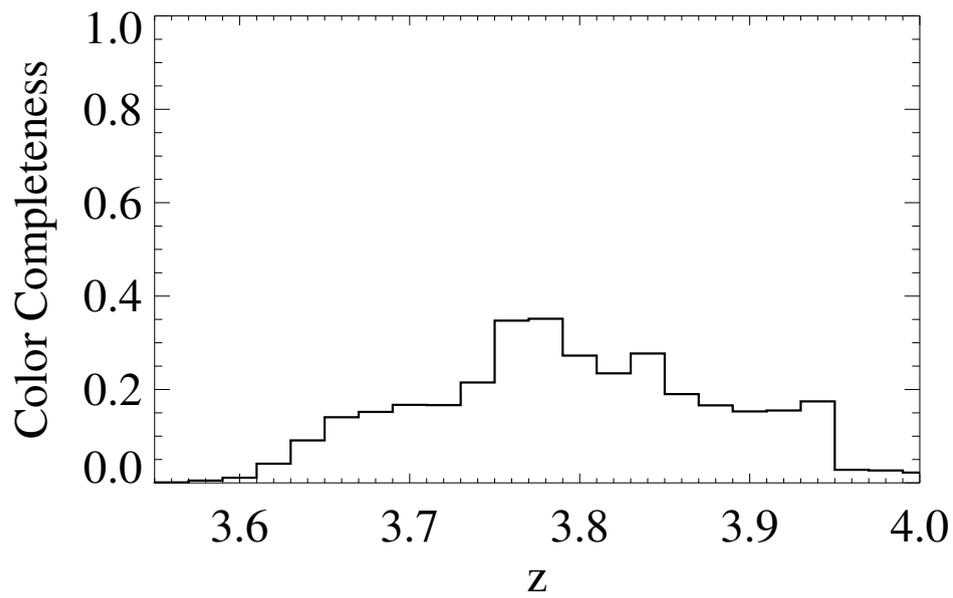


FIGURE 4.14: Completeness of the LBGs selection. The completeness was determined from 1000 simulated LBGs spectra with different EW and, continuum power law and magnitudes. This is calculated by computing the fraction of the simulated spectra, per redshift bin, which was selected by the selection region. This define the redshift selection function that is used for the clustering measurements in the next chapter.

Chapter 5

Clustering Analysis

In this chapter I analyze the clustering properties of LBGs around QSOs at $z \sim 4$. The auto-correlation of both LBGs and QSOs at $z \sim 4$ is well studied, so in a first step, I could be able to estimate the theoretical expectation for the QSO-LBG cross-correlation function from the individual auto-correlation functions which I present in section 5.1. This theoretical expectation is used to compare with our observational cross-correlation measurement in section 5.2 where I also estimate the real-space clustering parameters (r_0, γ) . As a complementary result, I present the LBGs auto-correlation function in QSO fields in section 5.3. Finally, in section 5.4, I describe a way to determine if the sample is contaminated based on the clustering measurements. I also present some arguments supporting the fact that our LBGs sample is indeed not highly contaminated.

5.1 Expected Clustering Signal

Our data does not include redshift measurements, and because of this, in principle we only have the 2D information for the position of LBGs in the sky. However, as we showed in section 4.4, our narrow band method ensures we are selecting LBGs in a very thin redshift range, which allows us to measure projected distances in physical units instead to angular units, as it is done in most studies that do not have redshift measurements of individual galaxies. This allows us to measure a volume averaged cross-correlation function as it was defined in section 2.3.2.

In order to estimate the theoretical expectation of the QSO-LBG cross-correlation function, we used the estimator presented in the equation (2.9) to estimate the volume averaged cross-correlation function in logarithmic spaced radial bins as:

$$\chi = \frac{\langle QG \rangle}{\langle QR \rangle} - 1 \quad (5.1)$$

where $\langle QG \rangle$ and $\langle QR \rangle$ are the number of QSO-LBG and QSO-random pairs in a radial bin. Note that we are not using the normalization by the number density as in the equation (2.9), because as the QSOs subject of this study are located on an overdense region in the Universe, this normalization is not valid in our case (see the section 2.4).

We can use the equation (2.5) to compute the number of LBGs at r distance from a QSO, in a volume element dV by:

$$\langle QG \rangle = \int_{\text{Vol}} n_G(z, < m_{\text{lim}}) C(z) [1 + \xi_{QG}(r)] dV \quad (5.2)$$

where $n_G(z, < m_{\text{lim}}) C(z)$ is the mean number density of LBGs at redshift z under the limit magnitude m_{lim} , weighted by $C(z)$, which refers to the completeness function estimated in section 4.4, and define the redshift selection function of our survey. $\xi_{QG}(r)$ is the QSO-LBG cross-correlation function. Here, r refers to the real space separation, but as in this study we do not have redshift measurements, we were able to measure only the projected distances R between galaxies, where $r = \sqrt{R^2 + Z^2}$, assuming they are roughly at the same redshift. Here Z refers to radial comoving distance:

$$dZ = \frac{c}{H(z)} dz = \frac{c}{H(z)} \frac{dv(1+z)}{c} = \frac{dv}{aH(z)} \quad (5.3)$$

where dv is the covered velocity range, $a = 1/(1+z)$ is the scale factor of the Universe, and $H(z)$ the Hubble constant at redshift z , which we assume does not varies in our covered redshift range.

We computed the expected $\langle QG \rangle$ value from equation (5.2) in a cylindrical volume with radius element dR and height dZ , mapped by our filter configuration (calculated in section 4.4), then $dV = 2\pi R dR dZ$. In these coordinates, the equation (5.2) becomes:

$$\langle QG \rangle = n_G(z, < m_{\text{lim}}) \int_{Z_{\text{min}}}^{Z_{\text{max}}} \int_{R_{\text{min}}}^{R_{\text{max}}} C(Z) [1 + \xi_{QG}(R, Z)] 2\pi R dR dZ \quad (5.4)$$

where we assumed that $n_G(z, < m_{\text{lim}})$ is constant in the considered redshift range. We define an effective volume as:

$$\begin{aligned}
V_{\text{eff}} &= \int_{Z_{\text{min}}}^{Z_{\text{max}}} \int_{R_{\text{min}}}^{R_{\text{max}}} C(Z) 2\pi R dR dZ \\
&= \pi (R_{\text{max}}^2 - R_{\text{min}}^2) \int_{Z_{\text{min}}}^{Z_{\text{max}}} C(Z) dZ
\end{aligned} \tag{5.5}$$

Using this notation, the equation (5.4) can be written as:

$$\langle QG \rangle = n_G(z, < m_{\text{lim}}) V_{\text{eff}} \left(1 + \frac{\int_{Z_{\text{min}}}^{Z_{\text{max}}} \int_{R_{\text{min}}}^{R_{\text{max}}} C(Z) \xi_{QG}(R, Z) 2\pi R dR dZ}{V_{\text{eff}}} \right) \tag{5.6}$$

For the V_{eff} calculation, we have truncated our completeness function at values in which the completeness is negligible, in order to avoid an increment of the noise in our estimation. We chose $z_{\text{min}} = 3.61$ and $z_{\text{max}} = 3.95$ corresponding to a velocity range of $\Delta v \sim 21,300 \text{ km s}^{-1}$ and $\Delta Z \sim 188.6 \text{ cMpc}$. Choosing different R_{min} and R_{max} values allows to compute the cross-correlation in different radial bins, and the maximum R_{max} value will be limited by the images size.

The computation of the expected number of LBGs in QSO environments, $\langle QG \rangle$, requires the knowledge of the mean number density of LBGs $n_G(z, < m_{\text{lim}})$, which can be calculated from their luminosity function. We used the Schechter parameters from Ouchi et al. (2004a) who studied the photometric properties based on a large sample of ~ 2200 LBGs at $z \sim 4$. The values used are $\phi^* = 2.8 \times 10^{-3} \text{ h}_{70}^3 \text{ Mpc}^{-3}$, $M_{1700}^* = -20.6$ mag, and $\alpha = -1.6$. We integrated the luminosity function in the limits given by our LBGs selection, corresponding to a luminosity range of $0.76 \lesssim L/L_* < 3.5$, and we obtained, $n_G = 3.3 \times 10^{-3} \text{ h}^3 \text{ cMpc}^{-3}$. (Note that each field has a slightly different limit magnitude, then we computed n_G for each individual field).

Finally, we assumed that the LBG-QSO cross-correlation function follows a power law form:

$$\xi_{QG}(R, Z) = \left(\frac{\sqrt{R^2 + Z^2}}{r_0^{QG}} \right)^{-\gamma} \tag{5.7}$$

The cross-correlation length r_0^{QG} can be estimated using the individual auto-correlation lengths of both QSO and LBGs (e.g. Coil et al., 2009). If we assume that both LBGs and QSOs trace the same underlying dark matter, and a linear bias such that $\delta_G = b_G \delta_{DM}$, and $\delta_Q = b_Q \delta_{DM}$ then we can write the galaxy auto-correlation function as in the equation (2.4):

$$\xi_{GG}(r) = \langle \delta_G(r) \delta_G(r) \rangle = b_G^2 \langle \delta_{DM} \delta_{DM} \rangle = b_G^2 \xi_{DM} \quad (5.8)$$

and similarly for QSOs:

$$\xi_{QQ}(r) = \langle \delta_Q(r) \delta_Q(r) \rangle = b_Q^2 \xi_{DM} \quad (5.9)$$

Then the cross-correlation function is given by:

$$\begin{aligned} \xi_{QG}(r) &= \langle \delta_Q(r) \delta_G(r) \rangle \\ &= b_Q b_G \xi_{DM} \\ &= \sqrt{\frac{\xi_{QQ}}{\xi_{DM}}} \sqrt{\frac{\xi_{GG}}{\xi_{DM}}} \xi_{DM} \\ &= \sqrt{\xi_{QQ} \xi_{GG}} \end{aligned} \quad (5.10)$$

Assuming a power law for both ξ_{QQ} and ξ_{GG} with the same slope γ , then we have:

$$r_0^{QG} = \sqrt{r_0^{QQ} r_0^{GG}} \quad (5.11)$$

Then, having the auto-correlation information of both quasars and galaxies, we can compute, assuming a linear bias, an estimation for the cross-correlation between them.

We used the auto-correlation lengths values $r_0^{GG} = 4.1 \text{ h}^{-1} \text{ cMpc}$ for LBGs at $z \sim 4$ (Ouchi et al., 2004b), and $r_0^{QQ} = 22.3 \text{ h}^{-1} \text{ cMpc}$ for QSOs at $z \sim 4$. This last value was calculated using the correlation measurements from (Shen et al., 2007) for QSO in the redshift range $z > 3.5$ with a fix $\gamma = 1.8$. The resulting expected r_0^{QG} value is then $r_0^{QG} = 9.6 \text{ h}^{-1} \text{ cMpc}$ for $\gamma = 1.8$.

In the particular case in which LBGs are randomly distributed around QSOs, $\xi_{QG}(r) = 0$, and the QSO-random number pairs at R distance from a QSO, in a volume V_{eff} , is given by:

$$\langle QR \rangle = n_G(z, < m_{\text{lim}}) V_{\text{eff}} \quad (5.12)$$

If we replace equation (5.6) and (5.12) in (5.1), we can write χ as:

$$\chi = \frac{\int_{Z_{\min}}^{Z_{\max}} \int_{R_{\min}}^{R_{\max}} C(Z) \xi_{QG}(R, Z) 2\pi R dR dZ}{V_{\text{eff}}} \quad (5.13)$$

which is exactly the definition we presented in section 2.3.2: a dimensionless estimator which corresponds to a volume averaged correlation function, integrated in both redshift and radial bin space.

We computed the theoretical expectation χ^{expected} for different radial bins, and for each field. Then, we stacked this measurement for our six fields. This result is shown as a dashed line in Fig. 5.1 together with the observational results, as it is explained in the next section.

In order to have a first approach of the LBGs overdensity in QSO fields, we calculated the expected number of LBGs on our whole fields if they were in random locations, so that we could compare with what we find in section 4.3. We computed this for the same magnitude range covered by the LBGs sample, and we obtained $\langle QR \rangle \sim 37$ for the six stacked fields, which is ~ 2 times lower than the number of LBGs we detected in our fields (73 LBGs), which is a first indication that our fields show a LBGs overdensity compared with what we expected from random fields. The number density of the sample is $8.71 \times 10^{-5} \text{ arcsec}^{-2}$. In the next section, we measure how they are clustered around the central QSO.

5.2 QSO-LBG Cross-Correlation Measurement

We computed the observational χ value using equation 5.1, where $\langle QG \rangle$ is the QSO-LBG pairs in a radial bin which is directly measured on our masked images, and then stacked it for our six fields.

As we mention in section 2.4, in our case we could not measure the number density of LBGs from the data itself, as it is typically done for clustering studies. Because of this $\langle QR \rangle$ could not be measured from our images, but we could use the theoretical computation of the number density of LBGs in random fields.

The estimation of $\langle QR \rangle$ was done by using the equation (5.12), which requires a precise knowledge of the redshift selection function, encapsulated in the term $C(z)$. For this reason we computed this so carefully in section 4.1. We computed $\langle QR \rangle$ for each individual field and we chose an arbitrary re-scaling factor F such that the number of LBGs per field becomes very large ($\sim 10,000$). Then, we created those galaxies over our masked images in order to consider the radial selection function. Finally, we measured

TABLE 5.1: QSO-LBG Cross-Correlation Function.

| R(h^{-1} cMpc) | $\langle QG \rangle$ | $\langle QR \rangle$ | χ |
|-------------------|----------------------|----------------------|------------------------------|
| 0.188 | 1 | 0.049 | $19.513^{+47.179}_{-16.964}$ |
| 0.383 | 3 | 0.217 | $12.833^{+13.455}_{-7.530}$ |
| 0.777 | 3 | 0.983 | $2.051^{+2.968}_{-1.661}$ |
| 1.576 | 15 | 4.092 | $2.666^{+1.212}_{-0.936}$ |
| 3.200 | 26 | 16.330 | $0.592^{+0.377}_{-0.310}$ |
| 6.497 | 25 | 15.127 | $0.653^{+0.401}_{-0.329}$ |

$\langle QR \rangle$ in the each radial bin and we re-scaled $\langle QR \rangle$ down by F factor to be in the correct units.

We repeated this process for our six fields and then we stacked the $\langle QR \rangle$ counts and measured the χ value in radial bins as in equation 5.1. The result of this measurement is shown in Fig. 5.1 (filled circles) together with the best fit of our measurement, which is described in the next section. The tabulated values of $\langle QG \rangle$, $\langle QR \rangle$ and χ are showed in Table 5.1. The statistical errors are estimated using:

$$\Delta\chi_i = \frac{(\Delta\langle QG \rangle)_i}{\langle QR \rangle_i} \quad (5.14)$$

where $(\Delta\langle QG \rangle)_i$ is the error in the LBG-QSO pair counts for the i th radial bin. This is estimated using one-side Poisson confidence intervals for small number statistics (Gehrels, 1986). However, note that this Poisson error bars are actually a underestimation of the real error bars. As galaxies are not normally distributed in the Universe, but they have a non zero clustering length, their position are correlated to each other. This implies that the measurement of $\langle QG \rangle$ in each bin is not independent, and in this case, the Poisson estimator for the error bars breaks down.

An used alternative is to estimate the errors bars using the bootstrap technique, however, if we assumed that the correlation between bins is not so strong then we can use error bars from Poisson.

5.2.1 Fitting r_0 Parameter

Given the projected cross-correlation function measurement, we are now interested in determining the real-space cross-correlation parameters r_0^{QG} and γ that best fit our data.

For this fitting, we used a maximum likelihood estimator (MLE), which is designed to estimate the models parameters given the data. Basically, this method searches for the

parameters which maximize the probability to produce the data we in fact observe. By definition, the likelihood function is given by:

$$\mathcal{L}(\vec{\theta}|x_1\dots x_N) = \prod_{i=1}^N f(x_i|\vec{\theta}) \quad (5.15)$$

where x_i is the data measured in the i th bin and $f(x_i|\vec{\theta})$ is the probability distribution function (PDF) of the data, given the vector of parameters $\vec{\theta} = (r_0, \gamma)$.

In our case, we are counting LBGs in the space given by the $\langle QG \rangle$ value. Since this is a counting process with small number counts (see Table 5.1), this is well modeled by Poissonian statistics. Then we assumed a Poisson distribution for the counts, and therefore a Poisson PDF for the equation (5.15):

$$\mathcal{L} = \prod_{i=1}^N \frac{e^{-\lambda_i} \lambda_i^{x_i}}{x_i!} \quad (5.16)$$

where the product goes for each radial bin, with N being the number of bins, and λ_i is the expected number counts value for the i th bin. If we take the natural logarithm at both sides of the equation (5.16), we obtain:

$$\ln \mathcal{L} \propto - \sum_{i=1}^N \lambda_i + \sum_{i=1}^N x_i \ln(\lambda_i) \quad (5.17)$$

The proportionality is because we have dropped the terms which are model independent. In our case x_i is the observed $\langle QG \rangle_i$ value (data points in Fig. 5.1), and then λ_i is the expected $\langle QG \rangle_i^{expected}$ value, given by the equation (5.6) for each radial bin and for a given (r_0, γ) model. Then we re-write the equation (5.17) as:

$$\ln \mathcal{L} \propto \sum_{i=1}^N \left[\langle QG \rangle_i \ln \left(\langle QG \rangle_i^{expected} \right) - \langle QG \rangle_i^{expected} \right] \quad (5.18)$$

We calculated the log-likelihood for a grid of (r_0, γ) values ranging from $0.0 \leq \gamma \leq 3.0$ and $1.0 \leq r_0 \leq 100.0$ and we obtained the most probable (r_0, γ) model, given our measurement by a maximization of the log-likelihood function.

The maximum likelihood values we obtain are $r_0 = 9.91_{-1.79}^{+3.28} h^{-1}$ cMpc and $\gamma = 2.05_{-0.46}^{+0.20}$ which are plotted in the top panel of Fig. 5.1. We also computed the 1 and 2 σ 2D confidence regions for these parameters which are shown in a $r_0 - \gamma$ plane in Fig. 5.2

in blue and red color, respectively. The signal to noise of our data is enough to fit both parameters, but we also calculated the maximum likelihood estimate for a fixed $\gamma = 1.8$ in order to compare with the expected r_0 estimation calculated in section 5.1. The resultant maximum likelihood estimator is $r_0 = 11.56_{-1.41}^{+15.31} \text{ h}^{-1} \text{ cMpc}$. We plotted this fit in the bottom panel of Fig. 5.1 as a red solid line.

Note that as the bins are correlated with each other, the correlation function will be sensitive to the choice of bins. In principle, we should consider this correlation including the covariance matrix in our likelihood function. However, we assumed that our measurements are uncorrelated, where the covariance matrix is a diagonal matrix, and the likelihood function reduces to equation 5.16. This assumption is not correct, which will traduce itself in an underestimation of the errors in our fitted parameters, but we chose this because doing a computation of a covariance matrix in cases where the data set is as small as here, is very difficult (e.g Hennawi et al., 2015).

Our QSO-LBGs cross-correlation measurements show an agreement with the expected correlation computed in the section 5.1, where we obtained $r_0^{expected} = 9.6 \text{ h}^{-1} \text{ cMpc}$ for $\gamma = 1.8$. Our results suggest that QSOs would be tracing massive dark matter halos in the early Universe, where an enhanced of LBGs is detected.

From our cross-correlation measurement, we can also draw some conclusions related to the AGN feedback. Extra physics related to the presence of the AGN, could suppress the expected enhancement of galaxies around $z \sim 4$ QSOs predicted by hierarchical clustering. Some people have suggested that feedback from an AGN can regulate the growth of galaxies throughout the history of the Universe (Hopkins et al., 2006; Sijacki et al., 2007; Springel et al., 2005a), and Fanidakis et al. (2013) has recently shown how these effects manifest in the clustering of AGN. If feedback from the AGN suppresses galaxy formation in its $\lesssim \text{Mpc}$ scale environment, then one expects a reduction in the number of galaxies around QSOs. Recently several claims have been made along these lines from studies of QSO environments. For example Kashikawa et al. (2007) and Utsumi et al. (2010) studied 2 QSO fields at $z = 5$ and $z = 6.4$, respectively, and both of them detect overdensity of galaxies, but they report no detection of galaxies at $R < 2 - 4.5 \text{ Mpc}$ from the central QSO. However, these highly intriguing results are inconclusive given the very poor statistics.

In our case, we do report LBGs detected at $R < 2 \text{ Mpc}$ from the QSO, as can be seen in the stacked LBGs distribution of Fig. 4.12, and indeed, our cross-correlation measurement do not show a decrease in the number density of LBGs in the inner bins. So, we roughly conclude that QSO feedback would seem not affect the galaxy number counts in their inner environment, but a more detailed study is absolutely needed to draw stronger conclusions.

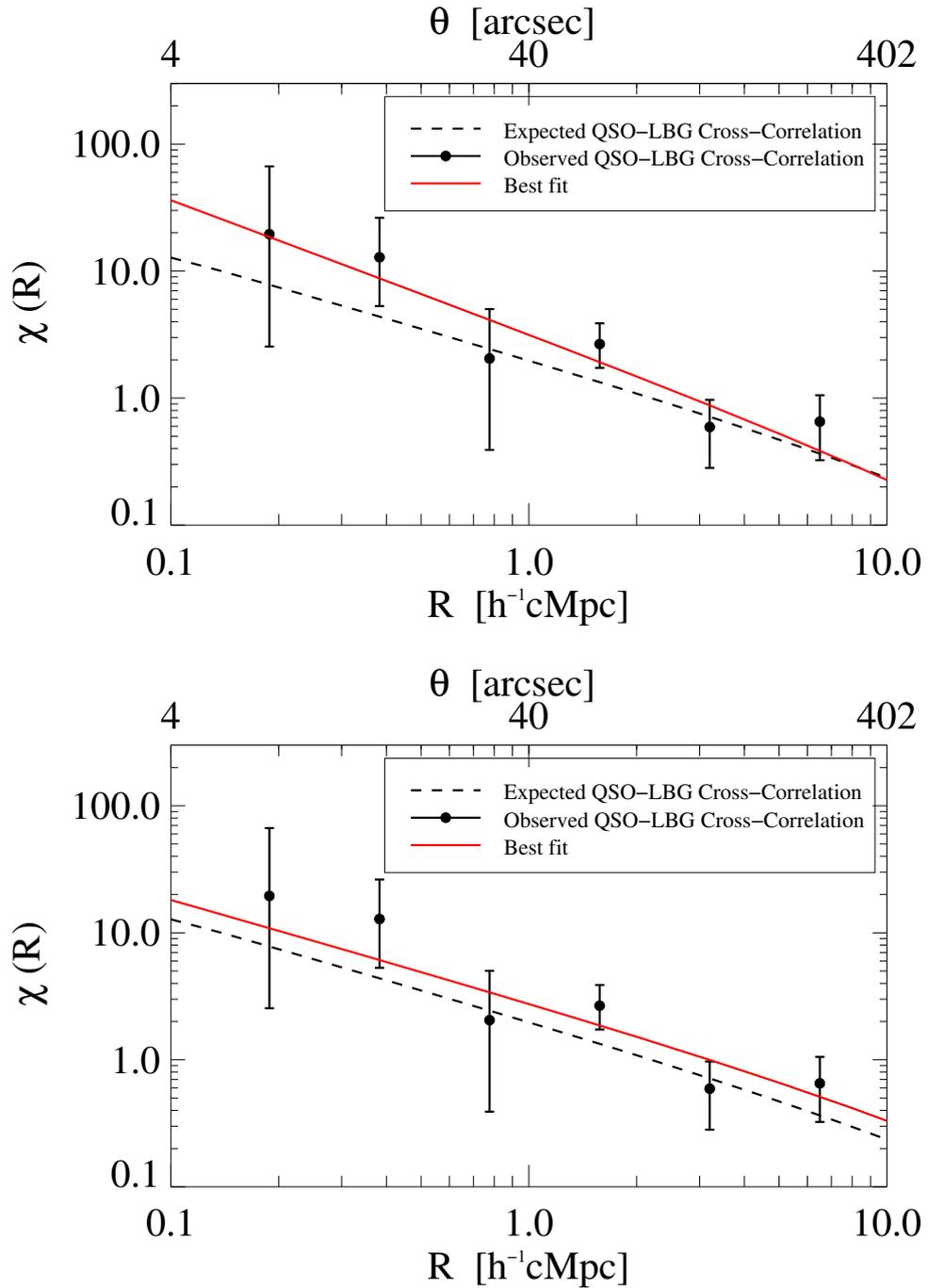


FIGURE 5.1: QSO-LBG cross-correlation function and the Maximum likelihood estimate. The filled circles are showing our measurement described in section 5.2 with 1σ Poisson error bars. The dashed black line shows the theoretical expectation of χ for our six stacked fields calculated from the QSO and LBGs auto-correlation functions. *Top:* The solid red curve shows the best MLE for both r_0 and γ as free parameters. We obtain $r_0 = 9.91^{+3.28}_{-1.79} h^{-1} \text{cMpc}$ and $\gamma = 2.05^{+0.20}_{-0.46}$. *Bottom:* The solid red curve shows the best MLE for r_0 with a fixed $\gamma = 1.8$. Our results are consistent, with the expectation value computed in the section 5.1.

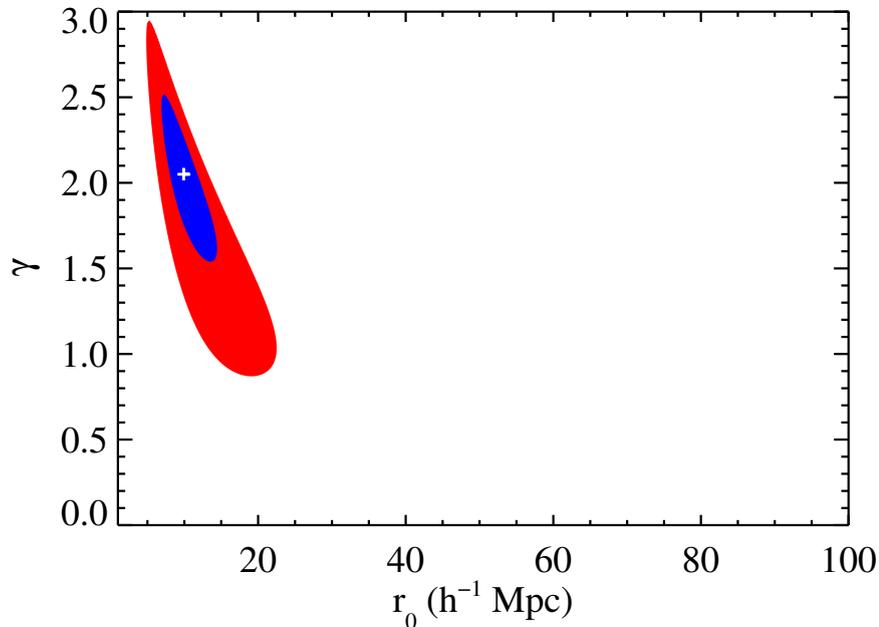


FIGURE 5.2: 1σ and 2σ confidence regions of r_0 and γ parameters (in blue and red respectively), determined using a maximum likelihood estimator. The best estimation is shown as a white cross. This result indicates that we can not perform a good estimation of those parameters simultaneously, which is in part due to the large error bars in our measurements. We then preferred to fit this data with a fixed γ value.

5.3 LBG Auto-Correlation Measurement

Another way to study the clustering of LBGs in QSO environments is measuring the LBGs auto-correlation function in our fields. If QSOs trace biased locations in the Universe, then we expect that LBGs around them be more clustered than LBGs in random fields. This implies that we should measure a slightly enhanced auto-correlation function in our fields. The auto-correlation function of $z \sim 4$ LBGs in random fields was measured by Ouchi et al. (2004b), then we used their results to directly compare them with ours.

To measure the LBGs auto-correlation function in our images, we used the estimator given in the equation (2.8):

$$\chi = \frac{\langle GG \rangle}{\langle RR \rangle} - 1 \quad (5.19)$$

where $\langle GG \rangle$ is the number of LBG-LBG pairs per radial bin, which is directly measured from our images, and $\langle RR \rangle$ is the number of random-random pairs per radial bin, measured using the random catalog created as I described in section 5.2.

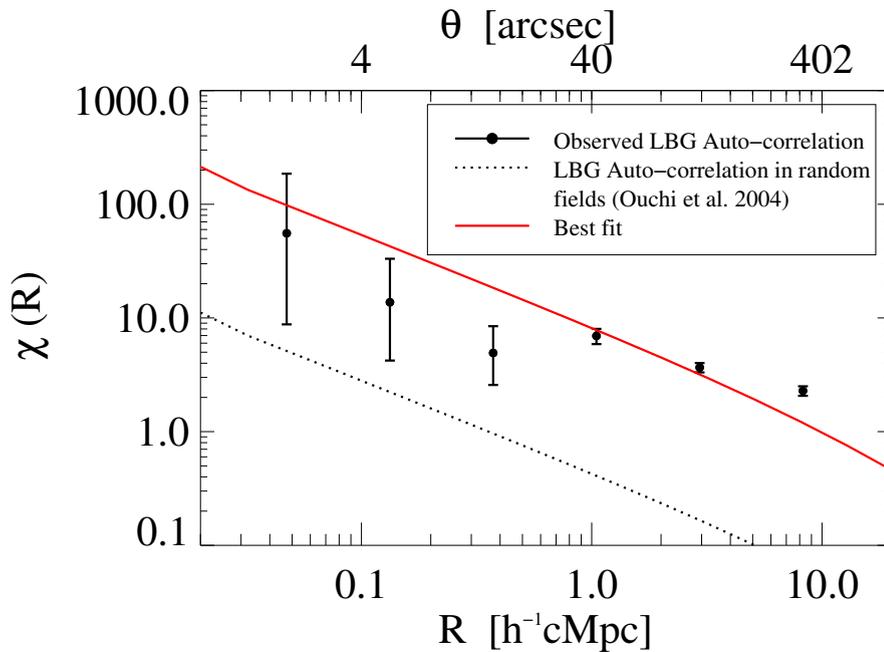


FIGURE 5.3: LBGs auto-correlation function in QSOs fields. The data points are showing our measurement described in section 5.3. The dotted black curve shows the results of LBGs auto-correlation in blank fields (Ouchi et al., 2004b). The solid red curve shows the best fit for our clustering measurements for a fix $\gamma = 1.8$. We find a higher clustering in our fields in comparison with random fields as measured by (Ouchi et al., 2004b), which suggests that our fields are located in overdense regions.

The equations for this case are very similar to the cross-correlation equations (5.6) and (5.12), but the difference is that here $\langle GG \rangle$ and $\langle RR \rangle$ are proportional to the square number density $n_G^2(z, < m_{\text{lim}})C^2(Z)$ and of course here we are measuring $\xi_{GG}(R, Z)$ instead of $\xi_{QG}(R, Z)$ in equation (5.6).

We computed χ for each field, and then we stacked the measurements. The results are presented in Fig. 5.3 and the numerical values in the table 5.2. The error bars are given by:

$$\Delta\chi_i = \frac{(\Delta\langle GG \rangle)_i}{\langle RR \rangle_i} \quad (5.20)$$

where $(\Delta\langle GG \rangle)_i$ is the error in the LBG-LBG pair counts for the i th radial bin estimated as in the section 5.2 using one-side Poisson confidence intervals for small number statistics.

In order to compare this clustering signal with the one computed in random fields, where QSO are not present, we overplotted the clustering measurement from Ouchi et al. (2004b). To draw this line, we took their r_0 and γ measurement and we computed

TABLE 5.2: LBGs Auto-Correlation Function.

| R(h^{-1} cMpc) | $\langle GG \rangle$ | $\langle RR \rangle$ | χ |
|-------------------|----------------------|----------------------|-------------------------------|
| 0.047 | 1 | 0.018 | $55.563^{+130.096}_{-46.778}$ |
| 0.133 | 2 | 0.136 | $13.737^{+19.438}_{-9.520}$ |
| 0.373 | 6 | 1.013 | $4.922^{+3.537}_{-2.349}$ |
| 1.048 | 56 | 7.041 | $6.953^{+1.063}_{-1.063}$ |
| 2.943 | 183 | 39.225 | $3.665^{+0.345}_{-0.345}$ |
| 8.264 | 222 | 67.563 | $2.286^{+0.221}_{-0.221}$ |

the value of χ using the equation 5.13 with the radial and redshift selection function of our images. This is shown as a dotted line in Fig. 5.3. The fact that our auto-correlation measurement is higher than the dotted line suggests that the LBGs in our fields are more clustered than LBGs in random fields, which indicates that our QSOs are tracing specially massive dark matter halos in the Universe.

As we did in the last section, we used a MLE to fit the r_0 and γ parameters. We obtained the best values to be $r_0 = 43.24^{+10.90}_{-4.74} h^{-1}$ cMpc and $\gamma = 1.22^{+0.07}_{-0.17}$. The corresponding 1 and 2σ 2D confidence regions are shown in a $r_0 - \gamma$ plane in Fig. 5.4 in blue and red color, respectively. In this case, r_0 is more poorly constrained if both parameters are free. Because of this, we fixed γ for the determination of the clustering amplitude. Although it makes sense to fix γ to its maximum likelihood value, we fixed this in $\gamma = 1.8$ in order to compare directly with the measurements of LBGs autocorrelation in random fields. When we fixed $\gamma = 1.8$, we obtained the best value to be $r_0 = 21.13^{+39.44}_{-14.78} h^{-1}$ cMpc, which is plotted in Fig. 5.3 as a red solid line.

Our results confirmed the findings of the last section. We detected a stronger clustering of LBGs in QSO fields in comparison with the LBGs clustering in random fields.

Although we expected this result, we should carefully consider them. Firstly, we must remember that the clustering properties are measured from a LBGs sample which is just $\sim 20\%$ complete.

Secondly, as I mention in section 4.2, our sample could be slightly contaminated, due to our novel technique to select LBGs seems to be sensitive to also select objects with intense lines in their spectra. We can not able to compute the exact contamination fraction in our sample, but as I describe in next section, our sample would not be highly contaminated. The purity of our sample can only be determined with spectroscopy on our LBGs candidates, or using an independent estimate of our number density, for example using control fields, or doing this study in much larger scales, where the background density could be computed far away from the quasar, outside of the overdensity.

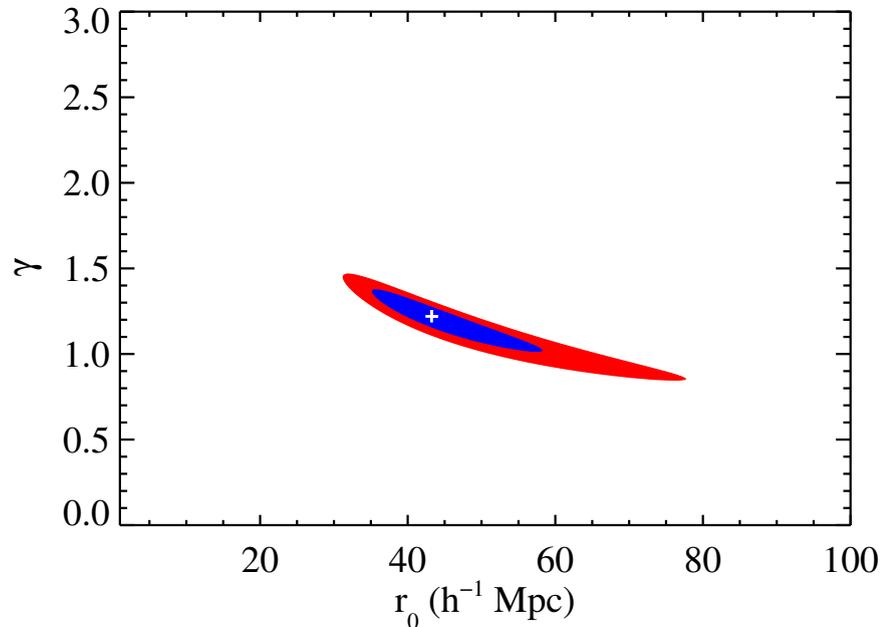


FIGURE 5.4: 1σ and 2σ confidence regions of r_0 and γ parameters (in blue and red respectively) determined using a maximum likelihood estimator for the LBGs auto-correlation measurement. The best estimation is shown as a white cross.

Thirdly, our clustering measurements depends on to the redshift selection function $C(z)$ we modeled in section 4.1, therefore, our results are sensitive to errors in that. If we work at high completeness, the error in the determination of $C(z)$ will be smaller and those will not affect importantly to the clustering estimation. On the opposite, if we work at low completeness (as is our case), the errors in the determination of $C(z)$ will be higher, affecting more importantly the clustering measurements. In other words, if we work at low completeness, our correlation function measurement is more sensitive to the lack of knowledge of the exact redshift selection function $C(z)$.

5.4 Contamination effects on the Clustering Measurements

An important issue to consider in LBGs selection is the purity of the sample. Since we are measuring clustering, we are interested in having a LBGs sample as clean as possible, avoiding contaminants. In section 4.3, the color cuts defining our selection region were chosen in order to increment the completeness of the sample as well as to decrease the low- z contaminants, just based on the color modeling of LBGs and on the mean evolutionary track of low- z galaxies. As I mention in that section, it is a challenge to measure the fraction of low- z contaminants in our LBGs sample, because as the filter set are not standard filters, we would have to fully model it, as we did with LBGs.

However, another independent method arisen, which allowed to check the contamination level in the sample, and it involved the use of the measured clustering properties. If we measure the cross-correlation function from the equation (5.1), with a highly contaminated sample, the observed number of LBGs around quasars $\langle QG \rangle$ will be overestimated, because the inclusion of low- z galaxies which will be counting as real LBGs. However, as the denominator $\langle QR \rangle$ is a theoretical expectation, which is computed based on the luminosity function of LBGs and on the redshift selection function of them, this value is not including those extra number counts due to contamination. This implies that the measured correlation does not behave like a power-law, this will show more like a flat shape in the correlation function.

In order to understand better this statement, we consider the measured value of χ as:

$$\chi = \frac{\langle QG^{obs} \rangle}{\langle QR^{exp} \rangle} - 1 \quad (5.21)$$

where $\langle QG^{obs} \rangle$ is measured directly from the images, and $\langle QR^{exp} \rangle$ is computed using the luminosity function and the redshift selection function, as we describe in the section 5.2. If we replace explicitly the equations (5.6) and (5.12) we have:

$$\chi = \frac{n_G^{obs}(z, < m_{lim})V_{eff}^{obs}}{n_G^{exp}(z, < m_{lim})V_{eff}^{exp}}(1 + \xi_{vol}) - 1 \quad (5.22)$$

where we define ξ_{vol} as the volume averaged correlation function:

$$\xi_{vol} = \frac{\int_{Z_{min}}^{Z_{max}} \int_{R_{min}}^{R_{max}} C(Z)\xi_{QG}(R, Z)2\pi R dR dZ}{V_{eff}} \quad (5.23)$$

If the number density of LBGs and the redshift selection function, encapsulated in the $C(Z)$ value, are well estimated, such that they well represent the observed quantities, then $n_G^{obs}(z, < m_{lim}) = n_G^{exp}(z, < m_{lim})$ and $V_{eff}^{exp} = V_{eff}^{obs}$ and we recover the equation (5.13), which have a power law shape. On the opposite, if none of them is well estimated, then we will have a numerical factor $X \neq 1$ in our equation, such that:

$$\chi = X(1 + \xi_{vol}) - 1 \quad (5.24)$$

which do not have a power law form anymore. We show in Fig. 5.5, the shape of χ for different values of X , where we assumed a power law shape for ξ_{vol} as in the equation (5.7). It is clear that the correlation function will flatten for increasing values of X . This

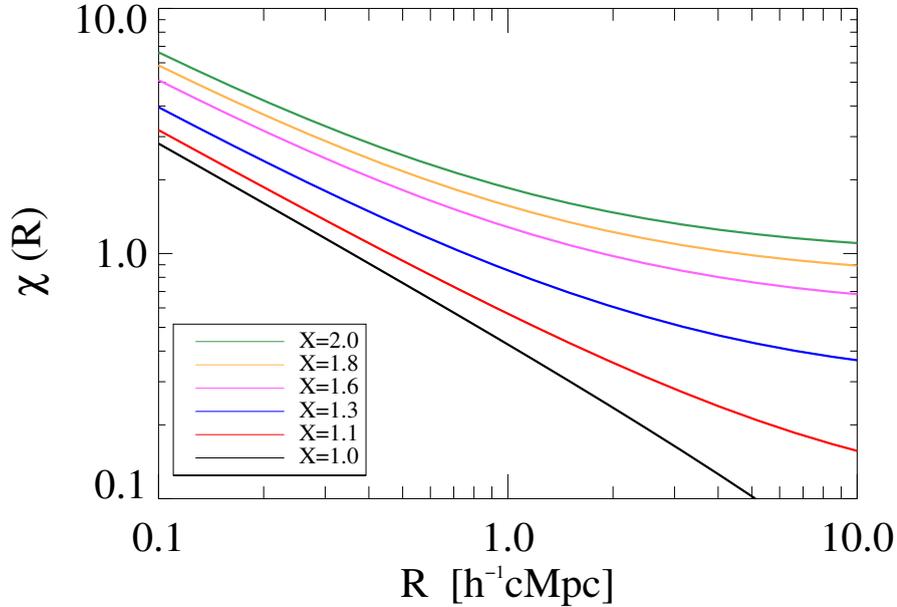


FIGURE 5.5: Shape of the function $\chi = X(1 + \xi_{vol}) - 1$ for different X values, where ξ_{vol} is modeled as in equation (5.23) using an arbitrary r_0 and γ value. When $X = 1$ we recover the power law shape of the function, but as X increase, the shape become increasingly flat.

flattening will be even stronger for the auto-correlation measurement because $\langle GG \rangle$ and $\langle RR \rangle$ are proportional to the square of number density.

Then, as a first sign of high contamination or a poor constrain of $C(Z)$ we should detect a flat cross and auto-correlation function in Fig. 5.1 and 5.3. Indeed, when we chose the preliminary selection region to select LBGs (which was bigger than the final one), we obtained a very flat measurement, then we iteratively reduced the size of the selection region until to obtain a shape consistent with a power-law.

However, note that the χ flattening could be due to either high contamination or poor constrain of the redshift selection function, but we could not know which is the responsible of the flattening, in other words, we could not disentangle both effects on the shape of the correlation. Considering this, another option to check contamination on our sample was available.

If we forget for a while that we are located on an overdense region, we can compute an angular correlation function using the standard procedure, where RR is computed from the data itself, as we described in section 2.3.3. Since we could be mixing galaxies at different redshifts, it is more appropriate to measure angular distances instead of the physical ones. If we use this method to measure the clustering, we avoid the uses of

theoretical expectations of the LBGs number density and the redshift selection function, and then in this case we always have $X = 1$ in the equation (5.22).

With this method we are measuring just the clustering of the sources, and our results will not depend on the modeling of the color selection. In that way, if the sample is highly contaminated, then we expect the correlation of this sample to be consistent with zero in all the angular scales, since the sample includes galaxies at several different redshifts, which should not be correlated at all. On the other hand, if we are not highly contaminated, we expect to measure a power law here. Note however that this is not a measurement of the clustering in our field, but this is just a way to test contamination. In other words, if we are not contaminated, this correlation function will show the right slope, but not the right clustering amplitude in the overdense regions.

We used our LBGs sample and we computed the angular auto-correlation function as:

$$\omega(\theta) = \frac{GG}{RR} \left(\frac{n_R}{n_G} \right)^2 - 1 \quad (5.25)$$

where n_G and n_R are the number density of objects in the data sample and the random sample respectively. Here the random catalog is built from the survey itself. That means that we computed the total number of LBGs candidates in all fields, then we divided that by the total unmasked area to get the number density on average. Then, we multiplied the unmasked area per image by this average number density to determine the number of galaxies expected in that field. Finally, we increased the number of counts for a big factor in order to decrease the noise in the measurement, and we randomly distributed those sources on the image and then we measured RR

We computed this angular auto-correlation for our selection region, and performed a power law fit given by $\omega(\theta) = A\theta^{-\beta}$ to quantify how consistent the measurement is with a flat shape (where $\beta = 0$). The results of this test are shown in Fig. 5.6.

Based on this test, we think that our LBGs sample is not highly contaminated, because the best fitted parameters are given by $A = 31.1$ and $\beta = 1.15$. Using this test we are not able to specifically constrain the contamination fraction of the sample. To compute that fraction, we would have to do a detailed modeling of the effect of contamination on clustering measurements. In other words, we would have to model how $\omega(\theta)$ looks if x% of the sample is contaminated.

To summarize, we claim we are not contaminated. Firstly, this is based on the fact that our volume averaged cross and auto correlation function are not showing a flat shape in Fig. 5.1 and 5.3. Given the possibility that this may be a consequence of the joint

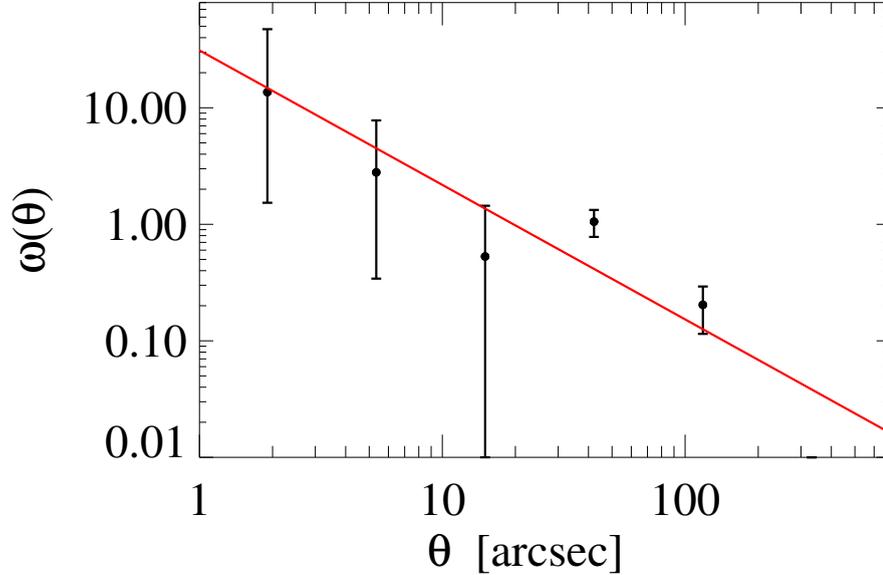


FIGURE 5.6: Angular cross-correlation function measurement for the LBGs sample. This measurement is used to test if the sample is contaminated. We expect that if the sample is highly contaminated $\omega(\theta) = 0$ at every scale and a power law fit with slope β is consistent with $\beta = 0$. We found that our LBGs sample is not highly contaminated, given their power law shape in this plot, which is well fitted by $\omega(\theta) = 31.1\theta^{-1.15}$.

contribution of redshift selection modeling and contaminants on the shape of the clustering, we performed a second test. This does not assume any modeling, the measured correlation is real, which is not consistent with zero, suggesting no contamination.

We recognize that this is an uncertainty in this study because the lack of an exact measurement of the contaminants fraction. However, this is very hard to quantify, but we are doing every effort to avoid contamination: we are modeling the low- z galaxies evolutionary track (in section 4.2), and avoiding their location in the color-color diagram, we are not considering sources brighter than $r_{\text{GUNN}} = 23.97$ in order to decrease the number of low- z contaminant in case they were affecting the sample. Finally, we tested the contamination effect on the clustering measurements, which showed to be consistent with no contamination.

Chapter 6

The First QSO-LAE Cross-Correlation Measurement at $z \sim 4$

In this Chapter I describe and show the results of the first QSO-LAE cross-correlation measurement at $z \sim 4$. I proposed the observation of this data during my PhD. We were allocated with 30 hours in the VLT/FORS2 which allowed us to obtain imaging on 14 QSO fields. This is the largest data sample at $z \sim 4$ so far, which enables us to tackle cosmic variance effects and doing clustering measurements with high signal to noise. This data complements the LBGs clustering measurements presented in this thesis and give us a complete understanding of the clustering properties of galaxies in QSO environments at $z \sim 4$ when the Universe was only $\sim 1,5$ Gyr.

The analysis is quite similar of what I describe in chapters 3, 4, and 5, with some small modifications due to the differences in the properties of LAEs and LBGs.

6.1 Data Description

Lyman alpha emitters (LAEs) are galaxies with a strong Lyman alpha emission line in their spectra. The LAEs selection is a little different from the LBGs selection because, in the case of LAEs, a narrow band filter is located just over the $\text{Ly}\alpha$ line and a broad band filter over the continuum. If the galaxy has a strong emission line, then it is expected to detect a flux excess in the narrow band filter compared with the broad band filter. This is a very used technique to detect LAEs at high redshift.

Additionally, a third broad band filter is usually located blueward of the Ly α line, in order to detect the Lyman break due to the neutral hydrogen absorption by the IGM, as is done for LBGs detection.

The differences in the techniques to select LAEs and LBGs causes that the observational strategy and photometry be slightly different. In this section, I provide detailed description about the QSO target selection, observations, and photometry performed to efficiently select LAEs at high redshift.

6.1.1 QSO Targets

For this study, we observed 14 QSO fields using a narrow-band with central wavelength at $\lambda = 5930\text{\AA}$. We have chosen this narrow band to identify LAEs at $z \sim 3.87$ associated with the central QSO. We thus selected QSOs from the SDSS/BOSS QSO catalog in such a way that Ly α lands within the core of our filter at $z \sim 3.87$. Since we have stacked the LAEs counts from all the fields to measure clustering, it is important to select QSOs with an accurate redshift determination and a spanning of a very narrow redshift slice.

QSO redshifts are determined using a custom line-centering code that corrects the known relative shifts between broad rest-frame UV emission lines, and are accurate up to 800km/s, which is much narrower than the narrow band filter width ($\sim 3200\text{km/s}$). Further, we selected only targets falling inside of the central 1000km/s of the narrow-band filter in order to ensure that LAEs selected with this band are associated to the central QSO in each field.

For the same reason that we explained in 3.2, we just selected QSOs without radio emission counterpart reported in the the VLA FIRST catalog. We also discarded QSOs with high extinctions ($A_\lambda > 0.2$). Finally, we selected the brightest QSOs, to ensure they have massive black holes. Our final sample is composed by QSOs in a thin redshift slice of $\Delta z \sim 0.02$ and with bright magnitudes $i < 20.6$. A summary of the QSOs properties are shown in the table 6.1.

6.1.2 Observations

We carried out imaging for 14 QSO fields with FORS2¹ instrument on the VLT telescope between September, 2014 and March, 2015. We observed each QSO field using the narrow band $HeI/2500 + 54$ ($\lambda = 5930\text{\AA}$, FWHM = 63\AA) and the broad bands g_{HIGH} and $R_{SPECIAL}$. This configuration enabled us to detect LAEs at $z \sim 3.87$ (see Fig. 6.1).

¹Focal Reducer and low dispersion Spectrograph 2 (Appenzeller and Rupprecht, 1992)

TABLE 6.1: Targeted Quasars properties.

| Name | RA (J2000) | DEC (J2000) | Redshift | i |
|-----------|--------------|--------------|----------|-------|
| SDSSJ0040 | 00:40:17.426 | +17:06:19.78 | 3.873 | 18.91 |
| SDSSJ0042 | 00:42:19.748 | -10:20:09.53 | 3.865 | 18.57 |
| SDSSJ0047 | 00:47:30.356 | +04:23:04.73 | 3.864 | 19.94 |
| SDSSJ0119 | 01:19:59.553 | -03:42:16.51 | 3.873 | 20.49 |
| SDSSJ0149 | 01:49:06.960 | -05:52:18.85 | 3.866 | 19.80 |
| SDSSJ0240 | 02:40:33.804 | +03:57:01.59 | 3.872 | 20.03 |
| SDSSJ1026 | 10:26:32.976 | +03:29:50.63 | 3.878 | 19.74 |
| SDSSJ1044 | 10:44:27.798 | +09:50:47.98 | 3.862 | 20.52 |
| SDSSJ1138 | 11:38:05.242 | +13:03:32.61 | 3.868 | 19.10 |
| SDSSJ1205 | 12:05:39.550 | +01:43:56.52 | 3.867 | 19.37 |
| SDSSJ1224 | 12:24:20.658 | +07:46:56.33 | 3.867 | 19.08 |
| SDSSJ1258 | 12:58:42.118 | -01:30:22.75 | 3.862 | 19.58 |
| SDSSJ2250 | 22:50:52.659 | -08:46:00.22 | 3.869 | 19.44 |
| SDSSJ2350 | 23:50:32.306 | +00:25:17.23 | 3.876 | 20.61 |

The total exposure time for $HeI/2500 + 54$, R , and g was 3660s, 360s, and 900s respectively. Those exposure times were chosen in order to reach a similar limit magnitude as in Ouchi et al. (2008), who computed the luminosity function of LAEs at $z = 3.7$, which is used for our background number density computation. We got dithered frames in order to fill the gap between the CCDs. Spectrophotometric stars were taken each night to calibrate $HeI/2500 + 54$ and g images, and photometric stars from Stetson fields were observed several times during the course of the night for R images calibration. In order to increase the chance of target observation, we relaxed the seeing requirement to 1.3 arcsec, then our images have a broad range of seeing corresponding to 0.6 - 1.3 arcsec.

6.1.3 Data Reduction

We performed the data reduction and stacking in the same way as we described in 3.4. However, the photometric calibration was done before the stacking process, because the individual frames were taken in different nights.

The photometric calibration for individual frames of HeI and g was done using spectrophotometric stars. For that, we computed the star magnitude convolving the filter transmission curve with the spectra and then we compared with instrumental magnitudes obtained using SExtractor on the science images. In the case of R images, we calibrated them comparing the tabulated standard stars magnitudes from the Stetson fields (Stetson, 2000), with instrumental magnitudes on the science images, after correcting the tabulated magnitudes by the color term to take into account the difference

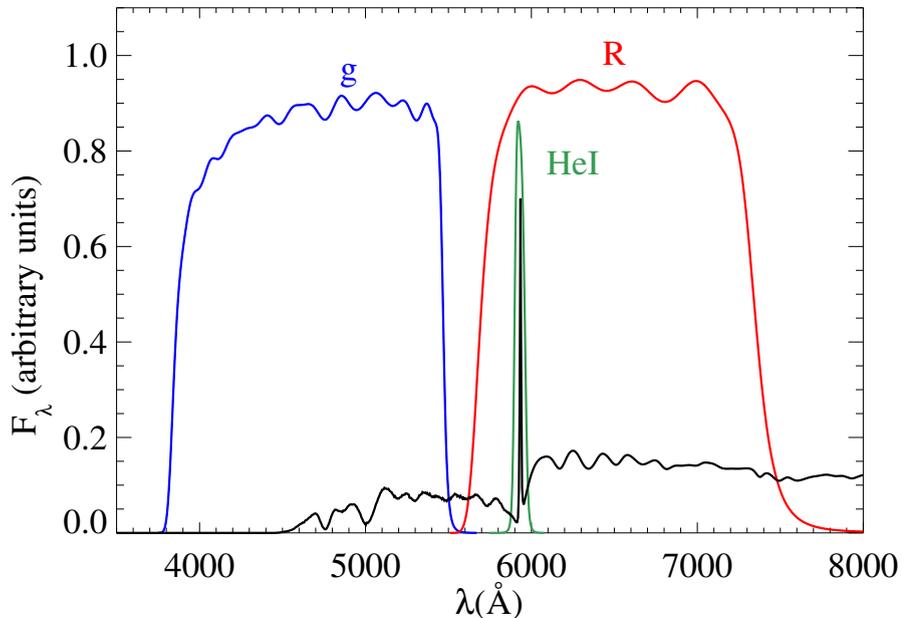


FIGURE 6.1: Filter configuration shown on a galaxy simulated spectra at $z \sim 3.87$.

between the Stetson and FORS2 R filter curve. We used all the observed stars of the field, and then we compute a median value for the final zeropoint value.

For the photometric calibration, we also included airmass and galactic extinction correction using the same procedure as we described in 3.4.

6.1.4 Photometry

Object detection and photometry were performed using SExtractor. Since we are interested in detection of LAEs, which have a strong $\text{Ly}\alpha$ line located in the core of the HeI narrow band, we used this image to detect objects, and then we performed photometry on the broad bands in the same positions and apertures.

The background was calculated in regions of 64 pixels in size, and then recomputed locally in an annulus area of 24 pixels of thickness centered around the object. In order to maximize the detection of faint sources, the detection image was smoothed by applying a Gaussian filter of seeing FWHM of 2 pixels and size 3×3 pixels. Every group of at least 7 contiguous pixels having a value above 1.8σ , was defined as an object. Those parameters were chosen after iterating with different values and checking how many spurious objects were detected, computed as in Hennawi et al. (2015).

TABLE 6.2: 5σ limit magnitudes per field measured in a $2''$ diameter aperture.

| Name | <i>HeI</i> | <i>g</i> | <i>R</i> |
|-----------|------------|----------|----------|
| SDSSJ0040 | 25.12 | 26.30 | 25.39 |
| SDSSJ0042 | 25.13 | 26.38 | 25.39 |
| SDSSJ0047 | 24.77 | 26.12 | 25.46 |
| SDSSJ0119 | 25.13 | 26.49 | 25.32 |
| SDSSJ0149 | 25.23 | 26.49 | 25.37 |
| SDSSJ0240 | 24.84 | 25.98 | 25.44 |
| SDSSJ1026 | 25.29 | 26.46 | 25.33 |
| SDSSJ1044 | 25.31 | 26.47 | 25.50 |
| SDSSJ1138 | 24.65 | 25.82 | 25.19 |
| SDSSJ1205 | 25.25 | 26.46 | 25.53 |
| SDSSJ1224 | 25.16 | 26.33 | 25.25 |
| SDSSJ1258 | 25.11 | 26.42 | 25.47 |
| SDSSJ2250 | 25.28 | 26.3 | 25.40 |
| SDSSJ2350 | 25.11 | 26.22 | 25.34 |

We estimated the objects magnitudes using aperture photometry after convolving the images with a Gaussian kernel to match the seeing value. The magnitudes were measured in a fix aperture of $2''$ of diameter. Magnitudes of objects not detected or detected with a signal to noise lower than 2 either in *g* or *R* were replaced by the corresponding 2σ limiting magnitude. The mean 5σ limiting magnitude of the reduced images were 25.10 for *HeI*, 26.30 for *g* and 25.38, for *R* for an $2''$ diameter aperture. The limit magnitudes per field are presented in table 6.2.

If we consider the limit magnitudes reached in the narrow band and in the *R* band, we can compute the smaller Ly α EW that we were able to detect on the images. The $EW_{Ly\alpha}$ is defined as:

$$EW = \frac{F_{Ly\alpha}}{f_{cont}} \quad (6.1)$$

where $F_{Ly\alpha}$ is the Ly α flux (in $\text{erg cm}^{-2} \text{s}^{-1}$ units) and f_{cont} is the density flux of the continuum (in $\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$ units). In our case, the broad band *R* is containing the continuum and the Ly α line, so, the density flux of the continuum has to be computed using the combined information from the narrow band and the broad band. Following the relations from Yang et al. (2009) we can write it as:

$$f_{cont}^{\lambda} = \frac{F_R - F_{NB}}{\Delta\lambda_R - \Delta\lambda_{NB}} \quad (6.2)$$

where $\Delta\lambda_R$ and $\Delta\lambda_{NB}$ are the FWHM of the R and NB filters respectively. Then the line flux can be written by:

$$F_{Ly\alpha} = F_{NB} - f_{cont}^{\lambda} \Delta\lambda_{NB} \quad (6.3)$$

If we use the mean 5σ limit magnitudes of both NB and R filters, we can compute the limit fluxes, and then $F_{Ly\alpha,limit}$ and $f_{cont,limit}^{\lambda}$. Replacing in the equation (6.1), we computed the limit EW that we were able to detect, obtaining $EW_{Ly\alpha,limit} \sim 38\text{\AA}$ (corresponding to an $EW_{Ly\alpha,RF} = 7.8\text{\AA}$ in rest frame). This means that in our images we were able to detect galaxies with $EW_{Ly\alpha}$ greater or equal to this value for a detection of 5σ . This shows us that the deep of our images is enough to detect LAEs, since they are typically defined as galaxies with $EW_{Ly\alpha} \gtrsim 20\text{\AA}$. Using the $F_{Ly\alpha,limit}$ value, we also computed the limit of the Ly α luminosity, obtaining $L_{Ly\alpha,limit} = 1 \times 10^{42} \text{ erg s}^{-1}$.

6.2 LAEs Selection

As we mentioned before, the main feature of LAEs is their strong Ly α emission line. For that reason, the technique to identify them is by the detection of a flux excess in the narrow band *HeI*, compared with the flux in the continuum traced by our broad band *R*, then we expect a blue $R - HeI$ color. On the other hand, the second broad band is located in order to detect the Lyman break, which can be detected by a red $g - R$ color. Considering this, a suitable way to select LAEs is studying the $g - R$, $R - HeI$ color-color diagram.

In this case, the process to choose the selection region is different as we did for LBGs selection. In the case of LAEs, the selection technique is standard, and the redshift range covered by the filters are very known and easy to calculate. Then, in principle if we adopt the same LAEs selection as previous workers, it is not necessary to calculate a completeness function to compute the clustering, because by construction we should get the same mean number density as them. In the case of LBGs, the redshift selection function was absolutely unknown, because nobody had used that filter configuration before. In that case, the completeness computation to measure the clustering was indispensable. However, as we discuss in chapter 5, this theoretical computation introduce uncertainties in the clustering measurement, so, here we chose to perform the same color as other workers, and therefore we avoid the use of a theoretical computation for the selection function.

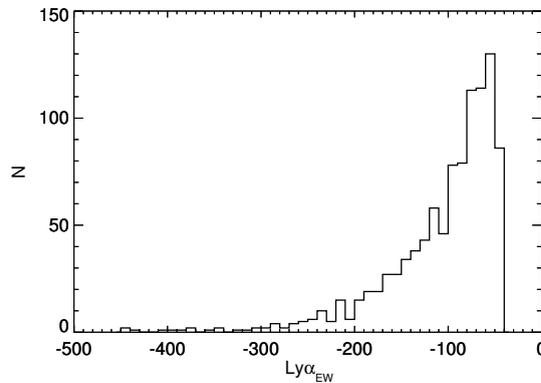


FIGURE 6.2: Distribution of the $EW_{Ly\alpha}$ used to create the simulated spectra. Negative values correspond to emission lines. EW is chosen from an exponential distribution with scale length of $W_0 = -64\text{\AA}$ (Ciardullo et al., 2012). The mean $EW_{Ly\alpha}$ of this distribution is $\sim 90\text{\AA}$.

Even though it is not necessary to compute the completeness of the sample, we use our Monte Carlo simulation described in section 4.1 for modeling the LAEs colors in order to have an intuition of the location of them in the color-color diagram.

6.2.1 Selection Region Choice

To simulate the colors of LAEs, we used the Monte Carlo simulation described in section 4.1, but the only difference is that this time we created each spectrum model assigning it a randomly chosen $EW_{Ly\alpha}$ taken from a different distribution. In the case of LBGs, we considered spectra with both cases emission and absorption Ly α line. In this case, we only considered spectra with the Ly α emission line, and then the distribution is built as an exponential function with rest-frame scale length of $W_0 = -64\text{\AA}$ (Ciardullo et al., 2012).

We took 1000 $EW_{Ly\alpha}$ values from that distribution, but we only considered $EW_{Ly\alpha,RF} \geq 44\text{\AA}$ since we only select LAEs with this EW limit (as I explain below). The $EW_{Ly\alpha}$ distribution used in this simulation is shown in Fig. 6.2.

We also included photometric errors in our simulated colors, but this time assigning to each spectrum a random value for the NB filter, taken from the Ly α luminosity function integrated in the magnitudes limits of our observations. We used the Schechter parameters from Ouchi et al. (2008), who measured the luminosity function for LAEs at $z \sim 3.7$, based on a photometric sample of 101 LAEs and a spectroscopic sample of 26 LAEs, with $EW_{Ly\alpha,RF} \gtrsim 44\text{\AA}$. The best parameters are given by $\phi^* = 3.4 \times 10^{-4} \text{ Mpc}^{-3}$, $L_{Ly\alpha}^* = -10.2 \times 10^{42} \text{ erg s}^{-1} \text{ mag}$, and $\alpha = -1.5$.

The results of this color modeling are shown in Fig. 6.3. We plotted the color of each simulated spectrum as color coded according to the redshift. We overplotted as a continuum curve the mean LAE evolutionary track, where the large filled circle is indicating the position of a LAE at $z = 3.87$. The region used to select LAEs is shown as a black dashed line.

The purple, blue, and some of the green points in this plot are LAEs spectra at $z < 3.87$, where both, the NB and R band are located over the continuum redward to the Ly α line. Since the continuum of LAEs is nearly flat, the magnitudes of NB and R are similar, then $R - NB = 0$. Orange and red points in this plot correspond to LAEs spectra at $z > 3.87$ and the NB is located blueward to the Ly α line, where the flux is strongly attenuated by the IGM. The R band is mostly over the continuum, then a red color is expected in the $R - NB$ axis. When the LAEs is at $z = 3.87$, the NB is just over the Ly α line, then we expected blue colors in $R - NB$. The mean evolutionary track shown in the figure, has an $EW_{\text{Ly}\alpha, \text{RF}} \sim 90 \text{\AA}$, then the color of any LAE with a more intense Ly α line will show bluer colors. Furthermore, the g band is located blueward to the Ly α line, then we expected a red color in $g - R$ due to the Lyman break detection. Considering this, we expect the LAEs be located in the upper right region in this diagram. The large scattering of $z = 3.87$ LAEs colors are due to the photometric errors. When we simulated them, we assigned a noise to the photometry which is greater for fainter magnitudes. The $z = 3.87$ LAEs are not detected or faintly detected in the g band, and for those faint magnitudes, the photometric error is big, which produces a large scattering in the $g - R$ axis of this plot.

We also studied the low- z galaxies location, using the same spectral templates as in section 4.2, and convolved them with our three filters for different redshift, ranging from 0 to 3. The results are shown in Fig. 6.4. In this case, the locus of the low- z galaxies is well isolated from the $z = 3.87$ LAEs location, then the contamination is not a problem and this allows us to have a sample with high completeness and purity.

As we explain in the next section, for the clustering measurements, we used the luminosity function of LAEs at $z = 3.7$ computed by Ouchi et al. (2008), who selected LAEs based on Subaru Suprime-Cam imaging data using the filter set B , NB , V . This configuration is similar to our configuration g , HeI, R, and their V band is also including Ly α , as our R band do. In order to avoid a redshift selection function computation, we adopted the same color cuts as their work, then this ensures us that our completeness and contamination is the same as their, and we can use directly their luminosity function to compute the number counts in random fields. Our NB limit magnitude is a bit deeper than the used in their study. We reached a median value of $HeI = 25.1$, while they

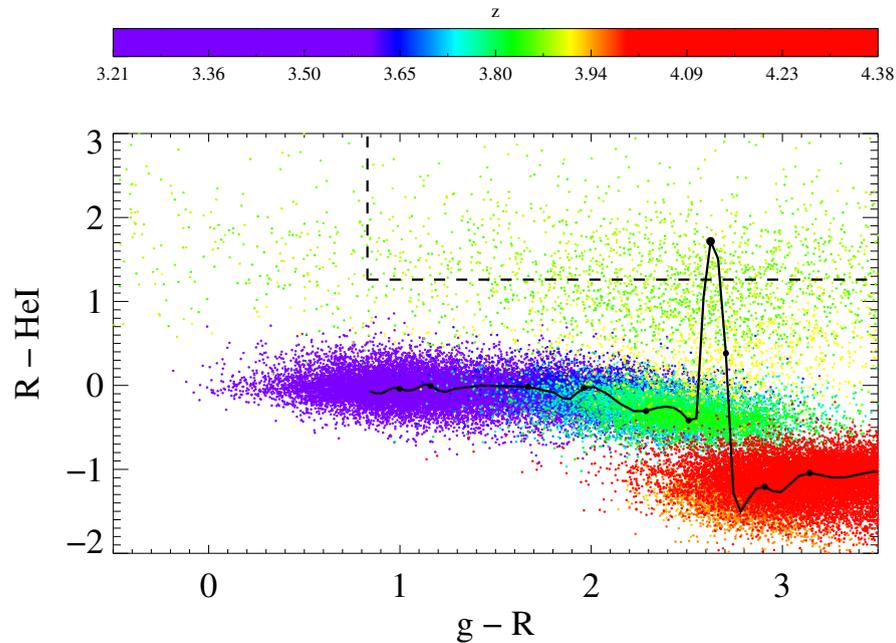


FIGURE 6.3: Color-color diagram showing the simulated colors of 1000 LAEs spectra, including photometric errors and plotted as redshift color-coded points according to the color bar. The mean LAE evolutionary track is also plotted as a black curve. The filled points over the curve from the left to the right indicate the redshift from 3.3 to 4.1. The larger circle shows the exact position of the mean $z = 3.87$ LAEs colors and the black dashed line is indicating the LAEs selection region.

reached $NB = 24.7$. However, they reached 1.4 mag deeper in their broad band V in comparison with our R band.

First we defined a lower limit for the detection of the $Ly\alpha$ line. This limit is basically defining a limit in the $EW_{Ly\alpha}$ of the detected LAEs. We chose this color cut in $R - HeI = 1.26$, which corresponds to an $EW_{Ly\alpha,RF} \sim 44\text{\AA}$ (computed using equations (6.2) and (6.3)). This value is chosen to match with the $EW_{Ly\alpha,RF}$ used in Ouchi et al. (2008).

The second color cut that we defined is related to the detection of the Lyman break, and we chose this cut in $g - R = 0.83$ which is equivalent to the color cut used in Ouchi et al. (2008), which is given by $B - V = 0.7$.

We summarize the color cuts in the equations:

$$R - HeI > 1.26 \quad (6.4)$$

$$g - R > 0.83 \quad (6.5)$$

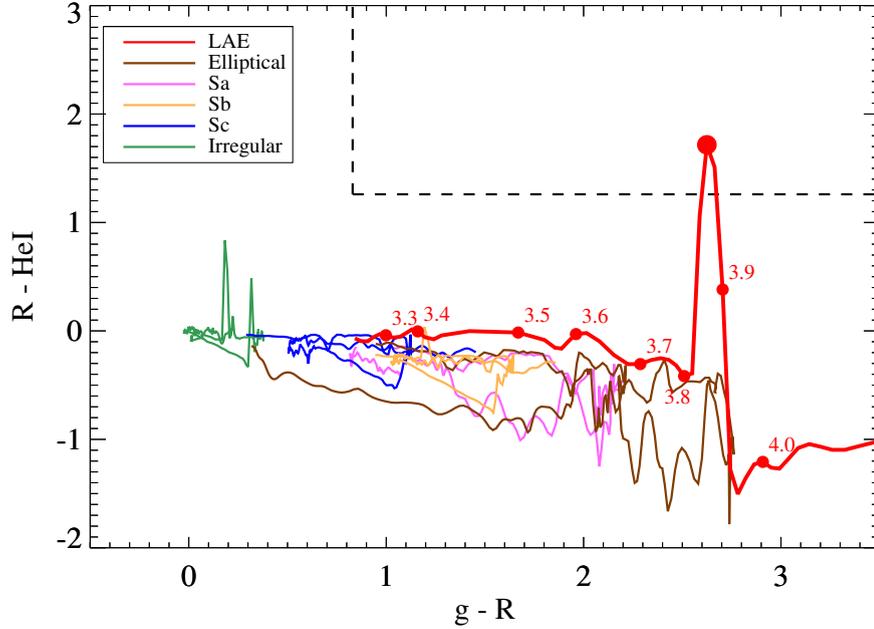


FIGURE 6.4: Low- z galaxies evolutionary tracks redshifted from $z = 0$ to $z = 3$. We plot as brown, magenta, orange, blue, and red curves the evolutionary track of elliptical, Sa, Sb, Sc, and irregular galaxies respectively. We overplotted the track of LAEs as a red curve. Filled circles over the red curve indicate colors of LAEs from redshift 3.3 to 4.0, and the largest red point indicate the exact position of the color of LAEs at $z = 3.87$. We overplotted the selection region as a black dashed line.

6.2.2 LAEs Sample and Completeness

In each field we selected objects fulfilling the color criteria described in the last section, and we only considered objects detected with $S/N \geq 5$ in the narrow band filter, in order to ensure the Ly α line detection. The noise is computed in the same way as we described in section 4.3. We show the color-color diagram of the 14 stacked fields in Fig. 6.5. We detected 17 LAEs candidates, which correspond to a number density of 2.88×10^{-2} LAEs arcmin $^{-2}$ and 1.55×10^{-4} LAEs Mpc $^{-3}$ (where we have computed the volume of the survey by multiplying the area of the survey by the FWHM of the NB, which is given by $\Delta z = 0.052$ corresponding to $\Delta Z = 28.07$ Mpc h $^{-1}$). The majority of them are not detected at 2σ in either R and g filters. In 5 fields we did not detect any LAEs.

Given that our R imaging is shallower than the V broad band used in Ouchi et al. (2008), our $R - HeI$ color cut is excluding 9 possible LAEs candidates which show $R - HeI$ lower limits (arrow pointing up in the diagram). If we had R images 1.4 mag deeper, the 2σ limit magnitude assigned in R when there is no detection, would be fainter, and then the lower limits would move to bluer colors, possibly fulfilling our $R - HeI$ color

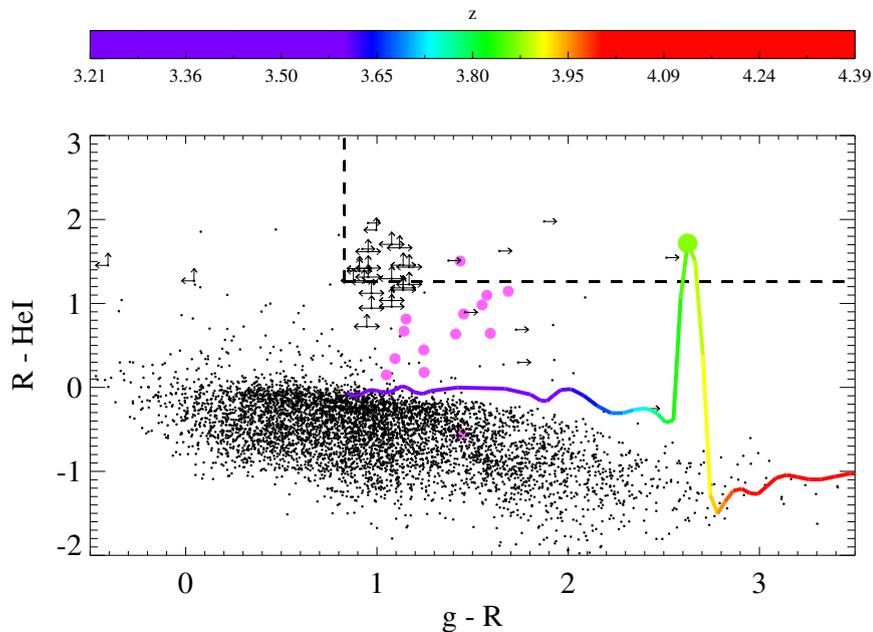


FIGURE 6.5: Color-color diagram for our 14 stacked QSO fields. Here the evolutionary track showed in Fig. 6.4 is plotted as redshift color-coded track according to the color bar. The magenta points indicate the color of the QSO in our filters. The magnitudes of cases in which the object was not detected in g or R filter at 2σ level, were replaced by the corresponding limit magnitude. Those cases are shown with arrows. Cases with no detection on either filter are shown with both arrows. The dashed line is indicating the selection region defined by equations (6.4) and (6.5).

requirement. Considering this, we compute our clustering including and non including those 9 LAEs in order to study the impact of them in the sample. However, it is needed deeper R images to confirm their nature. If we include those 9 new LAEs candidates, we obtain a total of 26 LAEs, corresponding to a number density of 4.41×10^{-2} LAEs arcmin^{-2} and 2.37×10^{-4} LAEs Mpc^{-3} .

In Table 6.3 the complete LAEs sample is presented, in Fig. 6.6 we show some examples of LAEs candidates, and in Fig. 6.7 we show the distribution of the LAEs for our 14 fields. In appendix B the images of all the LAEs are presented together with the color-color plots of the individual fields.

Finally, we computed the completeness of the sample for each field by computing the fraction of the simulated spectra in our Monte Carlo which are recovered by our color cuts for each redshift step. We show a median completeness function in Fig. 6.8. We obtained a completeness of $\sim 60\%$ at $3.86 < z < 3.88$ and our color cuts also select LAEs with high completeness (50%) in the redshift range $3.88 < z < 3.90$ and with a much lower completeness in the range $3.84 < z < 3.86$. This corresponds to the fraction of LAEs with $EW_{Ly\alpha,RF} = 44\text{\AA}$ and magnitude $HeI \lesssim 25.1$ recovered by our selection

TABLE 6.3: LAEs sample. The magnitudes correspond to AB magnitudes measured in a $2''$ diameter aperture for each filter

| ID | RA (J2000) | DEC (J2000) | R | g | HeI |
|--------------|---------------|----------------|--------|--------|-------|
| SDSSJ0040 1 | 10.0416 | 17.1033 | >26.39 | >27.30 | 25.00 |
| SDSSJ0040 2 | 10.0318 | 17.0782 | >26.39 | >27.30 | 24.97 |
| SDSSJ0119 1 | 20.0031 | -3.6507 | 26.53 | >27.49 | 24.57 |
| SDSSJ0119 2 | 19.9961 | -3.7038 | 25.61 | >27.49 | 23.63 |
| SDSSJ0119 3 | 19.9449 | -3.7202 | >26.32 | 27.31 | 24.44 |
| SDSSJ0119 4 | 20.0312 | -3.6762 | 24.97 | >27.49 | 23.43 |
| SDSSJ0119 5 | 20.0240 | -3.6775 | >26.32 | >27.49 | 24.88 |
| SDSSJ0119 6* | 19.9417 | -3.7107 | >26.32 | >27.49 | 25.09 |
| SDSSJ0149 1 | 27.3161 | -5.8986 | >26.36 | >27.48 | 24.70 |
| SDSSJ1026 1 | 156.6710 | 3.4884 | >26.32 | >27.46 | 24.87 |
| SDSSJ1026 2* | 156.6360 | 3.4466 | >26.32 | >27.46 | 25.16 |
| SDSSJ1026 3* | 156.6560 | 3.5526 | >26.32 | >27.46 | 25.14 |
| SDSSJ1044 1 | 161.0720 | 9.8350 | 25.82 | >27.46 | 24.20 |
| SDSSJ1044 2* | 161.1660 | 9.9040 | >26.49 | >27.46 | 25.55 |
| SDSSJ1044 3* | 161.1110 | 9.8730 | >26.49 | >27.46 | 25.37 |
| SDSSJ1205 1 | 181.4240 | 1.7269 | 26.54 | >27.46 | 24.89 |
| SDSSJ1205 2* | 181.4480 | 1.7262 | >26.53 | >27.46 | 25.27 |
| SDSSJ1224 1 | 186.1170 | 7.8037 | >26.24 | >27.32 | 24.54 |
| SDSSJ1224 2* | 186.0490 | 7.8310 | >26.24 | >27.32 | 25.21 |
| SDSSJ1224 3* | 186.0640 | 7.8136 | >26.24 | >27.32 | 25.28 |
| SDSSJ1258 1 | 194.6190 | -1.5410 | >26.46 | >27.42 | 25.15 |
| SDSSJ1258 2 | 194.6800 | -1.4501 | >26.46 | >27.42 | 25.04 |
| SDSSJ1258 3 | 194.6260 | -1.5227 | >26.46 | >27.42 | 24.84 |
| SDSSJ1258 4 | 194.7160 | -1.4827 | 26.04 | >27.42 | 24.53 |
| SDSSJ2250 1* | 342.7710 | -8.8220 | >26.39 | >27.34 | 25.67 |
| SDSSJ2350 1 | 357.6290 | 0.4162 | >26.33 | >27.21 | 25.06 |

* Correspond to 9 LAEs candidates which are not fulfilling our color cut, but they show a lower limit in the $R - HeI$ color. We would need deeper R images to confirm those are LAEs.

region. We computed an effective coverage in redshift as in section 4.4, and obtained $\Delta z = 0.027$, which corresponds to $\sim 1660 \text{ km s}^{-1}$ at $z = 3.87$.

Since we are doing the same selection as Ouchi et al. (2008), we should have the same completeness as them. For their faintest magnitude bin ($NB = 24.2 - 24.7$), they claim a completeness of $\gtrsim 50\% - 60\%$ which is in agreement with our completeness computation.

Regarding the contamination of the sample, Ouchi et al. (2008) used spectroscopy to estimate a contamination rate. They obtained a contamination within the range of $0\% - 14\%$ in their LAEs sample. Since we have the same selection function, we assume the same contamination. This implies that the LAEs sample here presented is much more complete and clean than the LBGs sample presented in chapter 4.

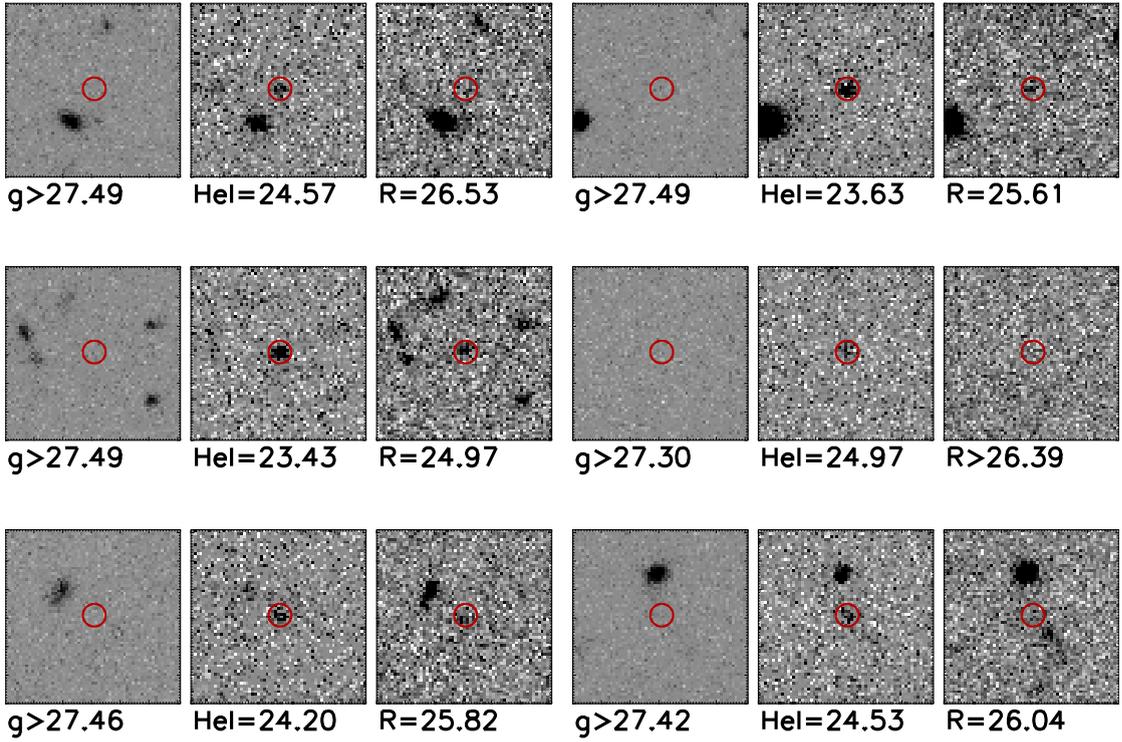


FIGURE 6.6: Images of some selected LAEs. From left to right we show the g , HeI , and R images. Each panel is $15''$ on a side. The red circle shows the position of the detected object, and the size correspond to the region in which the photometry was done ($2''$ in diameter). The magnitudes are indicated in each panel.

6.3 Clustering Measurements

In the same way as in section 5.1, we computed the expected clustering signal for the QSO-LAE cross-correlation function using the equation (5.6) and (5.12). In this case we compute $n_G(z, < m_{\text{lim}})$ from the luminosity function of LAEs at $z = 3.7$ (Ouchi et al., 2008), by integrating it from our limit of $Ly\alpha$ luminosity given by $L_{Ly\alpha} \sim 1.0 \times 10^{42} \text{ erg s}^{-1}$ (as computed in the section 6.1), up to infinity. This $L_{Ly\alpha}$ value is computed from the limit magnitudes of HeI and R for each field, respectively.

In this case we did not use the completeness function $C(z)$, but we assumed $C(z) = 1$ since the luminosity function was measured with a sample with the same completeness as ours (our LAEs sample have been built in that way). So, here the effective volume of the equation (5.5) is given by:

$$V_{\text{eff}} = \pi(R_{\text{max}}^2 - R_{\text{min}}^2)\Delta Z \quad (6.6)$$

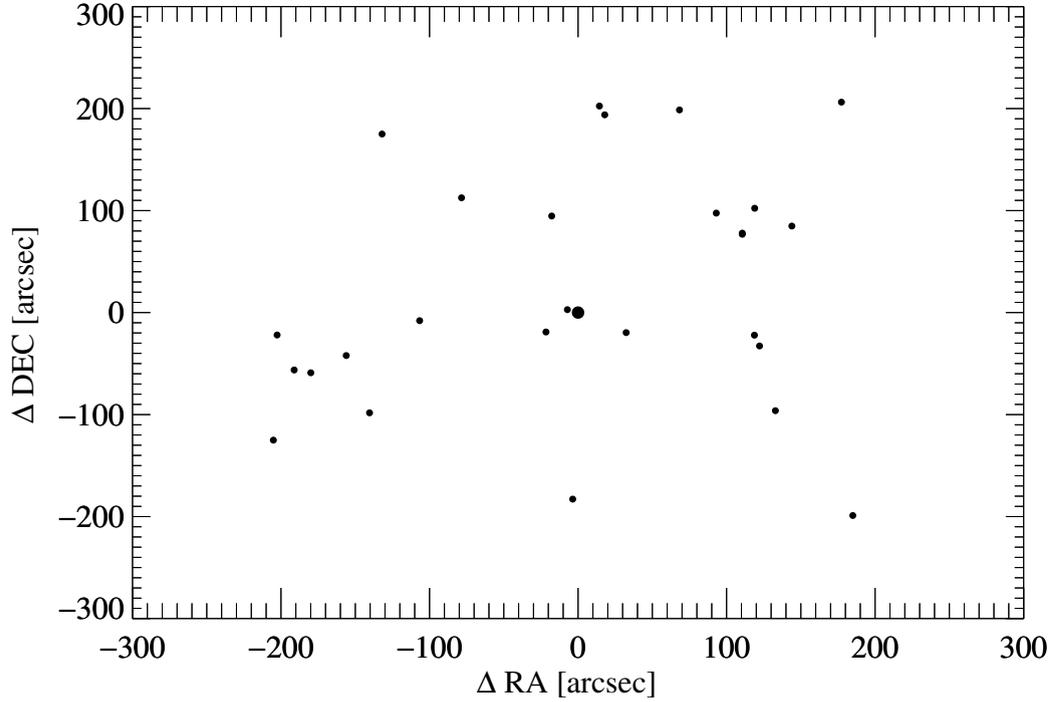


FIGURE 6.7: The distribution of LAEs around the quasar in the plane of the sky for the stacked 14 fields. The central QSOs is located in 0.0 and is plotted by a large black circle.

with ΔZ the comoving distance mapped by our filters, which is computed from the FWHM of the HeI band. We have $FWHM = 63.23$, then $dz = 0.052$ and $\Delta Z = 28.07 \text{ Mpc h}^{-1}$.

For the computation of the r_0^{QG} value, we used the auto-correlation lengths values for LAEs, given by $r_0^{GG} = 3.5 \text{ h}^{-1} \text{ cMpc}$ for a fixed $\gamma = 1.8$ value. This value corresponds to the auto-correlation of LAEs at $z = 4.86$ (Ouchi et al., 2003), but we assumed that the luminosity function do not show important evolution from 3.87 to 4.86. The resulting expected r_0^{QG} value is then $r_0^{QG} = 8.83 \text{ h}^{-1} \text{ cMpc}$ for a fixed $\gamma = 1.8$ value, which is a little lower than the expected for the QSO-LBG cross-correlation.

Following the same procedure as for LBGs, we computed the expected value for the QSO-LAE cross-correlation function, which is plotted in Fig. 6.9 as a dashed line. We also plotted our measured QSO-LAE cross-correlation function as data point, which is computed following the same procedure described in section 5.2. We have used the sample which include the 9 LAEs candidate with lower limits, located outside of our selection function. The tabulated values of $\langle QG \rangle$, $\langle QR \rangle$, and χ are shown in Table 6.4.

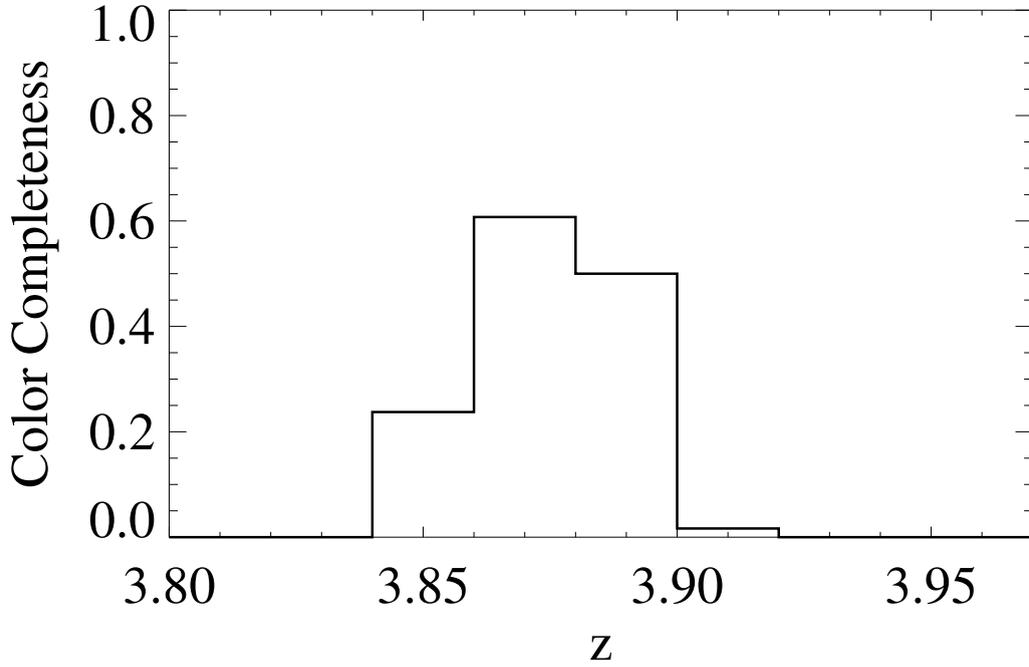


FIGURE 6.8: Completeness of the LAEs selection. The completeness was determined from 1000 simulated LAE spectra with different EW and, continuum power law and magnitudes. This is calculated by computing the fraction of the simulated spectra, per redshift bin, which was selected by the selection region.

TABLE 6.4: QSO-LAE Cross-Correlation Function.

| R(h^{-1} cMpc) | $\langle QG \rangle$ | $\langle QR \rangle$ | χ |
|-------------------|----------------------|----------------------|----------------------------|
| 0.294 | 1 | 0.179 | $4.593^{+12.864}_{-4.625}$ |
| 0.600 | 1 | 0.745 | $0.342^{+3.086}_{-1.110}$ |
| 1.225 | 1 | 3.017 | $-0.669^{+0.762}_{-0.274}$ |
| 2.502 | 4 | 12.183 | $-0.672^{+0.260}_{-0.157}$ |
| 5.107 | 19 | 29.353 | $-0.353^{+0.185}_{-0.147}$ |

As it is seen from Fig. 6.9 we have got a QSO-LAE cross-correlation measurement which is lower than the expected value. Indeed, for the last 3 bins we obtained negative values for χ , which means that the number of LAEs in QSO environments is slightly lower than the number of LAEs expected in random fields.

We explored this result, by computing the cumulative number density of LAEs in our 14 stacked fields (including and not including the 9 candidates outside of our selection region) and comparing it with the number density of LAEs measured by Ouchi et al. (2008) in random fields. We show our results in Fig. 6.10. The black points are the number density in random fields, and the red points correspond to our measurement

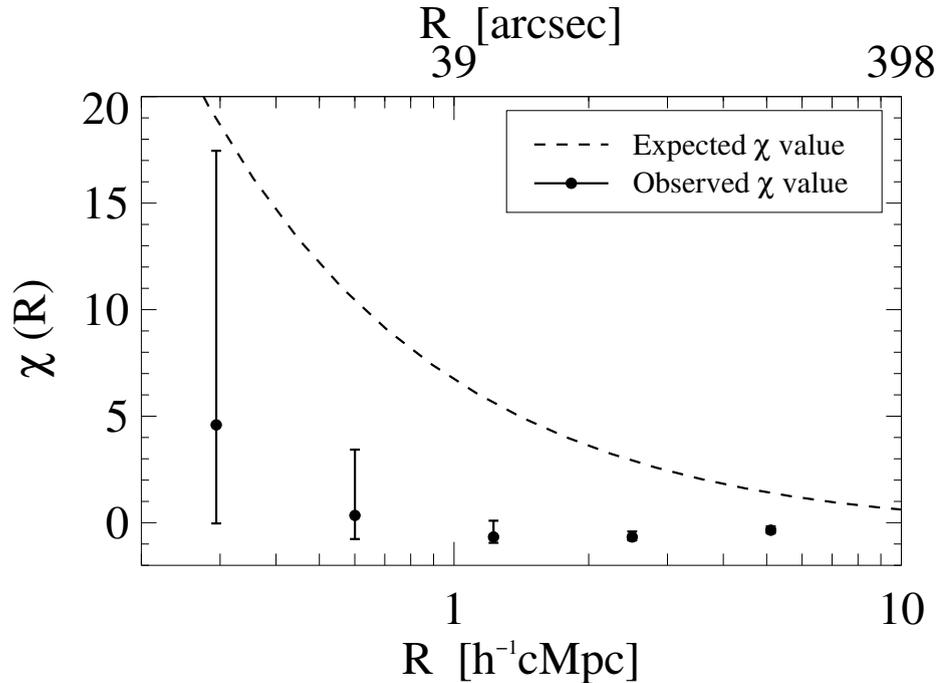


FIGURE 6.9: QSO-LAE cross-correlation function. The filled circles are showing our measurement with 1σ Poisson error bars. The dashed black line shows the theoretical expectation of χ for our 14 stacked fields calculated from the QSO and LAEs auto-correlation functions.

without the inclusion of the 9 LAEs. If we include them, we measure the same number density in the first 3 bins, but this variate slightly for the last 2 bins, which is plotted as green points (they were 0.1 mag shifted for clarity). We overplotted as a vertical red line the value corresponding to the mean magnitude of our 14 fields, so, the last bin is not complete for all our fields. The 9 extra LAEs correspond to faint LAEs which are not impacting importantly our results. Regardless of whether we include them or not, our results are consistent the number density in random fields, which suggest that QSOs environments are not especially overdense regions.

Although this result is theoretical unexpected, this is in agreement with some previous results of LAEs searches in QSOs environments at $z \sim 6$ (Bañados et al., 2013; Mazucchelli et al., 2015 submitted). In their cases, only 1 QSO field was studied, then the no-detection of an enhancement of LAEs, could have been explained by the low number statistics. When only one field is targeted some effect related with cosmic variance can be also affecting the results. In the case of our study, if we analyze our individual fields, we find 4 fields with none LAE, and others with several LAEs. In particular the most overdense field (SDSSJ0119) has 6 LAEs, corresponding to a numerical density of 0.135 LAEs/arcmin² (up to magnitude $HeI = 25.09$) which is roughly ~ 3 times higher than the expected number density in random fields. When we stack the signal of our 14

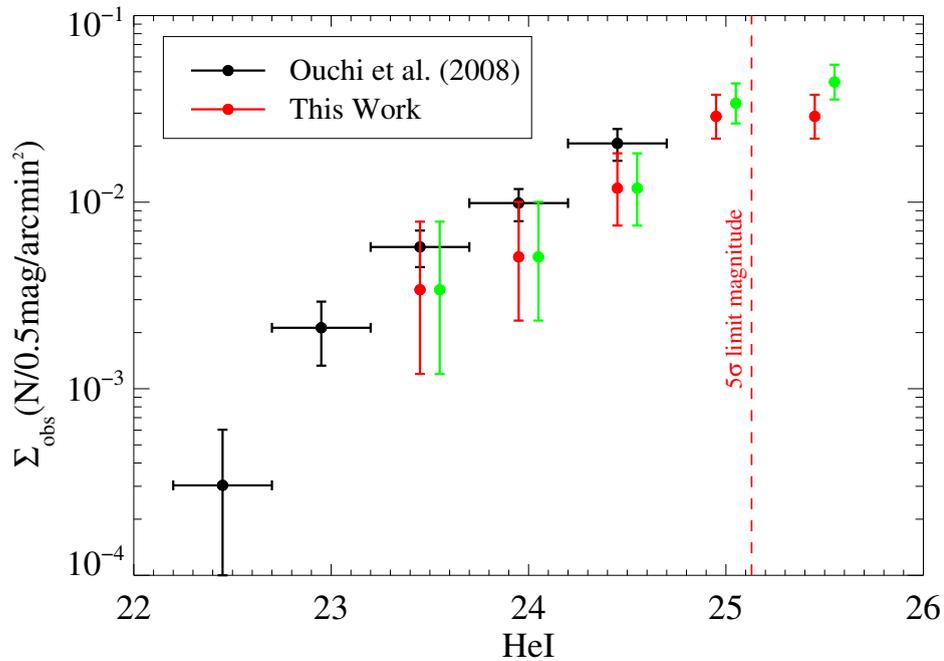


FIGURE 6.10: Cumulative LAEs number density for our 14 stacked fields, without the inclusion of the 9 LAEs located outside of our selection region (red points). If we include those extra 9 LAEs, they only influence the measurements in the last 2 bins, which is plotted as green points, shifted 0.1 mag for clarity. The black points are the LAEs number density in random fields (Ouchi et al., 2008). The vertical line is indicating the mean 5σ limit magnitude of our fields.

QSO fields, the cosmic variance effect is diluted, as well as the differences in the number density of each field, then we avoid misinterpretation of the results.

There are some possible explanations for our non-overdensity detection. A first one is that the QSO auto-correlation length at $z \sim 4$ that we used to compute the expected clustering in QSO environments is overestimated. Indeed, some discussion has been done about this topic. The QSO clustering at $z \sim 4$ measured by Shen et al. (2007) (and used in this thesis) implies dark matter halos masses of $> 8 \times 10^{12} M_{\odot}$. However, some simulations (with QSO feedback included in the model) suggest effective dark matter halos of $\sim 1.5 \times 10^{12} M_{\odot}$ (Fanidakis et al., 2013), implying that the clustering of QSOs at $z \sim 4$ could be overestimated. However, note that in that case, we expect that QSO reside in dark matter halos with slightly lower masses, but we still expect a small enhancement of LAEs in QSOs fields over the standard value. We did not detect that enhancement, then we also have to explore other explanations.

Another possibility is that LAEs could be avoiding QSO environments, at least on scales of $\lesssim 10 \text{Mpc} h^{-1}$. Some simulations have suggested an extended overdensity in QSOs environments at high redshift, ranging from few to several tens arcmin, and showing

variation in the number density across the sky (Muldrew et al., 2015; Overzier et al., 2009). Indeed, some people have found overdensities of galaxies in QSOs environments when they study their environments in large scales of $3000\text{-}4000\text{cMpc}^2$ (Morselli et al., 2014; Utsumi et al., 2010). It is necessary to perform an study in wider field-of-view in order to explore this possibility.

Some authors have also suggested that QSO feedback could be suppressing the star formation in the vicinity of the QSO (e.g Bruns et al., 2012; Francis and Bland-Hawthorn, 2004). Studies of the ionized regions around $z \sim 6$ quasars have suggested that this affects at scales of $\sim 1.5\text{-}5.2\text{pMpc}$ (Venemans et al., 2015). In this case we should detect a reduced number of galaxies in this fields. However, other authors claim the opposite effect, suggesting that a luminous QSO could enhance the star formation. In particular, the UV radiation from the QSO could illuminate the neutral gas of nearby galaxies allowing them to fluorescently emit $\text{Ly}\alpha$ photons (Borisova et al., 2015; Cantalupo et al., 2012). It has been found that possibly this effect could be affecting up to projected distances of $\lesssim 15\text{pMpc}$ around $z \sim 3$ QSOs (Borisova et al., 2015). Note that both mentioned physical processes could affect LAEs differently than LBGs, and then it could be also responsible for the disagreement in the results of LAEs overdensities and LBGs overdensities in QSO fields. A much more deep study of the physical effect of QSOs over nearby galaxies is needed to clarify this picture.

Chapter 7

Summary and Conclusions

Luminous quasars at $z \gtrsim 4$ are thought to inhabit the most massive dark matter halos in the early Universe. This would imply a higher galaxy number density in QSO environments in comparison with what we might expect in a random location in the Universe. Most of the studies have been focused in the search of LAEs or LBGs overdensities in the environs of the most distant QSOs ($z \gtrsim 5$), which has shown a disagreement in the results: some studies claim galaxy overdensities in QSO environments, and others find a galaxy numerical density consistent with random fields.

One deficiency of those studies is that most of them are based in LBGs samples selected over a broad redshift range ($\Delta z \sim 1$), which introduce line of sight contaminants to the number counts. Additionally, the lack of a well known galaxy luminosity function and clustering properties of both QSO and galaxies at $z \gtrsim 5$, prevent a good interpretation and analysis of the results in terms of clustering models.

In this thesis I studied $z \sim 4$ QSOs environments, where QSOs clustering predict dark matter halos masses of $M_{\text{halo}} \sim 10^{13} M_{\odot}$, which implies a strong galaxy-QSO cross-correlation function. We targeted QSOs in a narrow redshift range in order to measure, for the first time, the QSO-galaxy cross-correlation function at $z \sim 4$. In that way we did not only search for overdensities in individual fields as it is done in the other studies, but we stacked the galaxy counts of every QSO field to measure their clustering properties, encapsulated in the parameter r_0 and γ . These represent the amplitude and slope of a clustering described by a power law function.

Firstly, we used data from the ESO archive (Program ID: 079.A-0644) to study the environs of 6 luminous QSOs at $z = 3.78 \pm 0.04$, observed with the FORS1 instrument on the VLT in 2007. The QSO fields were observed using a set of two narrow band filters

designed specifically for this program and the broad band R in order to select LBGs with a novel technique which selects them in a narrow redshift range ($\Delta z \sim 0.2$).

Since we used a non-standard filter set to select LBGs, there were not available studies about the criteria to select them, and their redshift selection function was completely unknown. For that reason, we performed a Monte Carlo simulation to model the LBGs colors, as well as to measure the completeness of our sample, which is an essential quantity needed for clustering computation.

We obtained that the $z = 3.78$ LBGs colors are quite similar to the low- z galaxies colors, so they cannot be isolated efficiently. This is in part due to the fact that the narrow bands are too sensitive to the features in the object spectra, which makes them to have extreme colors for intense absorption/emission lines, which are similar to the excess color expected in LBGs due to the Lyman break. Considering this scenario, we chose a selection region where only the 20% of LBGs at $z = 3.78$ are selected, but it was the only way to ensure a non highly contaminated sample. Considering this selection function, we detected a total of 73 LBGs in our 6 stacked fields (which correspond to a 0.31 LBGs/arcmin²).

We measured the QSO-LBG volume averaged cross-correlation function $\chi(R)$, by counting the LBGs around the QSO in radial bins and comparing it with the theoretically expected LBGs counts in random fields. This expectation was computed using the LBG luminosity function at $z \sim 4$ and the redshift selection function computed using our Monte Carlo simulation. Our results show an overdensity of LBGs in QSO fields, in agreement with the expected overdensity computed from the individual QSO and LBGs auto-correlation function at $z \sim 4$ and assuming a linear bias model. We fitted our results using a maximum likelihood estimator to constrain the real space parameters r_0 and γ . We obtained a clustering amplitude of $r_0 = 9.91_{-1.79}^{+3.28} h^{-1}$ Mpc and a slope $\gamma = 2.05_{-0.46}^{+0.20}$.

Complementary to this results, we also measured the LBGs auto-correlation function in our fields, which is expected to be stronger than in random fields. The clustering amplitude measured in random fields is $r_0 = 4.1 h^{-1}$ cMpc for a fixed $\gamma = 1.8$, and our data showed a clustering amplitude of $r_0 = 21.13_{-14.78}^{+39.44} h^{-1}$ cMpc, confirming that LBGs in QSOs fields are more clustered than LBGs in random fields.

The novel technique used in this work was efficient to select LBGs in a narrow redshift range, but it showed to be highly sensitive to the features in the spectra, which made it impossible to select LBGs with high completeness, and high purity at the same time. This implies that our LBGs sample is only 20% complete and it could be slightly contaminated, but based on the clustering measurements, we argued that this is not highly

contaminated. Additionally, our results are sensitive to the redshift selection function computed by our simulations, then our results should be carefully considered.

One of the main deficiencies of using this new filter set is the lack of a way to measure contamination in our LBGs sample and the uncertainties in the redshift selection model. The first problem could be solved by doing spectroscopy to the LBGs sample. The second problem could be solved by observing random fields with the same filter configuration, in order to compute the mean number density of LBGs in random fields, from the data itself instead of using a theoretical model.

Additionally, we measured the QSO-LAE cross-correlation function from a sample of 14 QSO fields at $z = 3.87 \pm 0.02$, which represent the largest sample available for the search of LAEs in QSO environments at $z \gtrsim 4$. This data was taken using the FORS2 instrument on the VLT in 2015. In this case, the LAEs selection function is known, and they can be well isolated from low- z galaxies. We selected them using the same criteria as a previous study in which the luminosity function was measured. In that way we ensured that our sample selection had the same completeness and contamination as that study, and then a color modeling is not necessary in this case. Our LAEs sample is much more complete ($\sim 60\%$) and cleaner (contamination fraction $\sim 0\% - 14\%$) than the LBGs sample. We selected 26 LAEs with $EW > 44\text{\AA}$ in our 14 fields, which correspond to a number density of 4.41×10^{-2} LAEs/arcmin².

We measured the cross-correlation function in the same way as for LBGs, and we obtained that QSOs and LAEs are not correlated. We also found that our mean LAEs number density is consistent with the one expected in random fields, then an enhancement of LAEs is not found in QSO environments. This could mean either that the QSO auto-correlation length at $z \sim 4$ is overestimated, or that LAEs preferentially avoid QSO environments on $\lesssim 10 \text{ Mpc h}^{-1}$ scales.

7.1 Future Work

The results of this thesis place the first observational constraints in the structure formation at $z \sim 4$. From the second set of data mentioned above, we achieved to build a complete and clean LAEs sample, which allowed us to conclude that the QSO environments do not show an overdensity of galaxies as opposed to what is expected from theoretical predictions. This is in agreement with previous searches of LAEs in 2 individual QSO fields at $z \sim 6$. This result opens important questions, and could require a re-thinking of galaxy formation in dense environments.

One possibility that arises to explain this result is that QSOs are tracers of massive dark matter halos, but for some reason, galaxies are avoiding the immediate QSO environment, and the overdensity could be present in greater scales. The way to explore this scenario, is doing the imaging of QSO fields in much greater scales. Some authors have pointed this idea, finding overdensities in large scales. For example, Morselli et al. (2014) studied 4 QSO fields at $z \sim 6$ in a field of view of $\sim 3000 \text{cMpc}^2$, reporting overdensities in all the studied fields. Utsumi et al. (2010) reported overdensities of LBGs in one $z \sim 6$ QSO field, imagined in a field of view of $\sim 4600 \text{cMpc}^2$. The best way to explore this possibility is to do the imaging of a large sample of QSOs with large field of view in order to search for LAEs in their fields. The search of LAEs allows a cleaner sample, because the galaxy selection is over a much narrow redshift slice in comparison with LBGs searches. So far, only individual QSOs have been studied, but a larger sample of QSOs could put stronger conclusions. Because of this, it would be possible to measure a correlation function instead of only measure overdensity in individual fields. Additionally, it would avoid misinterpretation due to the cosmic variance, since this effect is diluted when the sample is larger.

The results that we obtained from the other set of data in this thesis are pointing to the opposite direction. We find an strong QSO-LBG cross-correlation at $z \sim 4$. However, because of the incompleteness and the possible contamination of the sample, this result has to be confirmed. The only way to do that is doing spectroscopy over our LBGs candidates. If the results presented in this thesis are confirmed, then important questions arise, and the differences of the LAEs and LBGs properties could be somehow affecting the disagreement in the number density of them in QSOs environments (e.g Kashikawa et al., 2007).

A complementary way to put observational constrains on the structure formation process is to use other objects as tracers of massive dark matter halos at high redshift. Some of the tracers that have been suggested are Ly α blobs (e.g Matsuda et al., 2004; Steidel et al., 2000; Yang et al., 2009), hot dust obscured galaxies (Assef et al., 2015) and submillimeter galaxies (Aravena et al., 2010). Another tracer that appears strongly are the binary quasars. Hennawi et al. (2010) presented the first sample of 27 binary QSOs at $z \sim 3-4$, which were selected in the SDSS imaging data. These structures are extremely rare, have a density of one per 10 Gpc^3 at $z \sim 4$ (Hennawi et al., 2010), then if our current understanding of the structure formation is right, we would expect to discover the most massive protoclusters yet known, and measure a huge cross-correlation between binary QSOs and galaxies. I am currently involved in a project to search for LBGs in 8 binary QSOs environments, for which we already have available data. The results of this project could also put strong constrains in the structure formation at high redshift.

Finally, to the extent that the luminosity function of galaxies at higher redshift $z \gtrsim 5$ be well studied, the natural next step is to push this study to higher redshifts, where the galaxy-QSO cross-correlation should be even stronger. Some QSOs survey (e.g Pan-STARRS¹) are currently enlarging the number of known QSOs in the Universe (Bañados et al., 2014), which will allow us to increment the statistic of QSOs environments studies at those high redshift.

¹(Kaiser et al., 2002, 2010)

Appendix A

LBGs Sample

In Fig. A.1 to A.7 I show the images of each LBG in the sample. From left to right I show the NB₅₇₁, NB₅₉₆ and r_{GUNN} images. Each panel is 15'' on a side. The red circle show the position of the detected object, and the size correspond to the region in which the photometry was done (2'' in diameter). The magnitudes are indicated in each panel.

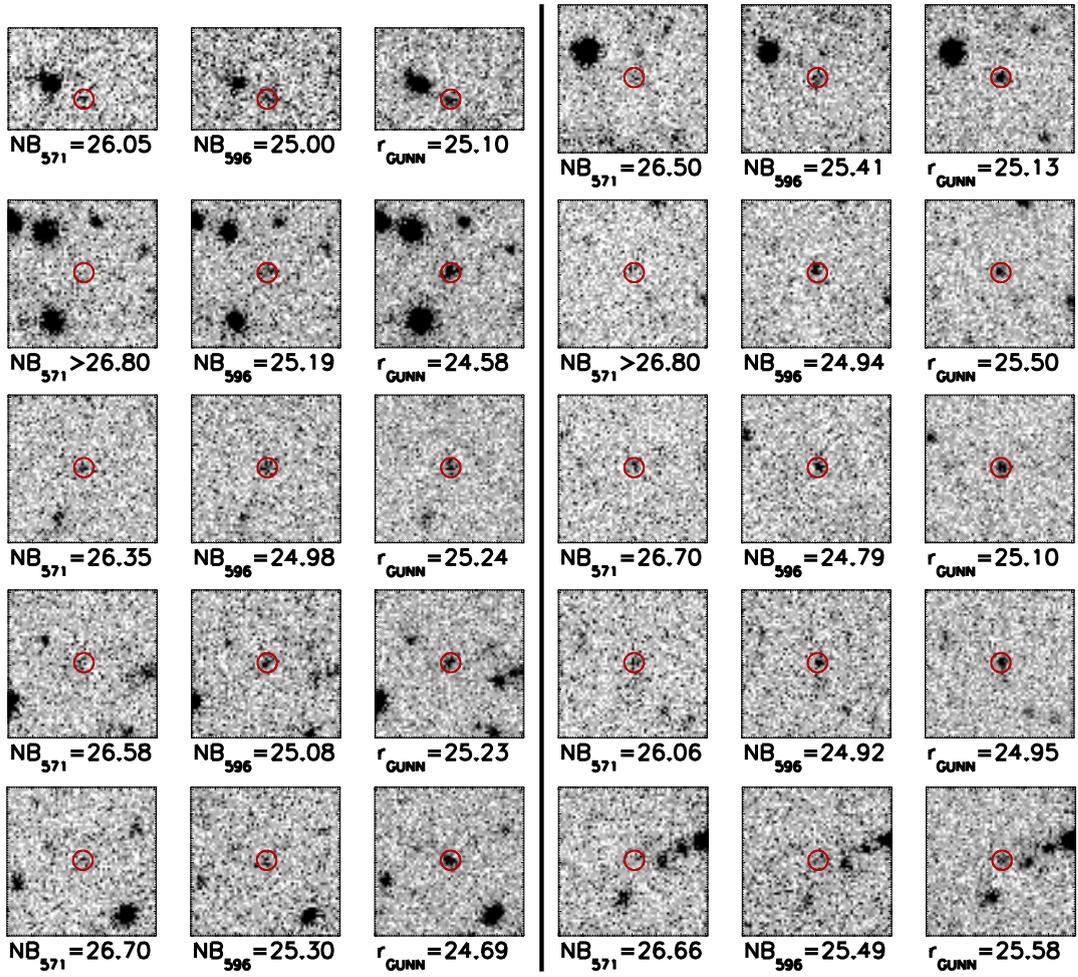


FIGURE A.1: Images of selected LBGs for 0124+0044.

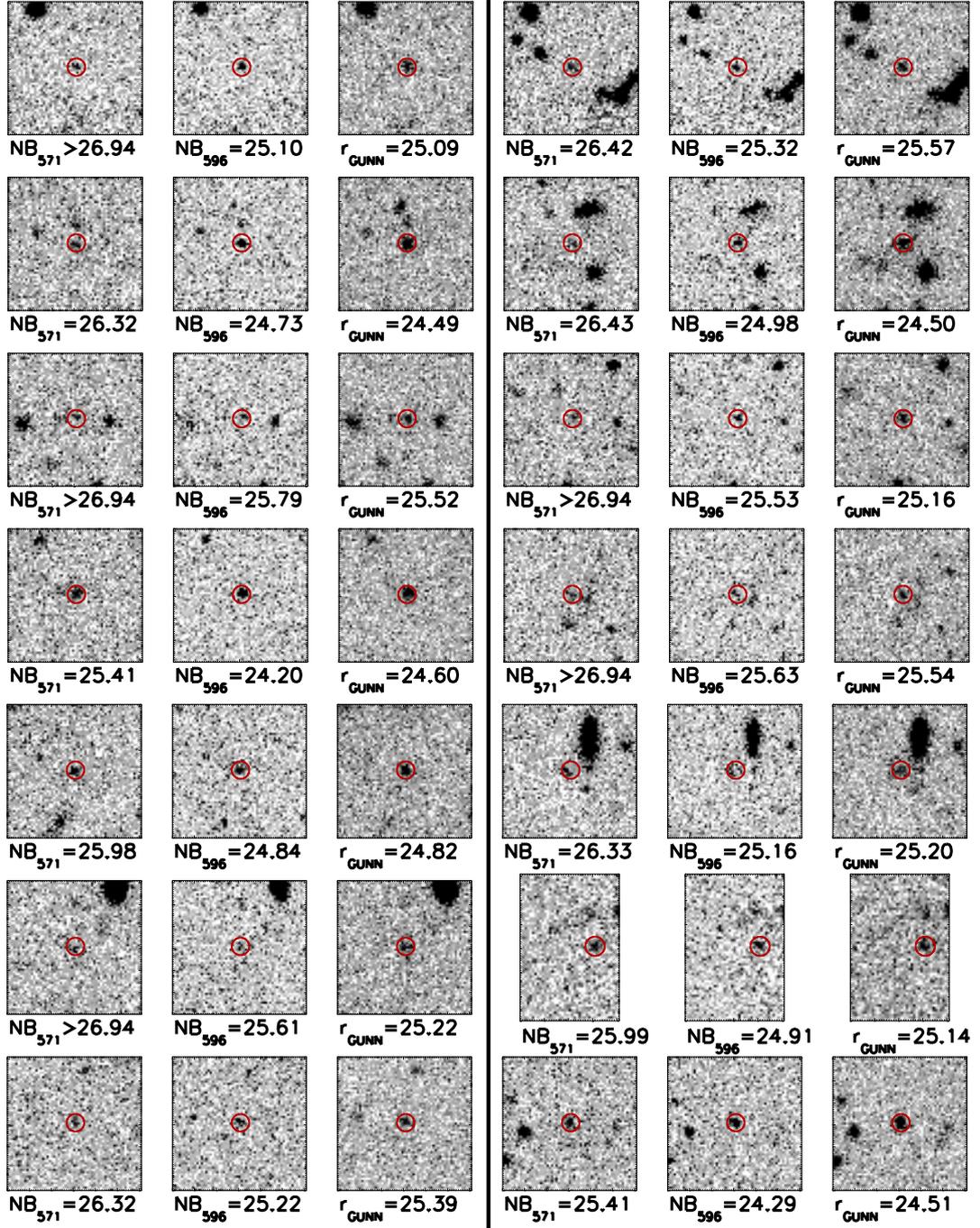


FIGURE A.2: Images of selected LBGs for 0213-0904 (continues).

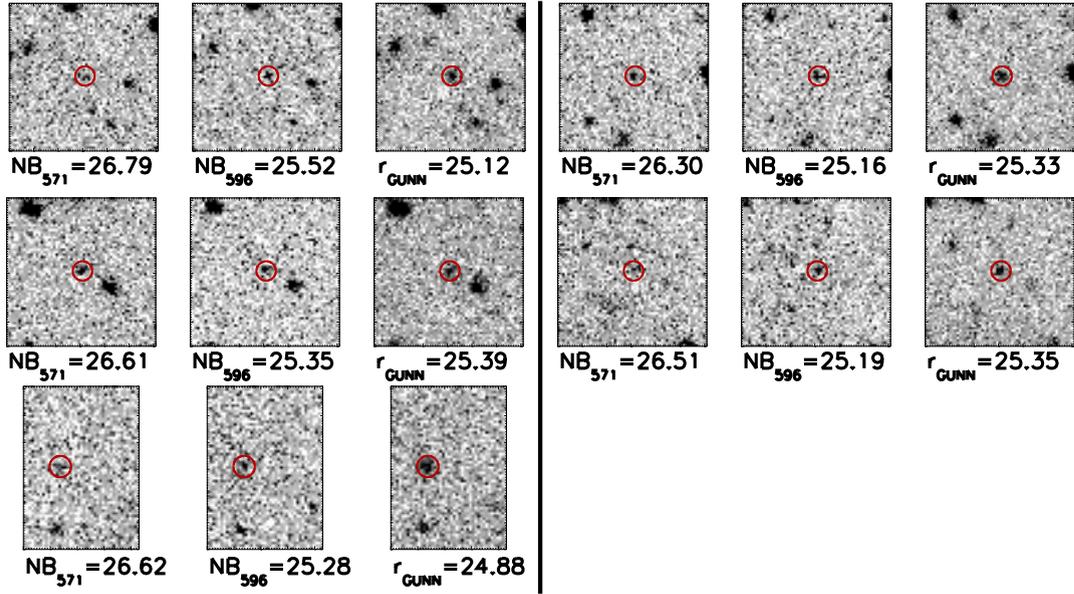


FIGURE A.3: Images of selected LBGs for 0213-0904 (continuation).

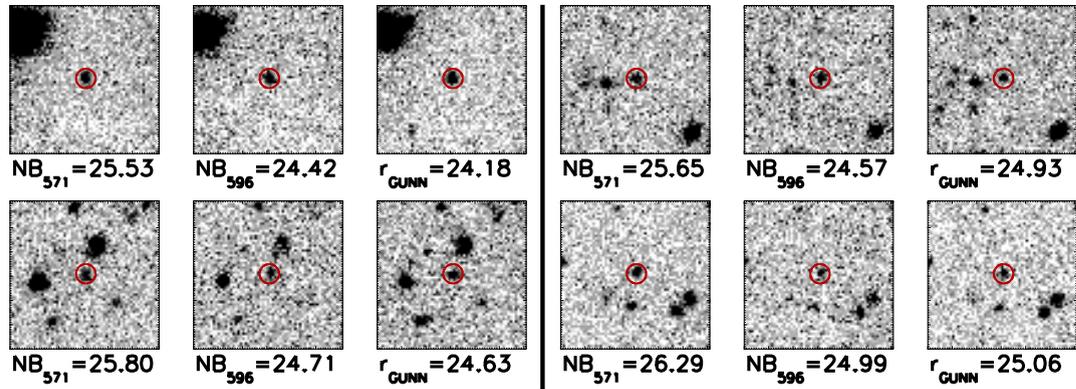


FIGURE A.4: Images of selected LBGs for 2003-33.

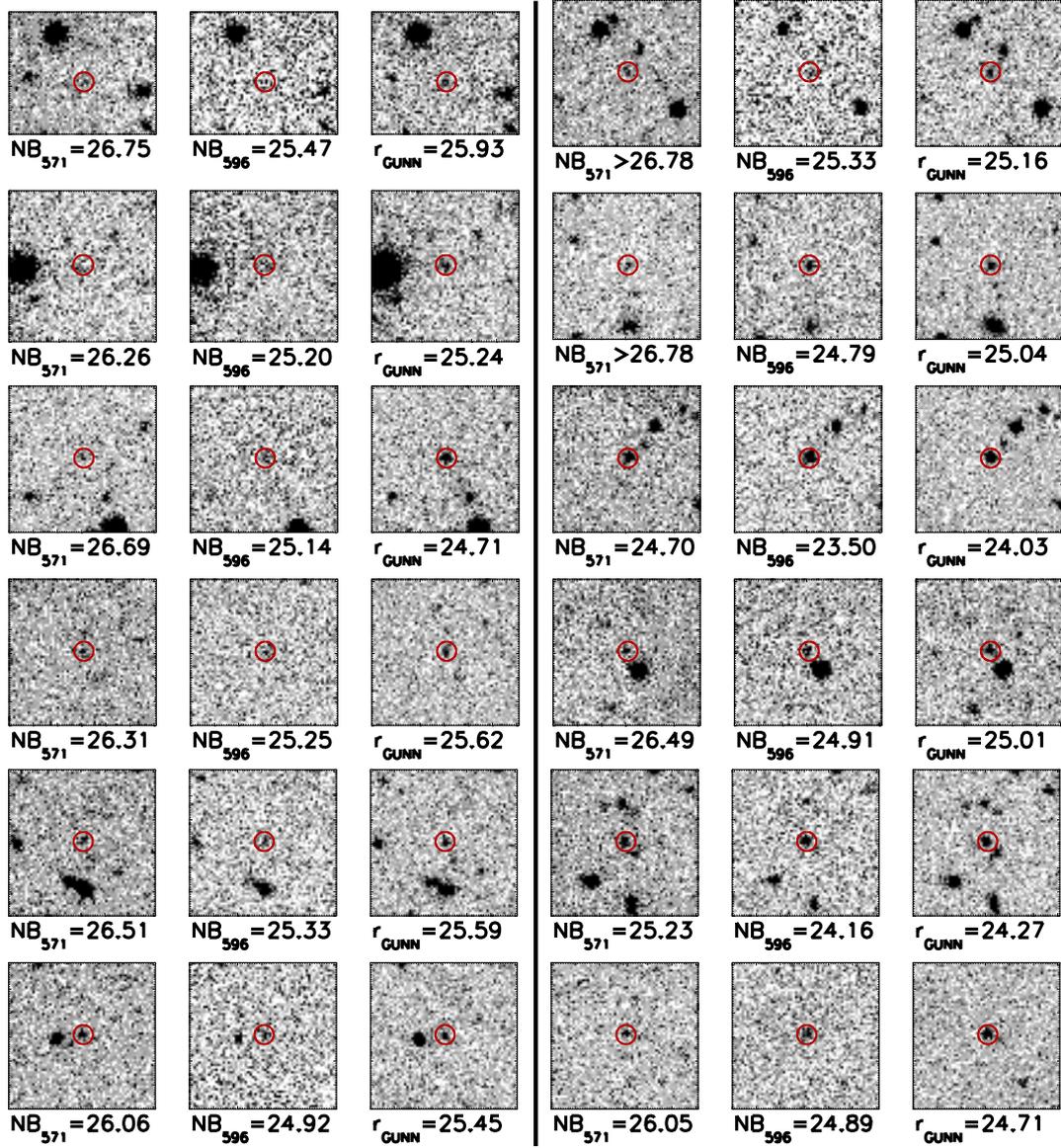


FIGURE A.5: Images of selected LBGs for 2207+0043.

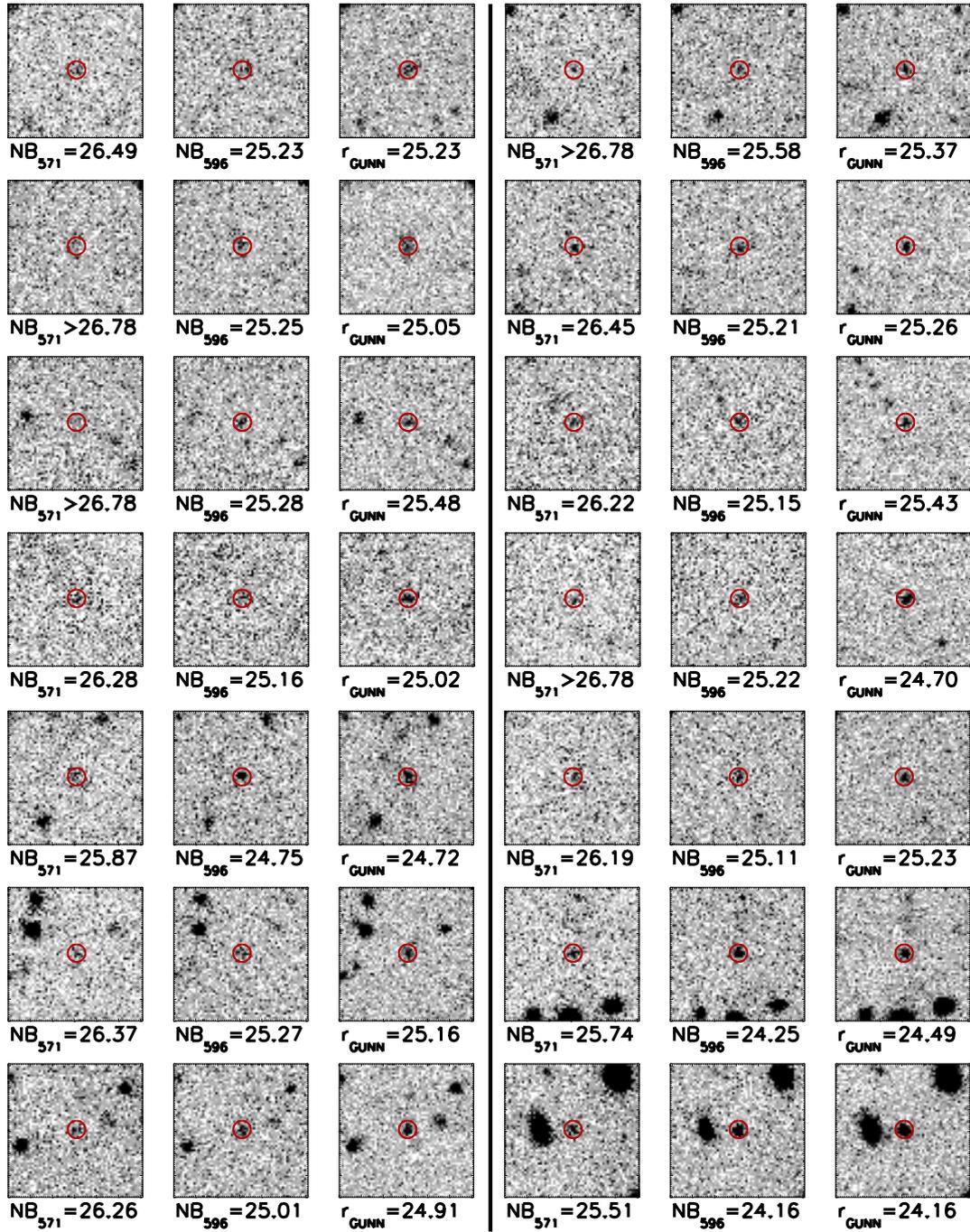


FIGURE A.6: Images of selected LBGs for 2301+0112.

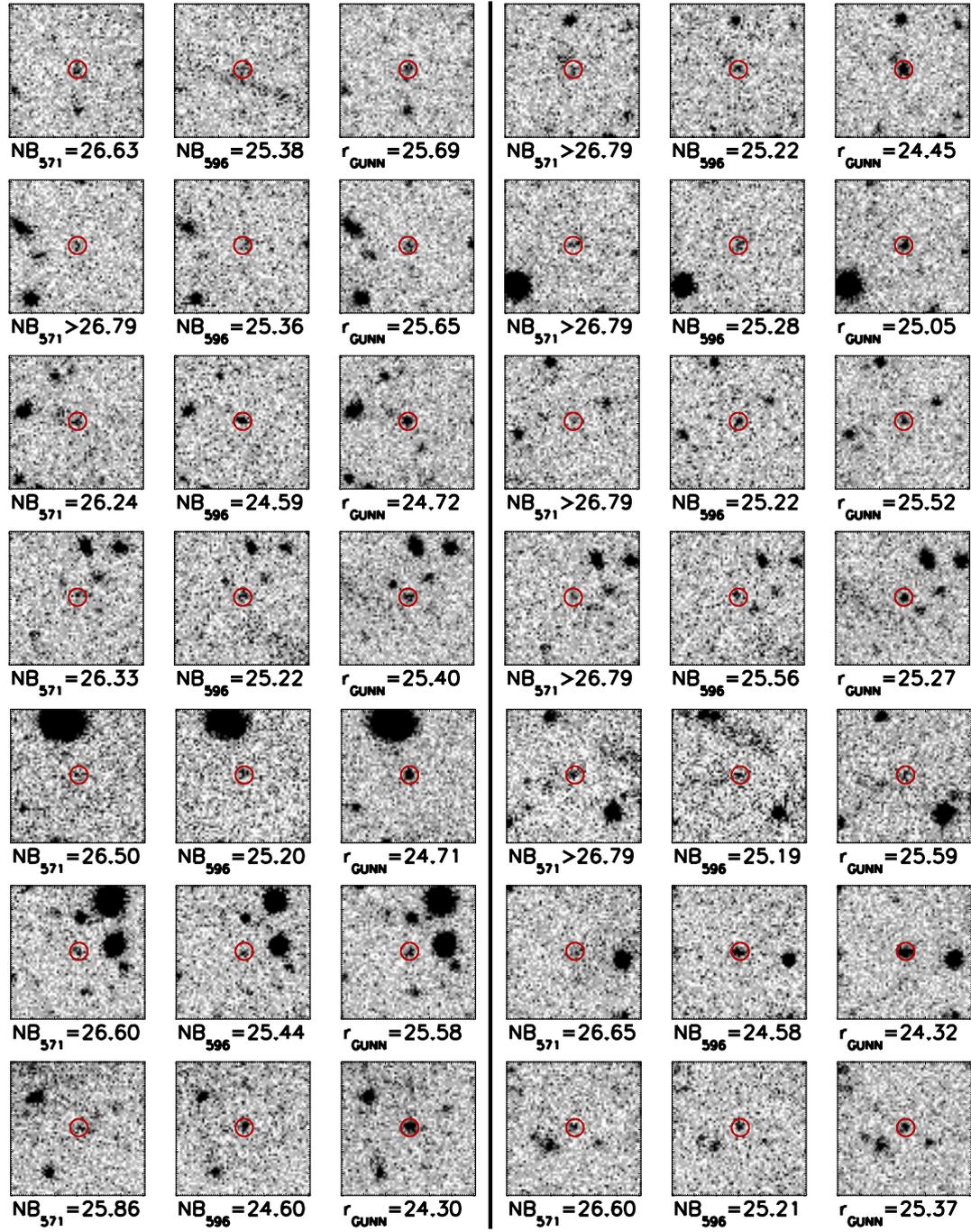


FIGURE A.7: Images of selected LBGs for 2311-0844.

Appendix B

LAEs Sample

In Fig. B.1 to B.10 I show the images of each LAE in the sample. From left to right I show the g , HeI and R images. Each panel is $15''$ on a side. The red circle show the position of the detected object, and the size correspond to the region in which the photometry was done ($2''$ in diameter). The magnitudes are indicated in each panel.

The color-color diagram for each individual field is presented in Fig. B.11 and B.12. The evolutionary track of LAEs is plotted as redshift color-coded track. The magenta point indicate the color of the QSO in our filters. Magnitudes of cases in which the object was not detected in g or R filter at 2σ level, were replaced by the corresponding limit magnitude. Those cases are shown with arrows. Cases with no detection on either filter are shown with both arrows. The dashed line is indicating the selection region used to select LAEs. At the top right of each plot the number of found LAEs inside of the selection region is shown.

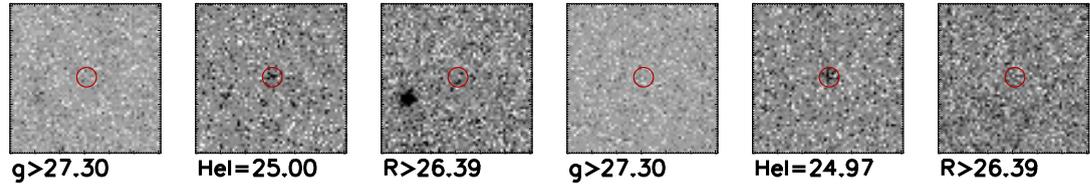


FIGURE B.1: Images of selected LAEs for SDSSJ0040.

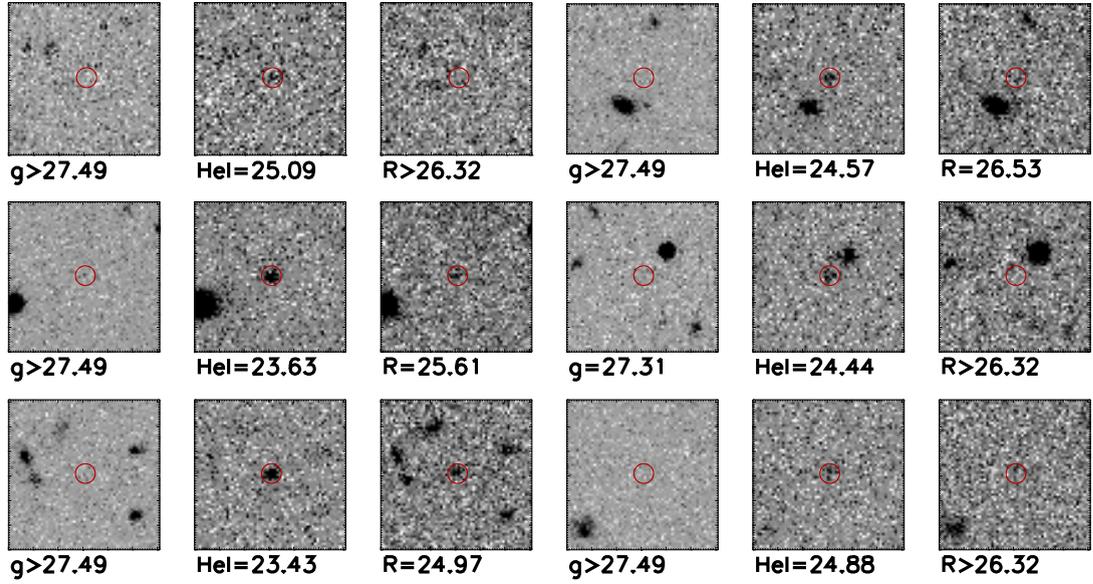


FIGURE B.2: Images of selected LAEs for SDSSJ0119.

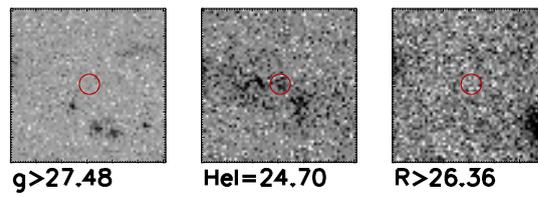


FIGURE B.3: Images of selected LAEs for SDSSJ0149.

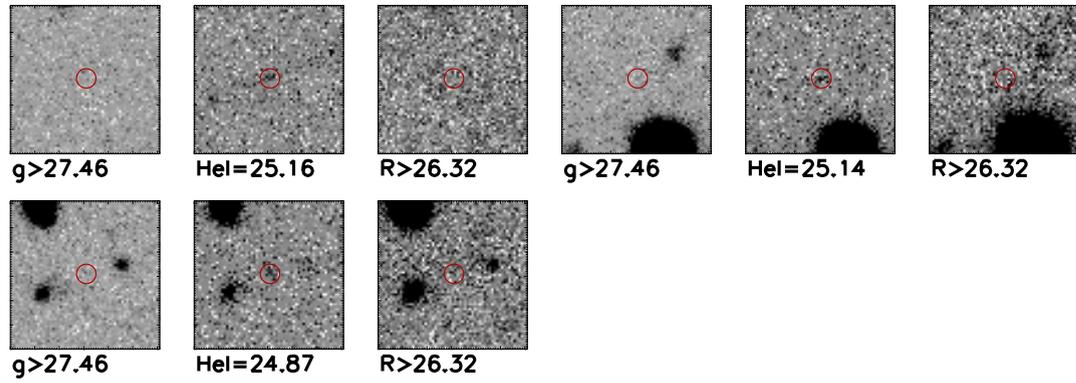


FIGURE B.4: Images of selected LAEs for SDSSJ1026.

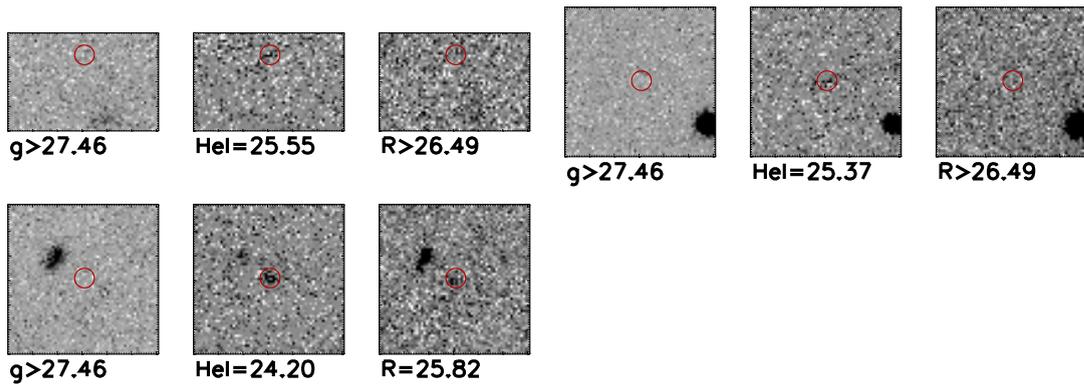


FIGURE B.5: Images of selected LAEs for SDSSJ1044.

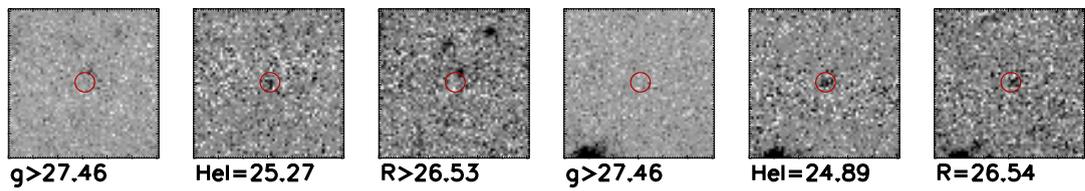


FIGURE B.6: Images of selected LAEs for SDSSJ1205.

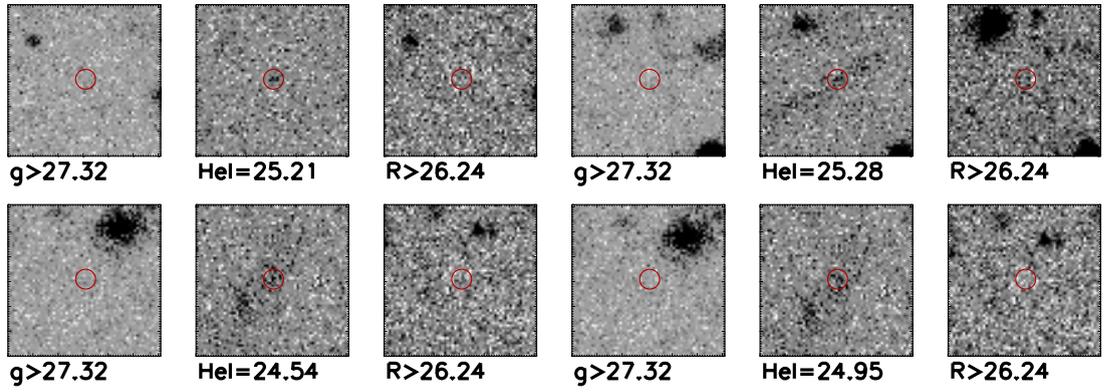


FIGURE B.7: Images of selected LAEs for SDSSJ1224.

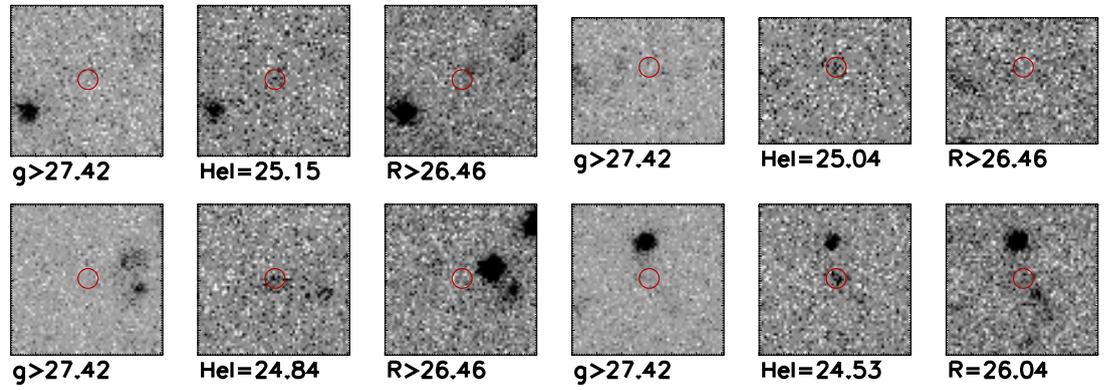


FIGURE B.8: Images of selected LAEs for SDSSJ1258.

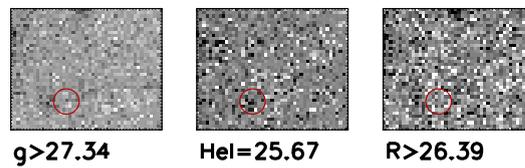


FIGURE B.9: Images of selected LAEs for SDSSJ2250.

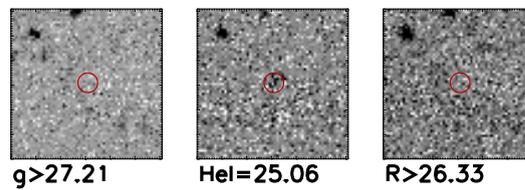


FIGURE B.10: Images of selected LAEs for SDSSJ2350.

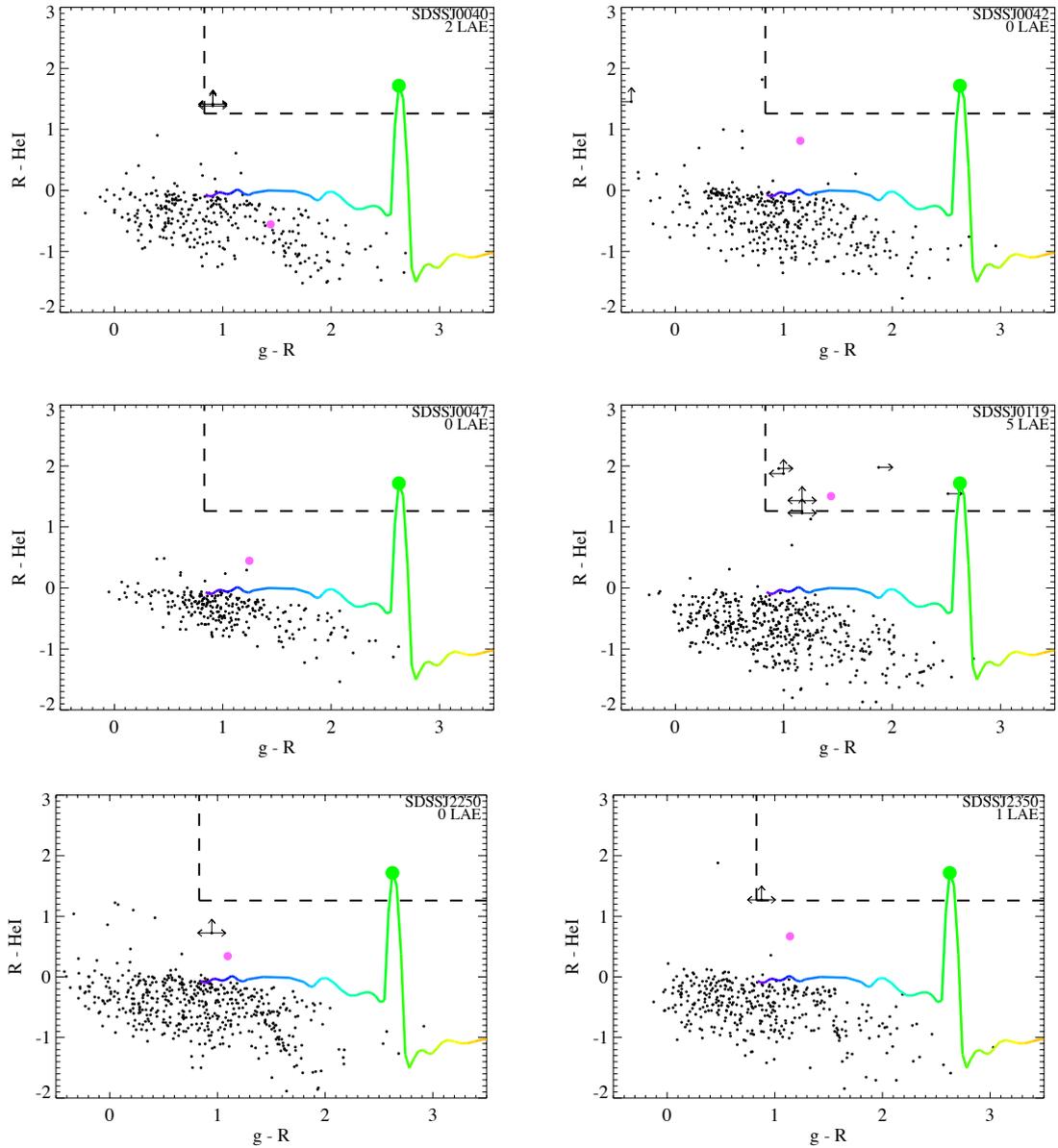


FIGURE B.11: Color-color diagram for our 14 individual QSO fields.

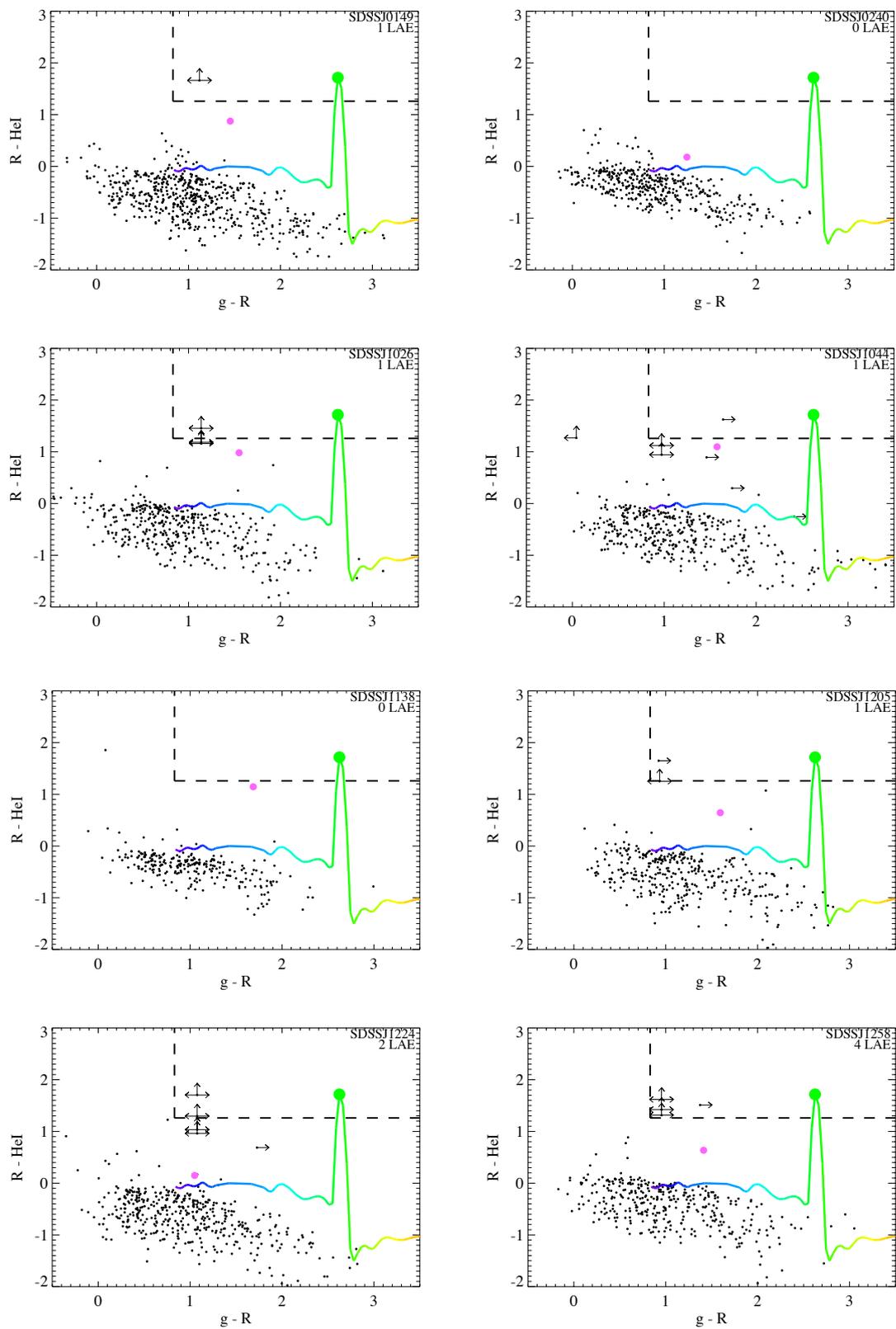


FIGURE B.12: Continuation

Bibliography

- Adams, S. M., Martini, P., Croxall, K. V., Overzier, R. A., and Silverman, J. D. (2015). Discovery of an overdensity of Lyman alpha emitters around a $z \sim 4$ QSO with the Large Binocular Telescope. *Monthly Notices of the Royal Astronomical Society*, 448:1335–1344.
- Ando, M., Ohta, K., Iwata, I., Akiyama, M., Aoki, K., and Tamura, N. (2006). Deficiency of Large Equivalent Width Ly α Emission in Luminous Lyman Break Galaxies at $z \sim 5$ -6? *The Astrophysical Journal Letters*, 645:L9–L12.
- Appenzeller, I. and Rupprecht, G. (1992). FORS - the focal reducer for the VLT. *The Messenger*, 67:18–21.
- Aravena, M., Bertoldi, F., Carilli, C., Schinnerer, E., McCracken, H. J., Salvato, M., Riechers, D., Sheth, K., Smölcic, V., Capak, P., Koekemoer, A. M., and Menten, K. M. (2010). Environment of MAMBO Galaxies in the COSMOS Field. *The Astrophysical Journal Letters*, 708:L36–L41.
- Assef, R. J., Eisenhardt, P. R. M., Stern, D., Tsai, C.-W., Wu, J., Wylezalek, D., Blain, A. W., Bridge, C. R., Donoso, E., Gonzales, A., Griffith, R. L., and Jarrett, T. H. (2015). Half of the Most Luminous Quasars May Be Obscured: Investigating the Nature of WISE-Selected Hot Dust-Obscured Galaxies. *The Astrophysical Journal*, 804:27.
- Bañados, E., Venemans, B., Walter, F., Kurk, J., Overzier, R., and Ouchi, M. (2013). The Galaxy Environment of a QSO at $z \sim 5.7$. *The Astrophysical Journal*, 773:178.
- Bañados, E., Venemans, B. P., Morganson, E., Decarli, R., Walter, F., Chambers, K. C., Rix, H.-W., Farina, E. P., Fan, X., Jiang, L., McGreer, I., De Rosa, G., Simcoe, R., Weiß, A., Price, P. A., Morgan, J. S., Burgett, W. S., Greiner, J., Kaiser, N., Kudritzki, R.-P., Magnier, E. A., Metcalfe, N., Stubbs, C. W., Sweeney, W., Tonry, J. L., Wainscoat, R. J., and Waters, C. (2014). Discovery of Eight $z \sim 6$ Quasars from Pan-STARRS1. *The Astronomical Journal*, 148:14.
- Bahcall, J. N., Kirhakos, S., Saxe, D. H., and Schneider, D. P. (1997). Hubble Space Telescope Images of a Sample of 20 Nearby Luminous Quasars. *The Astrophysical Journal*, 479:642–658.
- Bahcall, N. A. and Cen, R. (1992). Galaxy clusters and cold dark matter - A low-density unbiased universe? *The Astrophysical Journal Letters*, 398:L81–L84.
- Bahcall, N. A. and Fan, X. (1998). The Most Massive Distant Clusters: Determining Ω and σ_8 . *The Astrophysical Journal*, 504:1–6.
- Bertin, E. (2006). Automatic Astrometric and Photometric Calibration with SCAMP. In Gabriel, C., Arviset, C., Ponz, D., and Enrique, S., editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, page 112.
- Bertin, E. and Arnouts, S. (1996). SExtractor: Software for source extraction. *Astronomy and Astrophysics Supplement*, 117:393–404.

- Bertin, E., Mellier, Y., Radovich, M., Missonnier, G., Didelon, P., and Morin, B. (2002). The TERAPIX Pipeline. In Bohlender, D. A., Durand, D., and Handley, T. H., editors, *Astronomical Data Analysis Software and Systems XI*, volume 281 of *Astronomical Society of the Pacific Conference Series*, page 228.
- Borisova, E., Lilly, S. J., Cantalupo, S., Prochaska, J. X., Rakic, O., and Worseck, G. (2015). Constraining the lifetime and opening angle of quasars using fluorescent Ly α emission: the case of Q0420-388. *ArXiv e-prints*.
- Bouwens, R. J., Illingworth, G. D., Franx, M., Chary, R.-R., Meurer, G. R., Conselice, C. J., Ford, H., Giavalisco, M., and van Dokkum, P. (2009). UV Continuum Slope and Dust Obscuration from $z \sim 6$ to $z \sim 2$: The Star Formation Rate Density at High Redshift. *The Astrophysical Journal*, 705:936–961.
- Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., and Lemson, G. (2009). Resolving cosmic structure formation with the Millennium-II Simulation. *Monthly Notices of the Royal Astronomical Society*, 398:1150–1164.
- Brammer, G. B., van Dokkum, P. G., and Coppi, P. (2008). EAZY: A Fast, Public Photometric Redshift Code. *The Astrophysical Journal*, 686:1503–1513.
- Brodwin, M., Gonzalez, A. H., Moustakas, L. A., Eisenhardt, P. R., Stanford, S. A., Stern, D., and Brown, M. J. I. (2007). Galaxy Cluster Correlation Function to $z \sim 1.5$ in the IRAC Shallow Cluster Survey. *The Astrophysical Journal Letters*, 671:L93–L96.
- Bruns, L. R., Wyithe, J. S. B., Bland-Hawthorn, J., and Dijkstra, M. (2012). Clustering of Ly α emitters around luminous quasars at $z=2-3$: an alternative probe of reionization on galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 421:2543–2552.
- Bruzual, G. and Charlot, S. (2003). Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*, 344:1000–1028.
- Calzetti, D. (1997). UV opacity in nearby galaxies and application to distant galaxies. In Waller, W. H., editor, *American Institute of Physics Conference Series*, volume 408 of *American Institute of Physics Conference Series*, pages 403–412.
- Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., and Storchi-Bergmann, T. (2000). The Dust Content and Opacity of Actively Star-forming Galaxies. *The Astrophysical Journal*, 533:682–695.
- Cantalupo, S., Lilly, S. J., and Haehnelt, M. G. (2012). Detection of dark galaxies and circumgalactic filaments fluorescently illuminated by a quasar at $z = 2.4$. *Monthly Notices of the Royal Astronomical Society*, 425:1992–2014.
- Cardelli, J. A., Clayton, G. C., and Mathis, J. S. (1989). The relationship between infrared, optical, and ultraviolet extinction. *The Astrophysical Journal*, 345:245–256.
- Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. *Publications of the Astronomical Society of the Pacific*, 115:763–795.
- Ciardullo, R., Gronwall, C., Wolf, C., McCathran, E., Bond, N. A., Gawiser, E., Guaita, L., Feldmeier, J. J., Treister, E., Padilla, N., Francke, H., Matković, A., Altmann, M., and Herrera, D. (2012). The Evolution of Ly α -emitting Galaxies between $z = 2.1$ and $z = 3.1$. *The Astrophysical Journal*, 744:110.
- Coil, A. L. (2013). *The Large-Scale Structure of the Universe*, page 387.
- Coil, A. L., Georgakakis, A., Newman, J. A., Cooper, M. C., Croton, D., Davis, M., Koo, D. C., Laird, E. S., Nandra, K., Weiner, B. J., Willmer, C. N. A., and Yan, R. (2009). AEGIS: The Clustering of X-Ray Active Galactic Nucleus Relative to Galaxies at $z \sim 1$. *The Astrophysical Journal*, 701:1484–1499.

- Coil, A. L., Hennawi, J. F., Newman, J. A., Cooper, M. C., and Davis, M. (2007). The DEEP2 Galaxy Redshift Survey: Clustering of Quasars and Galaxies at $z = 1$. *The Astrophysical Journal*, 654:115–124.
- Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Deeley, K., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Price, I., Seaborne, M., and Taylor, K. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts. *Monthly Notices of the Royal Astronomical Society*, 328:1039–1063.
- Croom, S. M., Richards, G. T., Shanks, T., Boyle, B. J., Strauss, M. A., Myers, A. D., Nichol, R. C., Pimblet, K. A., Ross, N. P., Schneider, D. P., Sharp, R. G., and Wake, D. A. (2009). The 2dF-SDSS LRG and QSO survey: the QSO luminosity function at $0.4 < z < 2.6$. *Monthly Notices of the Royal Astronomical Society*, 399:1755–1772.
- Davis, M. and Peebles, P. J. E. (1983). A survey of galaxy redshifts. V - The two-point position and velocity correlations. *The Astrophysical Journal*, 267:465–482.
- De Lucia, G. and Blaizot, J. (2007). The hierarchical formation of the brightest cluster galaxies. *Monthly Notices of the Royal Astronomical Society*, 375:2–14.
- Dodelson, S. (2003). *Modern cosmology*.
- Eke, V. R., Cole, S., and Frenk, C. S. (1996). Cluster evolution as a diagnostic for Ω . *Monthly Notices of the Royal Astronomical Society*, 282.
- Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M. (2006). Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with $z \sim 6$ Quasars. II. A Sample of 19 Quasars. *The Astronomical Journal*, 132:117–136.
- Fanidakis, N., Macciò, A. V., Baugh, C. M., Lacey, C. G., and Frenk, C. S. (2013). The most luminous quasars do not live in the most massive dark matter haloes at any redshift. *Monthly Notices of the Royal Astronomical Society*, 436:315–326.
- Fernández-Soto, A., Lanzetta, K. M., and Chen, H.-W. (2003). Mission: impossible (escape from the Lyman limit). *Monthly Notices of the Royal Astronomical Society*, 342:1215–1221.
- Ferrarese, L. (2002). Beyond the Bulge: A Fundamental Relation between Supermassive Black Holes and Dark Matter Halos. *The Astrophysical Journal*, 578:90–97.
- Francis, P. J. and Bland-Hawthorn, J. (2004). The mysterious absence of neutral hydrogen within 1 Mpc of a luminous quasar at redshift 2.168. *Monthly Notices of the Royal Astronomical Society*, 353:301–309.
- Fukugita, M., Shimasaku, K., and Ichikawa, T. (1995). Galaxy Colors in Various Photometric Band Systems. *Publications of the Astronomical Society of the Pacific*, 107:945.
- Gehrels, N. (1986). Confidence limits for small numbers of events in astrophysical data. *The Astrophysical Journal*, 303:336–346.
- Giodini, S., Lovisari, L., Pointecouteau, E., Ettori, S., Reiprich, T. H., and Hoekstra, H. (2013). Scaling Relations for Galaxy Clusters: Properties and Evolution. *Space Science Reviews*, 177:247–282.
- Gioia, I. M., Henry, J. P., Mullis, C. R., Voges, W., Briel, U. G., Böhringer, H., and Huchra, J. P. (2001). Cluster Evolution in the ROSAT North Ecliptic Pole Survey. *The Astrophysical Journal Letters*, 553:L105–L108.
- Hamilton, A. J. S. (1993). Toward Better Ways to Measure the Galaxy Correlation Function. *The Astrophysical Journal*, 417:19.

- Hawkins, E., Maddox, S., Cole, S., Lahav, O., Madgwick, D. S., Norberg, P., Peacock, J. A., Baldry, I. K., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Jones, B., Lewis, I., Lumsden, S., Percival, W., Peterson, B. A., Sutherland, W., and Taylor, K. (2003). The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. *Monthly Notices of the Royal Astronomical Society*, 346:78–96.
- Hennawi, J. F., Myers, A. D., Shen, Y., Strauss, M. A., Djorgovski, S. G., Fan, X., Glikman, E., Mahabal, A., Martin, C. L., Richards, G. T., Schneider, D. P., and Shankar, F. (2010). Binary Quasars at High Redshift. I. 24 New Quasar Pairs at $z \sim 3$ –4. *The Astrophysical Journal*, 719:1672–1692.
- Hennawi, J. F., Prochaska, J. X., Cantalupo, S., and Arrigoni-Battaia, F. (2015). Quasar quartet embedded in giant nebula reveals rare massive structure in distant universe. *Science*, 348:779–783.
- Hennawi, J. F., Strauss, M. A., Oguri, M., Inada, N., Richards, G. T., Pindor, B., Schneider, D. P., Becker, R. H., Gregg, M. D., Hall, P. B., Johnston, D. E., Fan, X., Burles, S., Schlegel, D. J., Gunn, J. E., Lupton, R. H., Bahcall, N. A., Brunner, R. J., and Brinkmann, J. (2006). Binary Quasars in the Sloan Digital Sky Survey: Evidence for Excess Clustering on Small Scales. *The Astronomical Journal*, 131:1–23.
- Hogg, D. W., Baldry, I. K., Blanton, M. R., and Eisenstein, D. J. (2002). The K correction. *ArXiv Astrophysics e-prints*.
- Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Robertson, B., and Springel, V. (2006). A Unified, Merger-driven Model of the Origin of Starbursts, Quasars, the Cosmic X-Ray Background, Supermassive Black Holes, and Galaxy Spheroids. *The Astrophysical Journal Supplement Series*, 163:1–49.
- Hopkins, P. F., Richards, G. T., and Hernquist, L. (2007). An Observational Determination of the Bolometric Quasar Luminosity Function. *The Astrophysical Journal*, 654:731–753.
- Husband, K., Bremer, M. N., Stanway, E. R., Davies, L. J. M., Lehnert, M. D., and Douglas, L. S. (2013). Are $z \sim 5$ quasars found in the most massive high-redshift overdensities? *Monthly Notices of the Royal Astronomical Society*, 432:2869–2877.
- Intema, H. T., Venemans, B. P., Kurk, J. D., Ouchi, M., Kodama, T., Röttgering, H. J. A., Miley, G. K., and Overzier, R. A. (2006). Large-scale structure of Lyman break galaxies around a radio galaxy protocluster at $z \sim 4$. *arXiv:astro-ph*, 456:433–437.
- Jenkins, A., Frenk, C. S., Pearce, F. R., Thomas, P. A., Colberg, J. M., White, S. D. M., Couchman, H. M. P., Peacock, J. A., Efstathiou, G., and Nelson, A. H. (1998). Evolution of Structure in Cold Dark Matter Universes. *The Astrophysical Journal*, 499:20–40.
- Jones, T., Stark, D. P., and Ellis, R. S. (2012). Keck Spectroscopy of Faint $3 < z < 7$ Lyman Break Galaxies. III. The Mean Ultraviolet Spectrum at $z \sim 4$. *The Astrophysical Journal*, 751:51.
- Kaiser, N., Aussel, H., Burke, B. E., Boesgaard, H., Chambers, K., Chun, M. R., Heasley, J. N., Hodapp, K.-W., Hunt, B., Jedicke, R., Jewitt, D., Kudritzki, R., Luppino, G. A., Maberry, M., Magnier, E., Monet, D. G., Onaka, P. M., Pickles, A. J., Rhoads, P. H. H., Simon, T., Szalay, A., Szapudi, I., Tholen, D. J., Tonry, J. L., Waterson, M., and Wick, J. (2002). Pan-STARRS: A Large Synoptic Survey Telescope Array. In Tyson, J. A. and Wolff, S., editors, *Survey and Other Telescope Technologies and Discoveries*, volume 4836 of *Proceedings of the International Society for Optical Engineering*, pages 154–164.

- Kaiser, N., Burgett, W., Chambers, K., Denneau, L., Heasley, J., Jedicke, R., Magnier, E., Morgan, J., Onaka, P., and Tonry, J. (2010). The Pan-STARRS wide-field optical/NIR imaging survey. In *Ground-based and Airborne Telescopes III*, volume 7733 of *Proceedings of the International Society for Optical Engineering*, page 77330E.
- Kashikawa, N., Kitayama, T., Doi, M., Misawa, T., Komiyama, Y., and Ota, K. (2007). The Habitat Segregation between Lyman Break Galaxies and Ly α Emitters around a QSO at $z \sim 5$. *The Astrophysical Journal*, 663:765–773.
- Kerscher, M., Szapudi, I., and Szalay, A. S. (2000). A Comparison of Estimators for the Two-Point Correlation Function. *The Astrophysical Journal Letters*, 535:L13–L16.
- Kim, S., Stiavelli, M., Trenti, M., Pavlovsky, C. M., Djorgovski, S. G., Scarlata, C., Stern, D., Mahabal, A., Thompson, D., Dickinson, M., Panagia, N., and Meylan, G. (2009). The Environments of High-Redshift Quasi-Stellar Objects. *The Astrophysical Journal*, 695:809–817.
- Krumpe, M., Miyaji, T., Husemann, B., Fanidakis, N., Coil, A. L., and Aceves, H. (2015). The Spatial Clustering of ROSAT All-Sky Survey Active Galactic Nuclei. IV. More Massive Black Holes Reside in More Massive Dark Matter Halos. *The Astrophysical Journal*, 815:21.
- Lacey, C. and Cole, S. (1993). Merger rates in hierarchical models of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, 262:627–649.
- Landy, S. D. and Szalay, A. S. (1993). Bias and variance of angular correlation functions. *The Astrophysical Journal*, 412:64–71.
- Limber, D. N. (1953). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *The Astrophysical Journal*, 117:134.
- Madau, P. (1995). Radiative transfer in a clumpy universe: The colors of high-redshift galaxies. *The Astrophysical Journal*, 441:18–27.
- Matsuda, Y., Yamada, T., Hayashino, T., Tamura, H., Yamauchi, R., Ajiki, M., Fujita, S. S., Murayama, T., Nagao, T., Ohta, K., Okamura, S., Ouchi, M., Shimasaku, K., Shioya, Y., and Taniguchi, Y. (2004). A Subaru Search for Ly α Blobs in and around the Protocluster Region At Redshift $z = 3.1$. *The Astronomical Journal*, 128:569–584.
- McLeod, K. K. and Bechtold, J. (2009). Host Galaxies of $z = 4$ Quasars. *The Astrophysical Journal*, 704:415–438.
- Morganson, E., De Rosa, G., Decarli, R., Walter, F., Chambers, K., McGreer, I., Fan, X., Burgett, W., Flewelling, H., Greiner, J., Hodapp, K., Kaiser, N., Magnier, E., Price, P., Rix, H.-W., Sweeney, B., and Waters, C. (2012). The First High-redshift Quasar from Pan-STARRS. *The Astronomical Journal*, 143:142.
- Morselli, L., Mignoli, M., Gilli, R., Vignali, C., Comastri, A., Sani, E., Cappelluti, N., Zamorani, G., Brusa, M., Gallozzi, S., and Vanzella, E. (2014). Primordial environment of super massive black holes: large-scale galaxy overdensities around $z \sim 6$ quasars with LBT. *arXiv:astro-ph*, 568:A1.
- Muldrew, S. I., Hatch, N. A., and Cooke, E. A. (2015). What are protoclusters? - Defining high-redshift galaxy clusters and protoclusters. *Monthly Notices of the Royal Astronomical Society*, 452:2528–2539.
- Nilsson, K. K., Möller-Nilsson, O., Möller, P., Fynbo, J. P. U., and Shapley, A. E. (2009). On the dependence between UV luminosity and Ly α equivalent width in high-redshift galaxies. *Monthly Notices of the Royal Astronomical Society*, 400:232–237.
- Oke, J. B. (1974). Absolute Spectral Energy Distributions for White Dwarfs. *The Astrophysical Journal Supplement Series*, 27:21.

- Ota, K., Kashikawa, N., Malkan, M. A., Iye, M., Nakajima, T., Nagao, T., Shimasaku, K., and Gandhi, P. (2008). Overdensity of i' -Dropout Galaxies in the Subaru Deep Field: A Candidate Protocluster at $z \sim 6$. *ArXiv e-prints*.
- Ouchi, M., Shimasaku, K., Akiyama, M., Sekiguchi, K., Furusawa, H., Okamura, S., Kashikawa, N., Iye, M., Kodama, T., Saito, T., Sasaki, T., Simpson, C., Takata, T., Yamada, T., Yamanoi, H., Yoshida, M., and Yoshida, M. (2005). The Discovery of Primeval Large-Scale Structures with Forming Clusters at Redshift 6. *The Astrophysical Journal Letters*, 620:L1–L4.
- Ouchi, M., Shimasaku, K., Akiyama, M., Simpson, C., Saito, T., Ueda, Y., Furusawa, H., Sekiguchi, K., Yamada, T., Kodama, T., Kashikawa, N., Okamura, S., Iye, M., Takata, T., Yoshida, M., and Yoshida, M. (2008). The Subaru/XMM-Newton Deep Survey (SXDS). IV. Evolution of Ly α Emitters from $z=3.1$ to 5.7 in the 1 deg² Field: Luminosity Functions and AGN. *The Astrophysical Journal Supplement Series*, 176:301–330.
- Ouchi, M., Shimasaku, K., Furusawa, H., Miyazaki, M., Doi, M., Hamabe, M., Hayashino, T., Kimura, M., Kodaira, K., Komiyama, Y., Matsuda, Y., Miyazaki, S., Nakata, F., Okamura, S., Sekiguchi, M., Shioya, Y., Tamura, H., Taniguchi, Y., Yagi, M., and Yasuda, N. (2003). Subaru Deep Survey. II. Luminosity Functions and Clustering Properties of Ly α Emitters at $z=4.86$ in the Subaru Deep Field. *The Astrophysical Journal*, 582:60–68.
- Ouchi, M., Shimasaku, K., Okamura, S., Furusawa, H., Kashikawa, N., Ota, K., Doi, M., Hamabe, M., Kimura, M., Komiyama, Y., Miyazaki, M., Miyazaki, S., Nakata, F., Sekiguchi, M., Yagi, M., and Yasuda, N. (2004a). Subaru Deep Survey. V. A Census of Lyman Break Galaxies at $z \sim 4$ and 5 in the Subaru Deep Fields: Photometric Properties. *The Astrophysical Journal*, 611:660–684.
- Ouchi, M., Shimasaku, K., Okamura, S., Furusawa, H., Kashikawa, N., Ota, K., Doi, M., Hamabe, M., Kimura, M., Komiyama, Y., Miyazaki, M., Miyazaki, S., Nakata, F., Sekiguchi, M., Yagi, M., and Yasuda, N. (2004b). Subaru Deep Survey. VI. A Census of Lyman Break Galaxies at $z \sim 4$ and 5 in the Subaru Deep Fields: Clustering Properties. *The Astrophysical Journal*, 611:685–702.
- Overzier, R. A., Bouwens, R. J., Cross, N. J. G., Venemans, B. P., Miley, G. K., Zirm, A. W., Benítez, N., Blakeslee, J. P., Coe, D., Demarco, R., Ford, H. C., Homeier, N. L., Illingworth, G. D., Kurk, J. D., Martel, A. R., Mei, S., Oliveira, I., Röttgering, H. J. A., Tsvetanov, Z. I., and Zheng, W. (2008). Lyman Break Galaxies, Ly α Emitters, and a Radio Galaxy in a Protocluster at $z = 4.1$. *The Astrophysical Journal*, 673:143–162.
- Overzier, R. A., Guo, Q., Kauffmann, G., De Lucia, G., Bouwens, R., and Lemson, G. (2009). Λ CDM predictions for galaxy protoclusters - I. The relation between galaxies, protoclusters and quasars at $z \sim 6$. *Monthly Notices of the Royal Astronomical Society*, 394:577–594.
- Padmanabhan, T. (2006). Advanced Topics in Cosmology: A Pedagogical Introduction. In Daflon, S., Alcaniz, J., Telles, E., and de la Reza, R., editors, *Graduate School in Astronomy: X*, volume 843 of *American Institute of Physics Conference Series*, pages 111–166.
- Patat, F., Moehler, S., O’Brien, K., Pompei, E., Bensby, T., Carraro, G., de Ugarte Postigo, A., Fox, A., Gavignaud, I., James, G., Korhonen, H., Ledoux, C., Randall, S., Sana, H., Smoker, J., Stefl, S., and Szeifert, T. (2011). Optical atmospheric extinction over Cerro Paranal. *arXiv:astro-ph*, 527:A91.
- Peacock, J. A. (1999). *Cosmological Physics*.
- Penzias, A. A. and Wilson, R. W. (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *The Astrophysical Journal*, 142:419–421.
- Porciani, C. and Norberg, P. (2006). Luminosity- and redshift-dependent quasar clustering. *Monthly Notices of the Royal Astronomical Society*, 371:1824–1834.

- Sargent, W. L. W. and Turner, E. L. (1977). A statistical method for determining the cosmological density parameter from the redshifts of a complete sample of galaxies. *The Astrophysical Journal Letters*, 212:L3–L7.
- Schlegel, D. J., Finkbeiner, D. P., and Davis, M. (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *The Astrophysical Journal*, 500:525–553.
- Schneider, P. (2015). *Extragalactic Astronomy and Cosmology: An Introduction*.
- Shapley, A. E., Steidel, C. C., Pettini, M., and Adelberger, K. L. (2003). Rest-Frame Ultraviolet Spectra of $z \sim 3$ Lyman Break Galaxies. *The Astrophysical Journal*, 588:65–89.
- Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., and Erb, D. K. (2006). The Direct Detection of Lyman Continuum Emission from Star-forming Galaxies at $z \sim 3$. *The Astrophysical Journal*, 651:688–703.
- Shen, Y., Richards, G. T., Strauss, M. A., Hall, P. B., Schneider, D. P., Snedden, S., Bizyaev, D., Brewington, H., Malanushenko, V., Malanushenko, E., Oravetz, D., Pan, K., and Simmons, A. (2011). A Catalog of Quasar Properties from Sloan Digital Sky Survey Data Release 7. *The Astrophysical Journal Supplement Series*, 194:45.
- Shen, Y., Strauss, M. A., Oguri, M., Hennawi, J. F., Fan, X., Richards, G. T., Hall, P. B., Gunn, J. E., Schneider, D. P., Szalay, A. S., Thakar, A. R., Vanden Berk, D. E., Anderson, S. F., Bahcall, N. A., Connolly, A. J., and Knapp, G. R. (2007). Clustering of High-Redshift ($z \geq 2.9$) Quasars from the Sloan Digital Sky Survey. *The Astronomical Journal*, 133:2222–2241.
- Sijacki, D., Springel, V., Di Matteo, T., and Hernquist, L. (2007). A unified model for AGN feedback in cosmological simulations of structure formation. *Monthly Notices of the Royal Astronomical Society*, 380:877–900.
- Simon, P. (2007). How accurate is Limber’s equation? *arXiv:astro-ph*, 473:711–714.
- Simpson, C., Mortlock, D., Warren, S., Cantalupo, S., Hewett, P., McLure, R., McMahan, R., and Venemans, B. (2014). No excess of bright galaxies around the redshift 7.1 quasar ULAS J1120+0641. *Monthly Notices of the Royal Astronomical Society*, 442:3454–3461.
- Springel, V., Di Matteo, T., and Hernquist, L. (2005a). Modelling feedback from stars and black holes in galaxy mergers. *Monthly Notices of the Royal Astronomical Society*, 361:776–794.
- Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005b). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636.
- Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., and Giavalisco, M. (2000). Ly α Imaging of a Proto-Cluster Region at $z = 3.09$. *The Astrophysical Journal*, 532:170–182.
- Steidel, C. C., Pettini, M., and Hamilton, D. (1995). Lyman Imaging of High-Redshift Galaxies. III. New Observations of Four QSO Fields. *The Astronomical Journal*, 110:2519.
- Stetson, P. B. (2000). Homogeneous Photometry for Star Clusters and Resolved Galaxies. II. Photometric Standard Stars. *Publications of the Astronomical Society of the Pacific*, 112:925–931.
- Stiavelli, M., Djorgovski, S. G., Pavlovsky, C., Scarlata, C., Stern, D., Mahabal, A., Thompson, D., Dickinson, M., Panagia, N., and Meylan, G. (2005). Evidence of Primordial Clustering around the QSO SDSS J1030+0524 at $z = 6.28$. *The Astrophysical Journal Letters*, 622:L1–L4.
- Stone, R. P. S. (1996). Spectrophotometry of Flux Calibration Stars for Hubble Space Telescope. *The Astrophysical Journal Supplement Series*, 107:423.

- Toshikawa, J., Kashikawa, N., Ota, K., Morokuma, T., Shibuya, T., Hayashi, M., Nagao, T., Jiang, L., Malkan, M. A., Egami, E., Shimasaku, K., Motohara, K., and Ishizaki, Y. (2012). Discovery of a Protocluster at $z \sim 6$. *The Astrophysical Journal*, 750:137.
- Trainor, R. F. and Steidel, C. C. (2012). The Halo Masses and Galaxy Environments of Hyperluminous QSOs at $z \sim 2.7$ in the Keck Baryonic Structure Survey. *The Astrophysical Journal*, 752:39.
- Utsumi, Y., Goto, T., Kashikawa, N., Miyazaki, S., Komiyama, Y., Furusawa, H., and Overzier, R. (2010). A Large Number of $z > 6$ Galaxies Around a QSO at $z = 6.43$: Evidence for a Protocluster? *The Astrophysical Journal*, 721:1680–1688.
- Venemans, B. P., Bañados, E., Decarli, R., Farina, E. P., Walter, F., Chambers, K. C., Fan, X., Rix, H.-W., Schlafly, E., McMahon, R. G., Simcoe, R., Stern, D., Burgett, W. S., Draper, P. W., Flewelling, H., Hodapp, K. W., Kaiser, N., Magnier, E. A., Metcalfe, N., Morgan, J. S., Price, P. A., Tonry, J. L., Waters, C., AlSayyad, Y., Banerji, M., Chen, S. S., González-Solares, E. A., Greiner, J., Mazzucchelli, C., McGreer, I., Miller, D. R., Reed, S., and Sullivan, P. W. (2015). The Identification of Z-dropouts in Pan-STARRS1: Three Quasars at $6.5 < z < 6.7$. *The Astrophysical Journal Letters*, 801:L11.
- Venemans, B. P., Röttgering, H. J. A., Miley, G. K., van Breugel, W. J. M., de Breuck, C., Kurk, J. D., Pentericci, L., Stanford, S. A., Overzier, R. A., Croft, S., and Ford, H. (2007). Protoclusters associated with $z > 2$ radio galaxies . I. Characteristics of high redshift protoclusters. *arXiv:astro-ph*, 461:823–845.
- Vestergaard, M. (2002). Determining Central Black Hole Masses in Distant Active Galaxies. *The Astrophysical Journal*, 571:733–752.
- Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., Ebeling, H., Forman, W. R., Hornstrup, A., Jones, C., Murray, S. S., Nagai, D., Quintana, H., and Voevodkin, A. (2009). Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints. *The Astrophysical Journal*, 692:1060–1074.
- Wall, J. V. and Jenkins, C. R. (2003). *Practical Statistics for Astronomers*.
- White, S. D. M., Navarro, J. F., Evrard, A. E., and Frenk, C. S. (1993). The baryon content of galaxy clusters: a challenge to cosmological orthodoxy. *Nature*, 366:429–433.
- Willott, C. J., Percival, W. J., McLure, R. J., Crampton, D., Hutchings, J. B., Jarvis, M. J., Sawicki, M., and Simard, L. (2005). Imaging of SDSS $z > 6$ Quasar Fields: Gravitational Lensing, Companion Galaxies, and the Host Dark Matter Halos. *The Astrophysical Journal*, 626:657–665.
- Worseck, G. and Prochaska, J. X. (2011). GALEX Far-ultraviolet Color Selection of UV-bright High-redshift Quasars. *The Astrophysical Journal*, 728:23.
- Wyithe, J. S. B. and Loeb, A. (2002). A Physical Model for the Luminosity Function of High-Redshift Quasars. *The Astrophysical Journal*, 581:886–894.
- Yang, Y., Zabludoff, A., Tremonti, C., Eisenstein, D., and Davé, R. (2009). Extended Ly α Nebulae at $z \simeq 2.3$: An Extremely Rare and Strongly Clustered Population? *The Astrophysical Journal*, 693:1579–1587.
- Zehavi, I., Zheng, Z., Weinberg, D. H., Blanton, M. R., Bahcall, N. A., Berlind, A. A., Brinkmann, J., Frieman, J. A., Gunn, J. E., Lupton, R. H., Nichol, R. C., Percival, W. J., Schneider, D. P., Skibba, R. A., Strauss, M. A., Tegmark, M., and York, D. G. (2011). Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *The Astrophysical Journal*, 736:59.

Zheng, W., Overzier, R. A., Bouwens, R. J., White, R. L., Ford, H. C., Benítez, N., Blakeslee, J. P., Bradley, L. D., Jee, M. J., Martel, A. R., Mei, S., Zirm, A. W., Illingworth, G. D., Clampin, M., Hartig, G. F., Ardila, D. R., Bartko, F., Broadhurst, T. J., Brown, R. A., Burrows, C. J., Cheng, E. S., Cross, N. J. G., Demarco, R., Feldman, P. D., Franx, M., Golimowski, D. A., Goto, T., Gronwall, C., Holden, B., Homeier, N., Infante, L., Kimble, R. A., Krist, J. E., Lesser, M. P., Menanteau, F., Meurer, G. R., Miley, G. K., Motta, V., Postman, M., Rosati, P., Sirianni, M., Sparks, W. B., Tran, H. D., and Tsvetanov, Z. I. (2006). An Overdensity of Galaxies near the Most Distant Radio-loud Quasar. *The Astrophysical Journal*, 640:574–578.

List of Figures

| | | |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Simulation showing the dark matter distribution in the Universe at $z = 0$ at different scales. Each slice have a thickness of 15 Mpc h^{-1} . The lower slice shows an homogeneous Universe as the cosmological principle states, but the smaller the scale is, the more distinguishable the structure is, and the inhomogeneity at small scales become clear. This figure is taken from Springel et al. (2005b) (The background color of this figure was modified). | 2 |
| 1.2 | Distribution of galaxies in the Universe, from the 2dF galaxy redshift survey. The polar angle is indicating the right ascension and the radial axes indicates the redshift. The large scale structure formed by filaments and voids is clearly observed. This figure is taken from the 2dF image gallery. | 3 |
| 1.3 | Simulations of the dark matter distribution in the Universe for four different redshifts and the same scale. At higher redshifts the fluctuations in the density field were smaller and the large scale structure was less well defined in comparison with the lower redshift snapshots. This illustrates the standard structure formation scenario, where structures grow hierarchically through gravitational instability. This figures were made with the images available at www.mpa-garching.mpg.de/galform/virgo/millennium/ | 5 |
| 3.1 | Filter configuration shown on a LBG simulated spectrum at $z = 3.78$ (see section 4.1 for the simulated spectra details). The narrow bands were designed especially for this program, in order to identify LBGs at $z \sim 3.78$ in a quiet narrow redshift slice of $\Delta z \sim 0.2$ | 26 |
| 4.1 | Distribution of the UV continuum slope α and $\text{EW}_{\text{Ly}\alpha}$ parameters for the simulated spectra. <i>Left panel:</i> UV continuum slope (α) distribution for the 1000 simulated spectra. α was chosen from a Gaussian distribution with mean $\mu = -1.676$ and $\sigma = 0.39$ Bouwens et al. (2009). α was defined as the slope measured in the wavelength range $1300 - 2000\text{\AA}$. <i>Right panel:</i> $\text{Ly}\alpha$ EW distribution for the simulated spectra. Negative values correspond to emission lines. EW was chosen from a Gaussian distribution with mean $\mu = -25\text{\AA}$ and $\sigma = 40\text{\AA}$ Shapley et al. (2003) plus an exponential tail of high EW values with scale length of $W_0 = -64\text{\AA}$ Ciardullo et al. (2012). | 31 |
| 4.2 | Mean LBG spectrum at $z = 3.78$. This is computed as the median of 1000 simulated LBGs spectra, which have different continuum slopes and $\text{EW}_{\text{Ly}\alpha}$. . . | 33 |
| 4.3 | Color-color diagram showing the simulated colors for 1000 LBGs spectra, plotted as redshift color-coded points according to the color bar. The mean LBG evolutionary track is also plotted as a black curve. The filled points indicate the redshift from 3.6 to 4.2. The larger circle shows the exact position of the mean $z = 3.78$ LBG colors. | 34 |
| 4.4 | Spectral templates for 5 different types of galaxies used to figure out how low- z galaxies populate the color-color diagram. The galaxies shown from left to right and from top to bottom are Elliptical, Sa, Sb, Sc, and Irregular respectively. . . | 36 |

- 4.5 Low- z galaxies evolutionary tracks redshifted from $z = 0$ to $z = 3$. We plot as brown, magenta, orange, blue, and red curves the evolutionary track of elliptical, Sa, Sb, Sc, and Irregular galaxies, respectively. We overplotted the track of LBGs computed as was explained in section 4.1 as a red curve. Filled circles over the red curve indicate colors of LBGs from redshift 3.6 to 4.0, and the largest red point indicates the exact position of the color of LBGs at $z = 3.78$ 37
- 4.6 *Left panel:* Same as Fig. 4.5, but now using the B, R, i filters of the Suprim-cam on the Subaru Telescope. Filled circles on the low- z galaxies evolutionary tracks are indicating the redshifts 0, 1, 2, and 3. The dashed line is showing the color cuts used by Ouchi et al. (2004a) to selects LBGs at $z \sim 4$. *Right panel:* Similar plot but taken directly from Ouchi et al. (2004a). They show as a red line the track of a typical spectrum of star-forming galaxies from $z = 3$ to 4. Filled circles on the red line indicate the redshift from $z = 3.3$ to 3.9 with an interval of $\Delta z = 0.1$. Typical spectra of elliptical, Sbc, Scd, and irregular galaxies are redshifted from $z = 0$ to 3, shown by green, cyan, blue, and violet lines, respectively. Each line is marked by filled circles at $z = 0, 1, \text{ and } 2$. Yellow stars show 175 Galactic stars. 38
- 4.7 Examples of some interlopers that could be affecting our LBGs sample. *Top left:* The spectra of an elliptical galaxy at $z = 0.46$, where the Balmer break is located at $\lambda \sim 5840$. *Top right:* The spectra of a galaxy at $z = 1$, where the MgI and MgII absorption are located at $\lambda \sim 5650$. *Bottom:* The spectra of a spiral galaxy at $z = 0.59$, where the OII line is located at $\lambda \sim 5925$ 39
- 4.8 Same as Fig. 4.5 but we overplotted the selection region as a dashed line. 41
- 4.9 Color-color diagram for our 6 stacked QSO fields. Here the evolutionary track showed in Fig. 4.8 is plotted as redshift color-coded track according to the color bar. The magenta point indicate the color of the QSO in our filters. Arrows indicate lower limits for $\text{NB}_{571} - \text{NB}_{596}$ color. These are cases in which the object was not detected in NB_{571} filter at 2σ level and magnitude was replaced by the corresponding limit magnitude. 42
- 4.10 Color-color diagram for our individual 6 QSO fields. At the top right of each plot the number of LBGs found inside of the selection region is shown. 43
- 4.11 Images of some selected LBGs. From left to right we show the NB_{571} , NB_{596} and r_{GUNN} images. Each panel is $15''$ on a side. The red circle show the position of the detected object, and the size correspond to the region in which the photometry was done ($2''$ in diameter). The magnitudes are indicated in each panel. 45
- 4.12 The distribution of LBGs around the quasar in the plane of the sky for the stacked 6 fields. The central QSOs is located in 0.0 and is plotted by a large black circle. 46
- 4.13 Color-color diagram showing the simulated colors for 1000 LBGs spectra including photometric errors, plotted as redshift color-coded points according to the color bar. The mean LBG evolutionary track is plotted as a black curve. The filled points indicate the redshift from 3.6 to 4.2. The bigger circle shows the exact position of the mean $z = 3.78$ LBG colors. We overplotted the selection region as a dashed line. 47
- 4.14 Completeness of the LBGs selection. The completeness was determined from 1000 simulated LBGs spectra with different EW and, continuum power law and magnitudes. This is calculated by computing the fraction of the simulated spectra, per redshift bin, which was selected by the selection region. This define the redshift selection function that is used for the clustering measurements in the next chapter. 48
- 5.1 QSO-LBG cross-correlation function and the Maximum likelihood estimate. The filled circles are showing our measurement described in section 5.2 with 1σ Poisson error bars. The dashed black line shows the theoretical expectation of χ for our six stacked fields calculated from the QSO and LBGs auto-correlation functions. *Top:* The solid red curve shows the best MLE for both r_0 and γ as free parameters. We obtain $r_0 = 9.91^{+3.28}_{-1.79} h^{-1} \text{ cMpc}$ and $\gamma = 2.05^{+0.20}_{-0.46}$. *Bottom:* The solid red curve shows the best MLE for r_0 with a fixed $\gamma = 1.8$. Our results are consistent, with the expectation value computed in the section 5.1. 57

| | | |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 5.2 | 1σ and 2σ confidence regions of r_0 and γ parameters (in blue and red respectively), determined using a maximum likelihood estimator. The best estimation is shown as a white cross. This result indicates that we can not perform a good estimation of those parameters simultaneously, which is in part due to the large error bars in our measurements. We then preferred to fit this data with a fixed γ value. | 58 |
| 5.3 | LBGs auto-correlation function in QSOs fields. The data points are showing our measurement described in section 5.3. The dotted black curve shows the results of LBGs auto-correlation in blank fields (Ouchi et al., 2004b). The solid red curve shows the best fit for our clustering measurements for a fix $\gamma = 1.8$. We find a higher clustering in our fields in comparison with random fields as measured by (Ouchi et al., 2004b), which suggests that our fields are located in overdense regions. | 59 |
| 5.4 | 1σ and 2σ confidence regions of r_0 and γ parameters (in blue and red respectively) determined using a maximum likelihood estimator for the LBGs auto-correlation measurement. The best estimation is shown as a white cross. | 61 |
| 5.5 | Shape of the function $\chi = X(1 + \xi_{vol}) - 1$ for different X values, where ξ_{vol} is modeled as in equation (5.23) using an arbitrary r_0 and γ value. When $X = 1$ we recover the power law shape of the function, but as X increase, the shape become increasingly flat. | 63 |
| 5.6 | Angular cross-correlation function measurement for the LBGs sample. This measurement is used to test if the sample is contaminated. We expect that if the sample is highly contaminated $\omega(\theta) = 0$ at every scale and a power law fit with slope β is consistent with $\beta = 0$. We found that our LBGs sample is not highly contaminated, given their power law shape in this plot, which is well fitted by $\omega(\theta) = 31.1\theta^{-1.15}$ | 65 |
| 6.1 | Filter configuration shown on a galaxy simulated spectra at $z \sim 3.87$ | 70 |
| 6.2 | Distribution of the $EW_{Ly\alpha}$ used to create the simulated spectra. Negative values correspond to emission lines. EW is chosen from an exponential distribution with scale length of $W_0 = -64\text{\AA}$ (Ciardullo et al., 2012). The mean $EW_{Ly\alpha}$ of this distribution is $\sim 90\text{\AA}$ | 73 |
| 6.3 | Color-color diagram showing the simulated colors of 1000 LAEs spectra, including photometric errors and plotted as redshift color-coded points according to the color bar. The mean LAE evolutionary track is also plotted as a black curve. The filled points over the curve from the left to the right indicate the redshift from 3.3 to 4.1. The larger circle shows the exact position of the mean $z = 3.87$ LAEs colors and the black dashed line is indicating the LAEs selection region. | 75 |
| 6.4 | Low- z galaxies evolutionary tracks redshifted from $z = 0$ to $z = 3$. We plot as brown, magenta, orange, blue, and red curves the evolutionary track of elliptical, Sa, Sb, Sc, and irregular galaxies respectively. We overplotted the track of LAEs as a red curve. Filled circles over the red curve indicate colors of LAEs from redshift 3.3 to 4.0, and the largest red point indicate the exact position of the color of LAEs at $z = 3.87$. We overplotted the selection region as a black dashed line. | 76 |
| 6.5 | Color-color diagram for our 14 stacked QSO fields. Here the evolutionary track showed in Fig. 6.4 is plotted as redshift color-coded track according to the color bar. The magenta points indicate the color of the QSO in our filters. The magnitudes of cases in which the object was not detected in g or R filter at 2σ level, were replaced by the corresponding limit magnitude. Those cases are shown with arrows. Cases with no detection on either filter are shown with both arrows. The dashed line is indicating the selection region defined by equations (6.4) and (6.5). | 77 |
| 6.6 | Images of some selected LAEs. From left to right we show the g , HeI , and R images. Each panel is $15''$ on a side. The red circle shows the position of the detected object, and the size correspond to the region in which the photometry was done ($2''$ in diameter). The magnitudes are indicated in each panel. | 79 |

| | | |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 6.7 | The distribution of LAEs around the quasar in the plane of the sky for the stacked 14 fields. The central QSOs is located in 0.0 and is plotted by a large black circle. | 80 |
| 6.8 | Completeness of the LAEs selection. The completeness was determined from 1000 simulated LAE spectra with different EW and, continuum power law and magnitudes. This is calculated by computing the fraction of the simulated spectra, per redshift bin, which was selected by the selection region. | 81 |
| 6.9 | QSO-LAE cross-correlation function. The filled circles are showing our measurement with 1σ Poisson error bars. The dashed black line shows the theoretical expectation of χ for our 14 stacked fields calculated from the QSO and LAEs auto-correlation functions. | 82 |
| 6.10 | Cumulative LAEs number density for our 14 stacked fields, without the inclusion of the 9 LAEs located outside of our selection region (red points). If we include those extra 9 LAEs, they only influence the measurements in the last 2 bins, which is plotted as green points, shifted 0.1 mag for clarity. The black points are the LAEs number density in random fields (Ouchi et al., 2008). The vertical line is indicating the mean 5σ limit magnitude of our fields. | 83 |
| | | |
| A.1 | Images of selected LBGs for 0124+0044. | 92 |
| A.2 | Images of selected LBGs for 0213-0904 (continues). | 93 |
| A.3 | Images of selected LBGs for 0213-0904 (continuation). | 94 |
| A.4 | Images of selected LBGs for 2003-33. | 94 |
| A.5 | Images of selected LBGs for 2207+0043. | 95 |
| A.6 | Images of selected LBGs for 2301+0112. | 96 |
| A.7 | Images of selected LBGs for 2311-0844. | 97 |
| | | |
| B.1 | Images of selected LAEs for SDSSJ0040. | 100 |
| B.2 | Images of selected LAEs for SDSSJ0119. | 100 |
| B.3 | Images of selected LAEs for SDSSJ0149. | 100 |
| B.4 | Images of selected LAEs for SDSSJ1026. | 101 |
| B.5 | Images of selected LAEs for SDSSJ1044. | 101 |
| B.6 | Images of selected LAEs for SDSSJ1205. | 101 |
| B.7 | Images of selected LAEs for SDSSJ1224. | 102 |
| B.8 | Images of selected LAEs for SDSSJ1258. | 102 |
| B.9 | Images of selected LAEs for SDSSJ2250. | 102 |
| B.10 | Images of selected LAEs for SDSSJ2350. | 102 |
| B.11 | Color-color diagram for our 14 individual QSO fields. | 103 |
| B.12 | Continuation | 104 |

List of Tables

| | | |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Protoclusters in QSO environments at $z \gtrsim 5$. Summary: 21 studied QSO fields: 13 shows galaxy overdensities and 8 do not show overdensities. | 8 |
| 3.1 | Targeted Quasars properties. | 25 |
| 3.2 | 5σ limit magnitudes per field measured in a $2''$ diameter aperture. | 28 |
| 4.1 | LBGs sample. The magnitudes correspond to AB magnitudes measured in a $2''$ diameter aperture for each filter | 44 |
| 5.1 | QSO-LBG Cross-Correlation Function. | 54 |
| 5.2 | LBGs Auto-Correlation Function. | 60 |
| 6.1 | Targeted Quasars properties. | 69 |
| 6.2 | 5σ limit magnitudes per field measured in a $2''$ diameter aperture. | 71 |
| 6.3 | LAEs sample. The magnitudes correspond to AB magnitudes measured in a $2''$ diameter aperture for each filter | 78 |
| 6.4 | QSO-LAE Cross-Correlation Function. | 81 |

