UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO

PROPIEDADES DE AGLOMERACIÓN DE GALAXIAS NORMALES Y ACTIVAS A $z \sim 3$ EN LA EXPLORACIÓN MUSYC

HAROLD FRANCKE HENRIQUEZ

MIEMBROS DE LA COMISIÓN EVALUADORA

Prof. Paulina Lira Teillery, Profesor Guía
Prof. Eric Gawiser
Prof. Luis Campusano Brown
Prof. Eduardo Hardy Raskovan
Prof. Leopoldo Infante Lira

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS
MENCIÓN ASTRONOMÍA

SANTIAGO DE CHILE
MARZO 2009

RESUMEN

Las observaciones de galaxias a corrimientos al rojo menores que 1 ($z < 1$) han mostrado que durante los últimos 8 giga-años de existencia del universo, las propiedades generales de las galaxias se mantienen parecidas a las actuales. Durante la última década, sin embargo, se han podido descubrir grandes poblaciones de galaxias en el universo temprano. Con esto se hace posible hoy en día estudiar directamente a las progenitoras de las galaxias modernas en pleno proceso de formación. Adicionalmente, recientemente se ha descubierto que los agujeros negros supermasivos, que generan la energía que alimenta a los núcleos activos en galaxias, se encuentran de alguna manera íntimamente vinculados al proceso de formación y a la evolución de las galaxias que los alojan.

La distribución espacial de las galaxias ha sido estudiada clásicamente con el fin de restringir parámetros cosmológicos, sin embargo, también es muy útil para estudiar la formación de galaxias, dado que ésta comienza justamente con el colapso de materia en estructuras que se denominan halos. La aglomeracíon espacial de estos halos depende directamente de la cantidad de materia que contienen y, por lo tanto, es un indicador de la masa total típica para cada población de galaxias.

En este trabajo de tesis se propuso determinar las propiedades de aglomeración de galaxias con formación estelar activa a $z \sim 3$, una época en que el universo era bastante joven, a 2 giga-años del Big-Bang. Utilizando los datos ópticos y de rayos X de la exploración MUSYC (MUltiwavelength Survey by Yale-Chile) se seleccionaron galaxias Lyman-Break (LBGs), galaxias emisoras de radiación Lyman alfa (LAEs) y galaxias con núcleo activo (AGN). Luego se llevó a cabo un seguimiento espectroscópico con el fin de confirmarlas y medirles su corrimiento al rojo. Con esta información se dedujeron, asimismo, las propiedades físicas de los halos de materia oscura que alojan a estas galaxias, y se restringieron a los posibles ancestros y descendientes de ellas.

Nuestro estudio de galaxias Lyman-Break confirmó, por una parte, resultados preexistentes en la literatura con respecto a la masas de los halos de estas galaxias, y

a que muchos de los descendientes de éstas en el universo local deberían encontrarse entre halos que alojen a galaxias elípticas o a grupos de galaxias.

Para las galaxias que poseían un núcleo activo, en cambio, se encontró que tendían a hallarse típicamente entre halos un orden de magnitud más masivos que los de galaxias normales con colores similares. Adicionalmente, con este trabajo se mostró que es posible seleccionar exitosamente AGNs menos luminosos que un quasar típico, en un rango restringido de redshift, tan sólo mediante uso combinado de información óptica y de rayos X.

Las galaxias emisoras en Lyman alfa, en cambio, mostraron un nivel de aglomeración muy bajo, indicando que estarían alojadas en halos poco masivos, la población de halos menos masivos encontrados hasta ahora a este redshift. Esto abre la posibilidad de que, buscando a los descendientes de estas galaxias en nuestro universo local, podríamos encontrar un gran fracción de galaxias como la Vía Láctea.

Profesores Guía: Eric Gawiser y Paulina Lira

A mi familia.

Agradecimientos

Quiero agradecer a mis padres, Juan Francke y Astrid Henríquez, sin cuyo apoyo, cariño, paciencia y preocupación jamás hubiese podido llegar a este punto. A mis hermanos, Hardy y Astrid, por el infinito cariño y comprensión que tuvieron con su hermano que vive en las nubes. También quiero agradecer a mis profesores guía: Eric Gawiser y Paulina Lira. No sólo por lo que me enseñaron, sino que también por lo mucho que siempre me apoyaron y la paciencia que me tuvieron. A la familia Alcaíno Olivares: mis suegros Graciela y Hernán papá (qepd), y a mis cuñados Hernán hijo y Felipe, que me recibieron como si fuese uno de los suyos y me dejaron acompañarlos hasta en los momentos más duros. A los amigos que me acompañaron (en algún momento) en la Universidad de Chile: a Diego Muñoz, Christian Moni, Andrés Escala y Juan Cortés por las entretenidas peleas en la oficina de los renegados; a Liza Videla, Sonia Duffau, Guillermo Blanc, Manuel Merello, Nicolás Tejos, María Fernanda, José Gallardo y Melina Bersten, que convirtieron la oficina en una pequeña vecindad. Y los amigos con que compartí en Yale: Gabriel Brammer, Shanil Virani, Lissa Miller, Anne Abramson, Pedro Capelo, Ezequiel Treister (exportado de Chile) y David Herrera, gracias por recibirme tan calurosamente al glacial invierno de New Haven. A todos ellos, muchas gracias por compartir los consejos, el relajo, los enojos, las frustraciones y las alegrías. Quiero también agradecer a mi tío Ricardo Francke, que siempre que pasó por San Felipe trajo algún artefacto extraño, problema matemático o regalo que me picara la curiosidad científica. Por permitirme llevar a cabo este trabajo, también quisiera agradecer a MECESUP, a Fundación Andes y a CONICYT. Finalmente, quiero agradecerle a mi esposa, Pamela, cuya compañia, consejos y amor me han mantenido con un pie en la tierra durante todo este largo período de ensimismamiento y divagaciones astronómicas.

Table of Contents

List of Figures

List of Tables

Introduction

The discovery, in the last decades, of large galaxy populations at high redshift has shown some of the possible seeds of modern galaxies in the young universe. The Lyman-break selection technique, since the mid-90s, has yielded relatively large and clean samples of high redshift, starforming galaxies (Steidel et al., 1996a,b). These galaxies have stellar masses of about $10^{10} M_{\odot}$, comparable to an L* galaxy in the nearby universe, but with much higher star formation rates, between 10-100 $M_{\odot}$/yr (Shapley et al., 2001; Steidel et al., 2003; Reddy et al., 2005). Narrow-band searches of the Lyman-Alpha line coming from high redshift galaxies have also proven to be successful in generating relatively large samples of sources (Hu & McMahon, 1996; Cowie & Hu, 1998), probing the faint end of the galaxy luminosity function at high redshifts. The relation of these star-forming galaxies to other high redshift populations, and to local galaxies is not completely understood yet, but it seems that the majority of the UV continuum selected samples evolve into fairly massive galaxies today (Adelberger et al., 2005b). At the same time, in the last few years the discovery of correlations between the mass of supermassive massive black holes and other properties of the hosting galaxies have proven that the presence of an active nucleus has a deep impact in the way a galaxy forms and evolves (Magorrian et al., 1998; Gebhardt et al., 2000).

In this work, we will address the problem of galaxy formation and evolution from the point of view of clustering analysis. Clustering strength is one of the fundamental properties of a galaxy population, and a key quantity to include in any complete description of a sample. Galaxies are thought to form at density peaks, and thus, to be biased tracers of underlying matter field (Kaiser, 1984; Bardeen

et al., 1986). In the current paradigm of hierarchical cosmology, where the structure formation and evolution is driven by the dynamics of dark matter, the effect of gravity reflects directly on the distribution of galaxies at the largest scales. The analysis of clustering can tell us the place of a given set of galaxies in the hierarchy tree of the large scale structure of the universe, by giving a measurement of the masses and number densities of the dark matter halos that host these galaxies. Furthermore, following the dynamics of these halos can help us in drawing an evolutionary picture between the different sets of galaxies we observe in the distant and local universe. The advantage of this approach is that it gives information which is complementary to what is obtained from the light itself, i.e. from the Spectral Energy Distribution (SED) of the galaxies. It is relatively independent of the complex physics behind the formation of stars and the emission of light.

The current standard (or *concordance*) cosmological model, ΛCDM, is a model for a spatially flat, homogeneous and isotropic universe, originated from a hot and dense initial state, called the Big Bang. It is composed of ordinary matter and radiation, but also containing a huge amount of dark matter and showing the presence of a cosmological constant (also known as dark energy). This is a strange universe, in which the nature of most the content is far from understood: cold dark matter (CDM) and dark energy reach nearly 30% and 70% of the total density budget respectively, leaving only a few percent for ordinary matter and radiation. Within this theory, dark matter is an hypothesized matter component, currently detected only through its gravitational interaction (Blumenthal et al., 1984). The cosmological constant, on the other hand, is a quantity allowed by Einstein's equation to set the residual curvature of spacetime when devoid of matter and energy (Carroll et al., 1992). It acts like a negative pressure, and is responsible for the accelerated expansion of the universe detected in high-redshift supernova experiments, although it was initially postulated by Einstein to maintain the universe static, before he learned about Hubble's result on the expansion of the universe. Although our physical understanding of these last two components, this concordance cosmological model has become a standard nowadays, and has been tested extensively against data from supernovae, the cosmic microwave background, large-scale structure and light element abundances (Spergel, 2007). With this combination of observations, currently the parameters determining the model are known to a few percent.

Historically, the analysis of the spatial clustering of galaxies has been pursued with the aim of determining cosmological parameters, and the two-point correlation function (described in chapter 1) and its variants (such as the variance of the galaxy counts in cells) have been the preferred quantity to describe it. The first wide surveys in which these analyses were systematically performed were the Shane-Wirtanen catalog (or Lick Observatory catalog) and the Zwicky catalog. Limber (1954) first analysed the galaxy counts in Lick catalog and put constraints on the spatial correlation function $\xi(r)$ of galaxies having magnitudes $B \lesssim 19$. During the 70s and 80s, Lick and Zwicky's surveys were both extensively studied (Peebles & Hauser, 1974; Groth & Peebles, 1977, 1986b,a), and the two major conclusions from these studies were: first, that the data were consistent with a power-law autocorrelation function $xi(r)$, with a correlation length ~ 5 Mpc, and second, that at large scales (above ~ 10 Mpc) the power-law behaviour breaks and the correlation function drops to zero. These results were confirmed in the CfA redshift survey (Davis & Peebles, 1983), with a direct calculation of the spatial correlation function of galaxies with $B < 14.5$. Although these findings were confirmed by the experiments that followed, the exact position of this break was a matter of debate, but more recent surveys have clarified these discrepancies, most notably the APM survey (Maddox et al., 1996), the CfA survey (Huchra et al., 1990), the LCRS (Las Campanas redshift survey) (Shectman et al., 1996) and the COSMOS survey (Collins et al., 1989, 1992). The cosmological importance of the shape of the correlation function is that, given additional measurements of the cosmic microwave background (CMB) fluctuations, the Hubble parameter and rich cluster abundances, it constrains the total amount of matter allowed in the model. The position of this break in the correlation function scales with the parameter $\Gamma = \Omega_m h$, which by the mid-nineties was already constrained to be within 0.2-0.3. This fact ruled out the old, high density CDM cosmology that was the standard during the 1980s ($\Omega_m = 1$ and $h = 0.5$), in favor of a less dense CDM model, having $\Omega_m = 0.2$-0.3.

For this investigation, however, we will instead take advantage of the current precision to which the cosmological parameters and the power spectrum of the CMB fluctuations are known, and within this framework deduce physical properties of the dark matter halos hosting starforming galaxies at $z \sim 3$. We will attempt to answer questions like: What are the typical masses of these halos? How many of them

3

are there and how many are actually being occupied by the galaxies? How does the clustering strength of each of these galaxy populations compare to other known galaxy samples? What are the possible ancestors and descendants of these galaxies, given the behavior of their hosting halos? The dataset for this study will come from the MUltiwavelength Survey by Yale-Chile (MUSYC), which is a survey designed to study galaxies at $z \sim 3$, with deep optical imaging in the optical bands UBVRIz. In the Extended Chandra Deep Field South (ECDF-S), one of the four fields in the survey, there is also deep coverage with a narrow band filter at 5000 Å, and a unique X-ray dataset from Chandra. Additionally, there is an extensive spectroscopic follow-up program in the four fields of the survey. This dataset is especially suited for our purposes of studying clustering, given the wide coverage at this depth, and the uniformity of the data.

This thesis is organized as follows: in chapter 1, we start by describing the statistical measure we will use to measure galaxy clustering, as well as the techniques used to estimate it from the samples. We also describe the theoretical methods that will be used to infer the physical properties of the dark matter halos hosting the galaxy population examined in this work. Chapter 2 describes the dataset that will be employed in this thesis, a part of MUSYC. We give a general description of the main optical and X-ray imaging data, and also describe the spectroscopic data, including the reduction process. In chapter 3 we analyse the clustering of Lyman-Break Galaxies (LBG) at $z \sim 3$ found in MUSYC. We describe the general method for selection, and we put constraints on possibles systematic errors that might affect this measurement. We give our estimates for dark matter halo properties, compare our results to other works, and put constraints on the possible ancestors and descendants according to the theoretical framework described in chapter 1. Chapter 4 and 5 describe the analogous procedure followed with Active Galactic Nuclei (AGN) and Lyman-Alpha Emitters (LAE) also found in MUSYC. In our final chapter we summarize our main conclusions and describe some of the possible spin-offs of this work. All through this work we assume a ΛCDM cosmology consistent with WMAP results (Spergel, 2007) with $\Omega_m = 0.3, \Omega_\Lambda = 0.7$, $H_0 = 70$ km s^{-1} Mpc^{-1} and $\sigma_8 = 0.8$. All quantities are comoving: correlation lengths scale as h_{70}^{-1}, number densities as h_{70}^3 and halo masses as h_{70}^{-1}. Unless explicitly stated otherwise, all optical magnitudes are in the AB system.

4

Chapter 1

The measurement of clustering

1.1 The two-point correlation function

1.1.1 Definition

For a dataset of object coordinates in a given space, the two-point correlation function $\xi(r)$ is defined as the excess probability of finding two points in the volume elements δV_1 and δV_2 separated by a distance r_{12}:

$$\delta P_{12} = \rho^2(1 + \xi(r_{12}))\delta V_1 \delta V_2 \qquad (1.1)$$

where ρ is the mean number density of data points and δP_{12} is the number of pairs between the volume elements. The excess is defined relative to a uniform random Poisson distribution, for which $\xi(r)$=0. If object positions are correlated at scale r, $\xi(r) > 0$, and if they are anticorrelated, i.e. they avoid each other at that scale, $\xi(r) < 0$. The angular two-point correlation function is defined in the same way:

$$\delta P_{12} = \Sigma^2(1 + \omega(\theta_{12}))\delta\Omega_1 \delta\Omega_2 \qquad (1.2)$$

where now the separation θ_{12} is angular, Σ is a surface density and the differentials are solid angle elements. Figures 1.1 and 1.2 show examples of different random fields with their corresponding two-point correlation functions.

Naturally, this statistic captures only part of the information within the field, and many different configurations can yield the same correlation function. To get a full description of a given field, one would have to consider the whole set of N-point correlation functions, which are defined in an analog way, by means of point triplets, quadruplets, etc (Peebles, 1980). In the case of a gaussian field, all these higher-point statistics can be determined exactly from the two-point case. Since the maps of fluctuations in the Cosmic Microwave Background (CMB) and in Large Scale Structure (LSS) surveys have been measured to be nearly gaussian, historically authors have disregarded these higher-order measurements. For the high redshift populations we are concerned about in this study, the two-point function will capture all the spatial information we need and can reliably measure.

The two-point correlation function and variants from it have been fairly popular among authors for decades: e.g. the variance in the number counts in cells, used as early as in Hubble (1934), is an integral of the autocorrelation function. One reason for its popularity is that it is a simple, robust statistic, and it is very direct to think about the position of galaxies in space as a point process. Another reason is due to its simple deprojection from two to three dimensions. Although very basic, distance to an astronomical object is one of the toughest quantities to obtain, and Limber (1953) showed that there is a simple linear integral equation relating the angular correlation function and the spatial one, as we will see in section 1.3. This allows the decoupling of the distance distribution information from the angular distribution in the sky, something very appealing when dealing with real galaxy surveys.

When we have more than one galaxy population which spatially overlap, then we can define, in an analogous way, the cross-correlation function:

$$\delta P_{AB} = \rho_A \rho_B (1 + \xi(r_{AB})) \delta V_A \delta V_B \qquad (1.3)$$

which is the excess probability of finding a galaxy of type A and another galaxy of
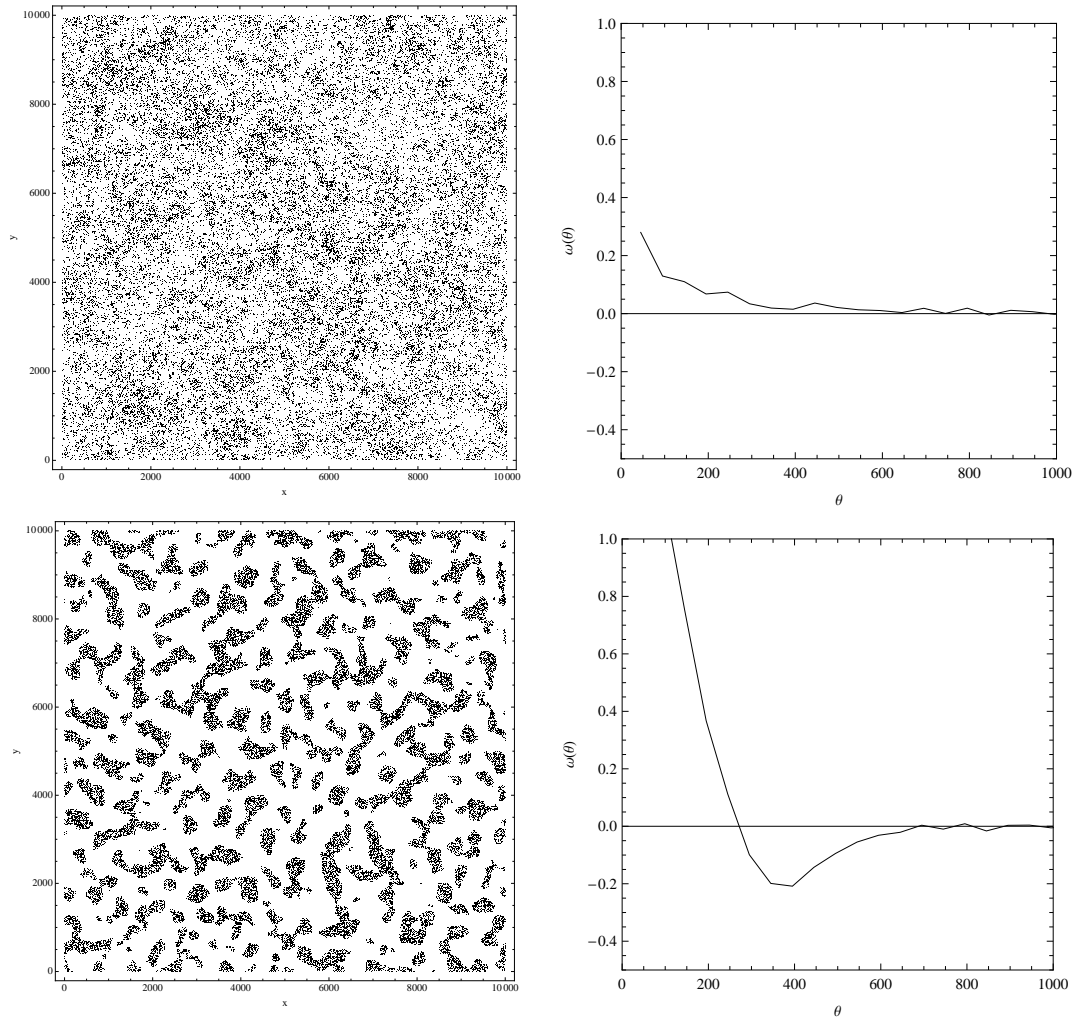
Figure 1.1 Simulated 2D fields with their corresponding two-point correlation function. Distance units are arbitrary. Each sample has exactly 50000 points, and some poisson noise can be seen in the measured correlation function. *Top panels:* field with a power-law behavior, similar to the one observed in the distribution of real galaxies. *Bottom panels:* a field where points tend to cluster on scales $\theta \lesssim 260$ and tend to avoid each other at scales $260 \lesssim \theta \lesssim 700$.
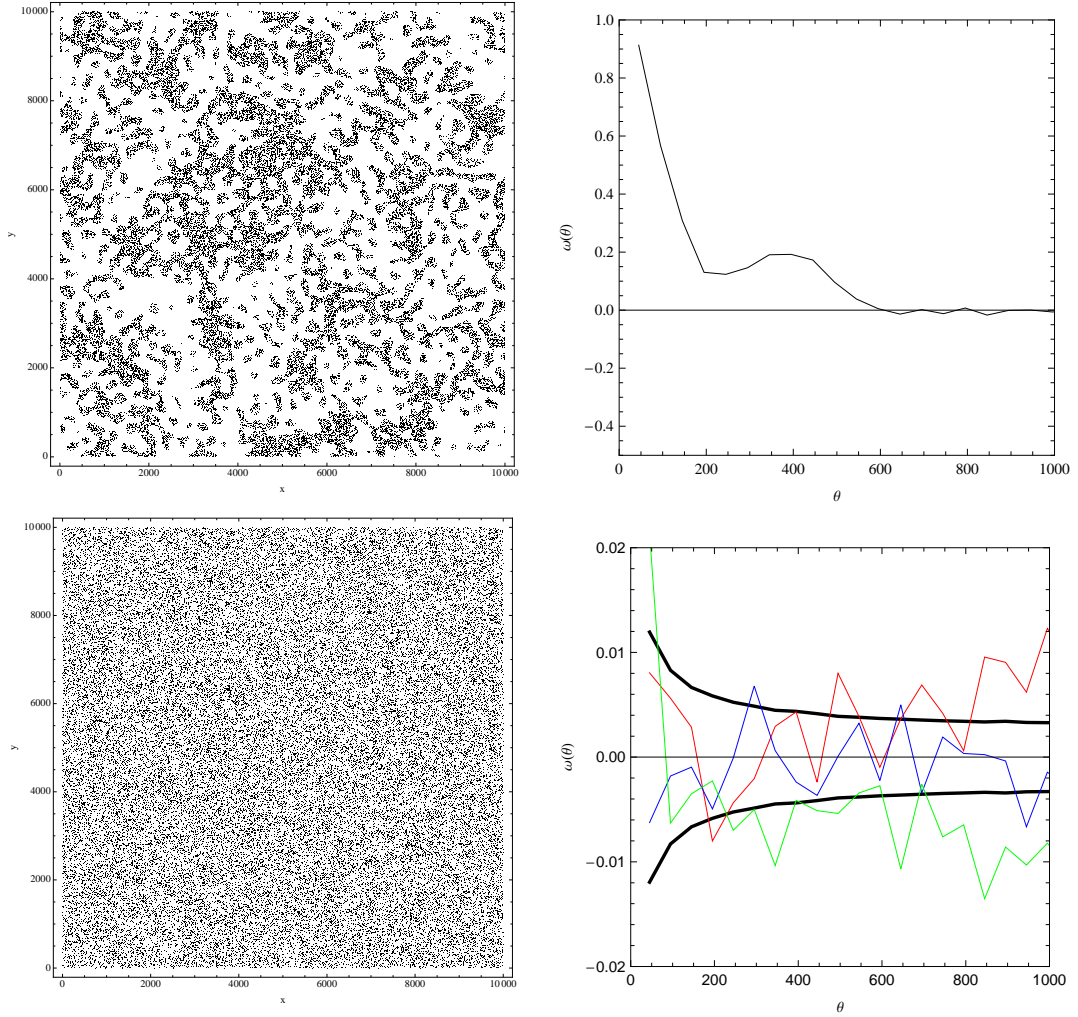
Figure 1.2 *Top panels:* field with clustering strength concentrated on two scales: sources tend to cluster at $\theta \lesssim 200$ and around $\theta \sim 400$. *Bottom panels:* reference field with a pure poisson point distribution, which by definition has $\omega(\theta) = 0$. The red, green and blue curves correspond to three measurements of $\omega(\theta)$ in three realizations of this random field with 10000 points. The black thick curve shows the standard deviation at each θ estimated from 1000 independent realizations. The noise in the measured autocorrelation function at this scale is basically poisson in the number of pairs that fall within each θ bin, but the bins are correlated. As can be deduced by the change in the y-scale, this noise is hard to see in any of the other plots due to the high number of points used in each field.

type B at a distance r_{AB} from each other. Now we have a density for each type, and the volume elements belong to different survey volumes. This occurs because in general, the survey for A galaxies doesn't have to coincide exactly with the survey for B galaxies. To measure $\xi(r_{AB})$ we need to design galaxy selection techniques that make this spatial overlap as big as possible, since it is from the intersection region where the signal is going to come from.

The definition of the angular cross-correlation function is analogous:

$$\delta P_{AB} = \Sigma_A \Sigma_B (1 + \omega(\theta_{AB})) \delta\Omega_A \delta\Omega_B \qquad (1.4)$$

Notice that in this case, the dependence of $\omega(\theta_{AB})$ on the amount of spatial overlap of the two samples is less obvious. Consider the following special situation: we have two sets of galaxies which have a strong spatial cross-correlation function at small scales, but that also decays rapidly at larger distances (as in any realistic case); and let the galaxies be within a couple of surveys that only overlap in their angular projection on the sky, but are actually at very different distances from the observer. Naturally any attempt to measure $\omega(\theta_{AB})$ would yield a null result, since the actual spatial distances being probed are as large as the spatial separation of the two survey volumes, and at those distances $\xi(r_{AB}) \sim 0$. This measurement says nothing about $\xi(r_{AB})$, as it will be clear after we look at the deprojection formula in section 1.3, but it rather leaves the clustering strength undetermined. From this reasoning we can also gain some insight on the effect of the survey shape on the angular autocorrelation function: if a survey is very elongated in the radial direction, we can obtain a fairly small amount of angular clustering, even if the spatial clustering is strong. This is just because in the projection from ξ to ω, at the small θ scales where most of the power is going to come from, we are adding together close pairs with pairs that come from large scales, and those last ones basically only add poisson noise.

One of the typical uses of the cross-correlation function is to allow the measurement of clustering strength for small populations of objects. As we will see later on, the autocorrelation functions of the individual populations are related to the cross-correlation function between them, and so even when the sample might be too small to allow the measurement of the autocorrelation function for a given set of galaxies,

we can still measure the cross-correlation function between those and another, more numerous set of neighboring galaxies. Since the bigger set of galaxies has presumably had its autocorrelation function measured with much greater precision and accuracy, we can deduce the clustering strength for the small sample from these two quantities. We will take advantage of this method in chapter 4.

From a more theoretical point of view, if we know the continuous density field $\rho(\boldsymbol{x})$ that generates the point distribution we have been talking about up to this point, we define the density contrast as:

$$\delta(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x}) - \bar{\rho}}{\bar{\rho}}, \tag{1.5}$$

where the coordinate $\boldsymbol{x}$ is in comoving coordinates and $\bar{\rho}$ is the average density of that field in the universe. The autocorrelation function of δ is obtained directly from the more familiar statistical definition:

$$\xi(\boldsymbol{r}) = \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x} - \boldsymbol{r}) \rangle \tag{1.6}$$

where the expectation value is taken over the $\boldsymbol{x}$ coordinate. Analogously, when we have two populations, the cross-correlation function takes the form:

$$\xi_{1,2}(\boldsymbol{r}) = \langle \delta_1(\boldsymbol{x})\delta_2(\boldsymbol{x} - \boldsymbol{r}) \rangle \tag{1.7}$$

These relations will be useful later on, when we connect the statistical properties the dark matter halos with their physical properties.

1.1.2 Measurement

Statistical estimates of the two-point correlation function are usually constructed from pair counts, which for the case of the angular autocorrelation function are

three:

$$DD(\theta) = \frac{\text{\# of unique galaxy pairs separated by } \theta}{n(n-1)/2}$$

$$DR(\theta) = \frac{\text{\# of galaxy-random points pairs separated by } \theta}{nn_r}$$

$$RR(\theta) = \frac{\text{\# of unique random point pairs separated by } \theta}{n_r(n_r-1)/2}$$

where n is the number of galaxies in the data catalog and n_r is the number of points in the random catalog. The latter is supposed to describe the survey shape, in the sense that its distribution should be the result of observing a field of points with uniform, poisson distribution over the same area, and in the same way the galaxy field was observed. These three functions are also normalized so that they integrate to unity between 0 and the maximum separation allowed by the shape of the survey, i.e. for a catalog in the shape of a perfect square of side A, they integrate to one between 0 and $\sqrt{2}A$. Notice that this last point puts a constraint on DD realtive to RR: if at some scale DD goes above RR, at some other scale it necessarily has to go below it. This issue will appear again when we refer to the "integral constraint" condition that appears while calculating the expectation value of statistical estimators of the correlation function.

The most simple and intuitive estimator for $\omega(\theta)$ can be inferred from isolating $\omega(\theta)$ in the equation of its own definition:

$$\omega(\theta_{12}) = \frac{\delta P_{12}}{\Sigma^2 \delta\Omega_1 \delta\Omega_2} - 1. \tag{1.8}$$

Since the quantity in the denominator is the number of pairs expected for a uniform random Poisson distribution, from this we can directly infer the estimator:

$$\widehat{\omega}(\theta) = \frac{DD(\theta)}{RR(\theta)} - 1, \tag{1.9}$$

which basically compares how the galaxies cluster around each other compared to an uncorrelated field having the same geometry as the galaxy survey. This estimator works fine and has been used extensively in the past, however, Landy & Szalay (1993) proposed another estimator, which has the same espectation value but a

lower variance, and thus more efficient:

$$\widehat{\omega}_{LS}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}. \tag{1.10}$$

To estimate the cross-correlation function, the estimators are analogous, but use cross-pairs instead. In chapter 4 we will use the following estimator:

$$\widehat{\omega}_{AB}(\theta) = \frac{D_A D_B(\theta)}{D_A R_B(\theta)} - 1. \tag{1.11}$$

where A is the small galaxy set and B the big set, and the pairs:

$$D_A D_B(\theta) = \frac{\# \text{ of galaxy A - galaxy B pairs separated by } \theta}{n_A n_B}$$

$$D_A R_B(\theta) = \frac{\# \text{ of galaxy A - random points B pairs separated by } \theta}{n_A n_{r,B}}$$

This now compares how the numerous set of galaxies (D_B) cluster around the small one (D_A), relative to an uncorrelated field having the same survey shape as the big set of galaxies (R_B). Now n_B, n_A and $n_{r,B}$ are the number of data points in each of these catalogs. The estimator given by eq. 1.11 is not the only option, but as far as the author knows, theoretical work regarding statistical properties of cross-correlation estimators is still lacking when compared to autocorrelation estimators. Nonetheless, eq. 1.11 has been used by many authors succesfully for datasets with characteristics similar to the ones we will be concerned with in chapter 4.

If we have full distance information, the estimators in three dimensions are formally the same, so we would simply replace θ by r and $\omega(\theta)$ by $\xi(r)$ in the equations above. Notice however, that in practice what is usually done is to separate the 3D catalog in distance slices, and then calculate the projected correlation function on each slice. This helps alleviate the "fingers of god" problem, which elongates structures artificially in the radial direction due to the large peculiar velocities of galaxies in dense regions.

1.1.3   The Integral Constraint

The effect of computing the clustering from a finite catalogue, with a mean density slightly different than the average in the universe, has long been recognized as a non-negligible bias that had to be accounted for. Landy & Szalay calculated the expectation value and the variance of their estimator (1.10) and obtained:

$$\langle \widehat{\omega}_{LS}(\theta) \rangle = \frac{\omega(\theta) - \omega_{\Omega}}{1 + \omega_{\Omega}} \tag{1.12}$$

$$\sigma^2_{LS}(\theta) = \frac{(1 + \langle \omega_{LS}(\theta) \rangle)^2}{(n(n-1)/2)RR(\theta)} \tag{1.13}$$

where n is the number of galaxies in the field, Ω is the field's area (or volume) and

$$\omega_{\Omega} = \frac{1}{\Omega^2} \int \int_{survey} \omega(\theta_{12}) d\Omega_1 d\Omega_2 = \int_0^{\theta_{MAX}} RR(\theta)\omega(\theta)d\theta \tag{1.14}$$

is the integral constraint factor, that in the calculation from Landy & Szalay appears simply as a normalization constant[1] (see equations 23-28 from Landy & Szalay, 1993). The second term in equation 1.14 give us a way to estimate this quantity from the random-random pairs, given a clustering strength $\omega(\theta)$ (Infante, 1994). The reason the factor ω_{Ω} appears can be understood qualitatively in the following way: let's consider first an unclustered, finite field of galaxies. Let's then increase the amount of clustering strength by moving galaxies closer together into a few clusters, keeping the number of galaxies constant. If the length of these movements is not negligible compared to the size of our observation field, then when we calculate $DD(\theta)$ we will see that DD at small scales increases together with clustering, but diminishes at scales comparable to the size of the field. The normalization we imposed on DD (that it should integrate one) is hiding the assumption that the density of galaxies measured in the survey is the same as the true density of these galaxies in the universe. If the field becomes bigger in comparison with the scales at which the galaxies cluster, then this factor gets smaller, as can be seen in figure 1.3. This is due to the overall distribution of pairs, which shifts to bigger scales, where $\omega(\theta)$ decreases.

---

[1] Notice that we dropped the "12" index after turning the double integration over the field into a single integration over the distance separation, just to simplify notation.
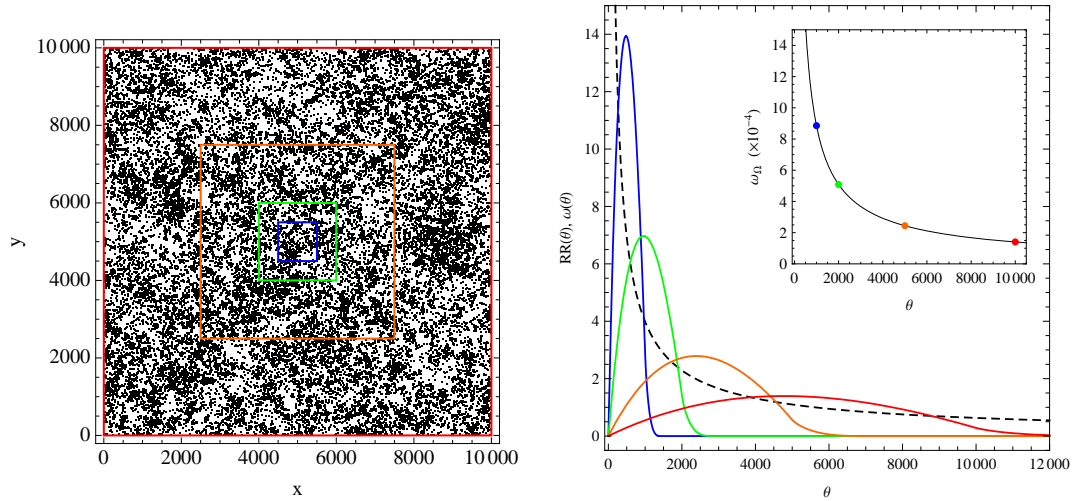
Figure 1.3 *Left*: Field with power-law autocorrelation function from figure 1.1, with square subfields drawn over it: blue is 1000 on a side, green is 2000, orange 5000 and red is the full field. *Right*: Functions in the integrand of eq. 1.14 are shown separated and for each subfield. Black dashed line is a power-law $\omega(\theta) = A\theta^{-\beta}$, arbitrarily scaled in the y-axis. Blue, green, orange and red curves are the different $RR(\theta)$ that correspond to each of the subfields. The inset shows the integral constraint that results from calculating 1.14 for a square field such as the ones shown here, as a function of the side of the square. The color points correspond to the particular subfields shown to the left. Notice that in the case of a power law correlation function, we always have $\omega(\theta) > 0$, therefore we will obtain $\omega_\Omega > 0$ for any survey.

Although many authors use Landy & Szalay's estimator, most do not fit their data with the expectation value calculated by Landy & Szalay, but use instead:

$$\langle \widehat{\omega}_{LS}(\theta) \rangle \approx \omega(\theta) - \omega_\Omega \qquad (1.15)$$

where the integral constraint constant ω_Ω is estimated from (1.14). The same approximation can also be found by looking at the same issue from another perspective. Hamilton (1993) showed that the following estimator has an expectation value of exactly $\omega(\theta)$:

$$\widehat{\omega}_{\text{ideal}}(\theta) = \frac{DD(\theta)(1+\delta)^2 - 2DR(\theta)(1+\delta) + RR(\theta)}{RR(\theta)} \qquad (1.16)$$

where δ is the unknown value of the mean contrast in the survey $(n_{obs} = \langle n \rangle(1+\delta))$, following the notation in Adelberger et al. (2005b). That is, if we knew exactly how much our measured galaxy density differs from the true density in the universe, we could "correct" the data-data and data-random pairs so that the Landy & Szalay estimator yields exactly $\omega(\theta)$, i.e. when $\omega_\Omega = 0$. Since we do not know the value of δ, this is only an idealized case.

If we want to estimate this bias of the Landy & Szalay estimator, we can rewrite eq. (1.16) to analyze the difference between the true angular correlation function and the estimated one. This difference is called I by Adelberger et al. (2005b) and reads:

$$\widehat{\omega}_{\text{ideal}}(\theta) - \omega_{LS}(\theta) = I = \frac{DD(\theta)\delta^2}{RR(\theta)} + 2\frac{(DD(\theta) - DR(\theta))\delta}{RR(\theta)} \qquad (1.17)$$

This is still a random variable, but we need to know the average over many statistical realizations. To calculate the expectation value for this, we will assume that δ is statistically independent from DD and DR at the scales of interest. Since $\langle \delta \rangle = 0$, the final term vanishes and we get the result that appears in Adelberger et al. (2005b), eq. (13):

$$\langle I \rangle = \omega(\theta) - \langle \omega_{LS}(\theta) \rangle = \frac{\langle DD(\theta) \rangle \langle \delta^2 \rangle}{RR(\theta)} \qquad (1.18)$$

The usual approximation at this point is to make $\langle DD \rangle \approx RR$ and get a constant for $\langle I \rangle$, since $\langle \delta^2 \rangle$ is already small. But the DD pairs mentioned here are the *observed*

pairs, and therefore the more consistent thing to do will be to include δ:

$$\langle DD(\theta)(1 + \delta^2) \rangle = (1 + \omega(\theta))RR(\theta) \tag{1.19}$$

$$\implies \frac{\langle DD(\theta) \rangle}{RR(\theta)} = \frac{(1 + \omega(\theta))}{1 + \langle \delta^2 \rangle}. \tag{1.20}$$

This means that

$$\langle I \rangle = \frac{(1 + \omega(\theta))\langle \delta^2 \rangle}{1 + \langle \delta^2 \rangle}, \tag{1.21}$$

and also that

$$\langle \omega_{LS}(\theta) \rangle = \frac{\omega(\theta) - \langle \delta^2 \rangle}{1 + \langle \delta^2 \rangle}. \tag{1.22}$$

Here we can identify $\langle \delta^2 \rangle = \omega_\Omega$. This tells us that the formula for ω_{LS}'s expectation value, as presented in Landy & Szalay (1993), is more consistent at this approximation level. Although ω_Ω is small, we prefer to avoid the Taylor series approximation that would yield to first order (1.15), since it's unnecessary.

1.1.4 Fitting procedure

To fit the measured angular correlation function, we assume the usual power law model for the true $\omega(\theta) = A\theta^{-\beta}$, and numerically evaluate eq. (1.14) for each (A,β). Naturally, this makes σ_{LS} and $\langle \omega_{LS} \rangle$ dependent on the parameters not only through $\omega(\theta)$, but also through ω_Ω. The best fit parameters are obtained with a traditional χ^2 minimization, and confidence intervals are obtained both from $\Delta\chi^2$ contours and jackknife estimates. Additionally, since $\omega(\theta)$ in the former equations is not the estimated but the true one, many people have tried an iterative scheme, as suggested by equations (1.12) and (1.14), to find the best fit to the model parameters. That method can be unstable and is not necessary. Instead, we replace the whole integral in the right hand side of (1.14) inside (1.12) and fit this complete expression to the measured ω_{LS} with a χ^2 minimization. This way the value for the obtained integral constraint is automatically consistent with the best fit model parameters.

1.1.5 Random Catalogs

The random catalogs used to compute the DR and RR counts have 100 times the number of point of the corresponding data catalog, to make the variance due to this random catalog negligible compared to the data catalog, and they are generated following the geometry of each of the fields, that is, excising regions blocked by bright stars. To accomplish this, we build a mask for each field, using the optical catalog itself in the following way:

1. We construct a matrix with a size 1/20 of the optical image, and put a 1 in any position of the matrix where the corresponding position in the catalog has at least one object with good photometry, otherwise we put 0.

2. We apply a moving average filter of length 1 pixel over each pixel of the matrix.

3. We assign a value of 0 for every pixel that is below 1.5σ of the mean value in the matrix, and 1 otherwise.

4. We finally apply a moving median filter of 2 pixels of length.

The first step simply makes a low resolution density map of the catalog. Although we could use a matrix of the same size of the image, we do not need so much resolution for this mask. The second step makes this map smoother, and the third step marks the lowest density regions. The last step is effective in removing noise and making the regions around stars grow slightly. At the end we have matrix of 0's and 1's, which mark the positions where no objects are present in the catalog.

This whole procedure is very ad-hoc and the parameters mentioned above have been fine tuned to visually reproduce the regions with no objects in the catalogs. Changing these parameters slightly does not result in great changes in the final clustering measurement, though, and a further refinement of the random catalog at the smallest details does not change the $RR(\theta)$ and $DR(\theta)$ functions in any noticeable way. Figure 1.4 shows the mask for the E-CDFS field as an example.
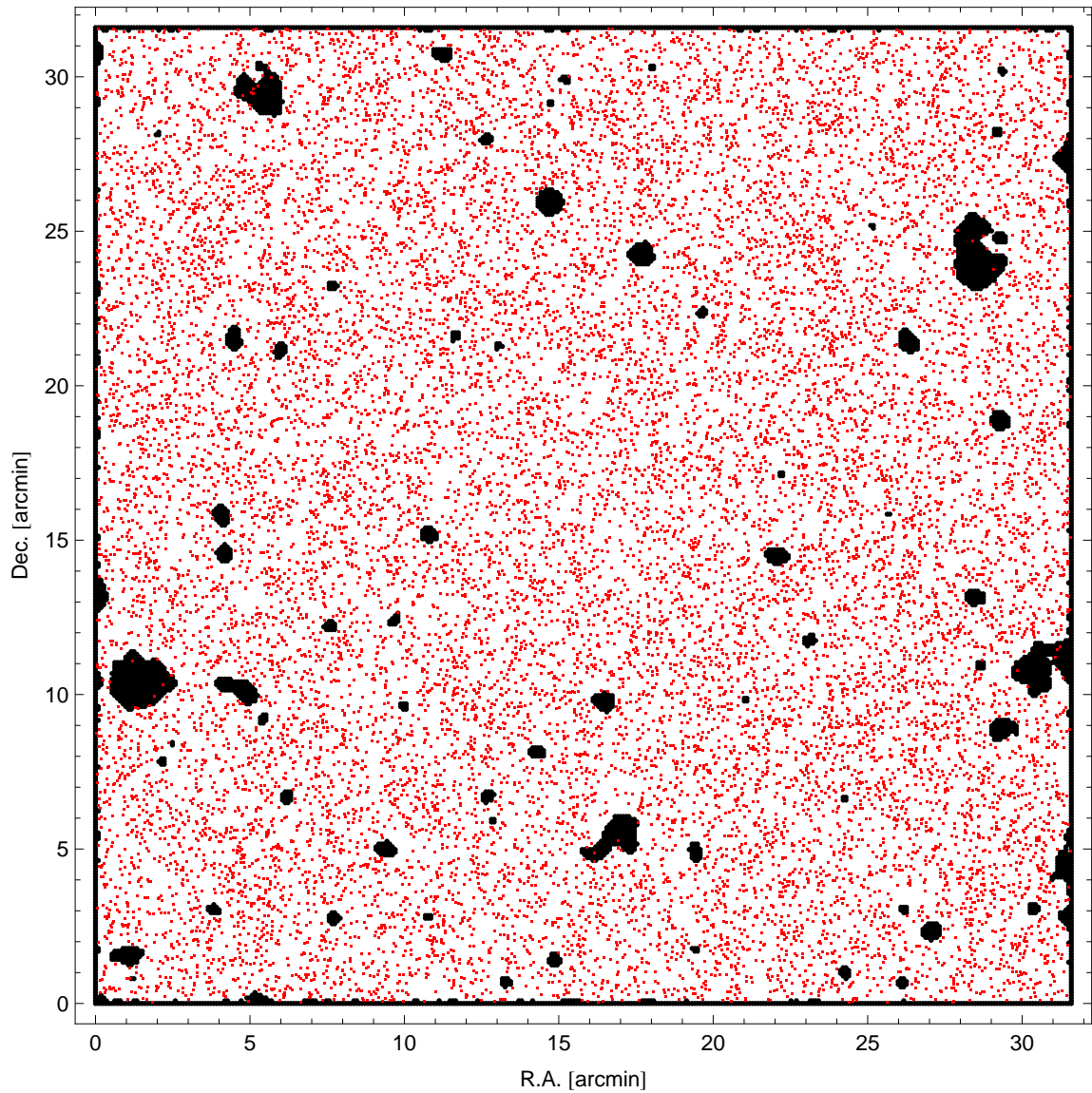
Figure 1.4 Mask for the generation of random catalogs in E-CDFS. Red dots are the positions of 20000 randomly selected objects from the optical catalog. The mask follows the geometry of the field fairly accurately.

1.2 Jackknife resampling

Although during the χ^2 minimization procedure we obtain confidence intervals for our fitting parameters of the angular correlation function, we would like to obtain these intervals from other methods too, in order to check for consistency. This is relevant since there is no guarantee that the errors are independent and normally distributed. As a matter of fact, the correlations between datapoints produce also correlations between different distance bins.

The jackknife resampling technique consist in dividing the full sample in N subsamples, and estimate the quantity of interest N times, each time excluding one of the subsamples. The variance in this very narrow distribution is then amplified by a factor $\sim$N to obtain an estimate of the real variance of the quantity measured in the full sample. See Lupton (1993) for a discussion on the approximations involved, and Zehavi et al. (2002) for an application on clustering analysis.

We have divided each of our 4 fields in 25 pieces, for a total of N=100 jackknife realizations, from which we estimate the variance in each quantity as

$$\sigma_t^2 = \frac{N-1}{N} \sum_{i=1}^{N} (t_{N-1,i} - \overline{t_{N-1}})^2, \tag{1.23}$$

where t can be $\omega(\theta)$ or the fit parameters (A, β), and $t_{N-1,i}$ is the estimation calculated excluding the i-th subset. Notice that this expression can be rewritten as

$$\sigma_t^2 = \frac{1}{N-1} \sum_{i=1}^{N} (\tilde{t}_{N,i} - t_N)^2 \tag{1.24}$$

$$\tilde{t}_{N,i} = t_N + (\sqrt{N} - 1/\sqrt{N})(t_{N-1,i} - \overline{t_{N-1}}), \tag{1.25}$$

i.e. the standard sample variance but computed over an artificial sample of jackknife realizations $\tilde{t}_{N,i}$. We have found this last quantity useful for plotting purposes when comparing the two dimensional confidence limits in the (A, β) plane. Additionally, we have also found that the linear extrapolation in the above formulas delivers more consistent results when using the variable log A instead of A, since the degeneracy

between A and β is linear in that space. The distributions of the parameters fitted to these realizations have also significantly long tails of outliers, so we used the median absolute deviation (MAD) to estimate σ_A and σ_β:

$$\sigma_t = 1.4826 \times \mathrm{median}(\tilde{t}_{N,i} - \mathrm{median}(\tilde{t}_{N,i})), \qquad (1.26)$$

where the numerical factor is necessary to obtain the standard deviation from the MAD for a gaussian distribution.

1.3 Deprojection of angular correlation functions

Naturally, the quantity that is physically meaningful is not the projected correlation function, but the spatial correlation function $\xi(r, z)$, since the former depends on the survey. If the time evolution of $\xi(r, z)$ is negligible within the redshift range of the sample, then the projection into $\omega(\theta)$ is given by: (Simon, 2007)

$$\omega(\theta) = \int_0^\infty \int_0^\infty dr_1 dr_2 p_1(r_1) p_2(r_2) \xi\left(\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta)}\right) \qquad (1.27)$$

where p_1 and p_2 are the comoving radial distance distributions of the samples being cross-correlated and in the case of deprojection of autocorrelation functions, $p_1 = p_2$. If the depth of the survey is much greater than the spatial correlation length, then one is allowed to use Limber's approximation of this integral (Limber, 1953, Peebles, 1980). If, in addition, the angular sizes in the catalog allow the small angle approximation, and we adopt the power law models $\omega(\theta) = A\theta^{-\beta}$ and $\xi(r) = (r/r_0)^\gamma$, this equation simplifies to:

$$A = r_0^\gamma B\left(\frac{1}{2}, \frac{\gamma - 1}{2}\right) \int_0^\infty d\bar{r}\ \bar{r}^{1-\gamma} p_1(\bar{r}) p_2(\bar{r}) \qquad (1.28)$$

and $\gamma = \beta + 1$. B(x,y) is a Beta function. To deproject the spatial clustering strength from this equation, we fit the comoving distance distribution for the spectroscopically confirmed galaxies with a smooth interpolating polynomial.

For error propagation purposes, we approximate the distributions of comoving distances with gaussians, and estimate the random error in r_0 due to the uncertainty in the first two moments of the distributions. Let's define:

$$J_k = \frac{1}{2\pi\sigma_1\sigma_2} \int_0^\infty dx\ x^{1-\gamma+k} \exp - \left(\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2} \right) \qquad (1.29)$$

where μ_1, μ_2 are the means and σ_1, σ_2 the standard deviations measured from the samples of p_1 and p_2. Propagating the standard error in the mean and the standard deviation we obtain:

$$
\begin{aligned}
\Delta J_{\mu,i} &= \frac{\partial J_0}{\partial \mu_i} \frac{\sigma_i}{\sqrt{N_i}} = \frac{J_1 - \mu_1 J_0}{\sqrt{N_i}\sigma_i} \\
\Delta J_{\sigma,i} &= \frac{\partial J_0}{\partial \sigma_i} \frac{\sigma_i}{\sqrt{2N_i}} = \frac{J_2 - 2\mu_i J_1 + (\mu_i^2 - \sigma_i^2)J_0}{\sqrt{2N_i}\sigma_i^2} \\
\Delta r_{0,p} &= \frac{r_0}{\gamma J_0} \sqrt{\Delta J_{\mu,1}^2 + \Delta J_{\mu,2}^2 + \Delta J_{\sigma,1}^2 + \Delta J_{\sigma,2}^2}
\end{aligned}
$$

This last error is added in quadrature with the errors coming from the A and β parameters from the fit to the angular correlation function:

$$\Delta r_0 = \sqrt{\Delta r_{0,p}^2 + \Delta r_{0,A,\beta}^2}. \qquad (1.30)$$

To correct for interlopers in the case of autocorrelation functions, we can make the following replacements in eq. (1.27):

$$p_1(r) = p_2(r) = f p_i(r) + (1-f) p_g(r) \qquad (1.31)$$

$$\xi(r) = \begin{cases} \xi_i(r) & p_i(r) > 0 \\ \xi_g(r) & p_g(r) > 0 \end{cases} \qquad (1.32)$$

Assuming that p_i and p_g do not overlap. Here f is the interloper fraction, p_i is the interloper distribution, p_g is the distribution of sources of interest, $\xi_i(r)$ is the interloper correlation function and $\xi_g(r)$ is the correlation function of the sources of interest. We then obtain the following relation:

$$\omega_g(\theta) = \frac{\omega_{obs}(\theta) - f^2\omega_i(\theta)}{(1-f)^2} \qquad (1.33)$$

Naturally, if the contaminants are stars, $\omega_i = 0$, and this simplifies to a constant correcting factor $(1 - f)^{-2}$. If the the contaminants are galaxies, on the other hand, they will show some clustering. We can rewrite the previous equation in terms of the ratio ω_i/ω_g to see more clearly the systematic effect introduced by the angular clustering of the interlopers:

$$\omega_g(\theta) = \frac{1}{(1 - f)^2 + f^2 \left(\frac{\omega_i(\theta)}{\omega_g(\theta)} \right)} \omega_{obs}(\theta) \tag{1.34}$$

Although typically the clustering strength of the interlopers $\omega_i(\theta)$, is unknown, we can put limits on this quantity and the systematic uncertainty due to this issue can be constrained. Let's consider varying the ratio ω_i/ω_g between 0 and 10, assuming that ω_i has the same slope as ω_g for simplicity. Figure 1.5 shows the correction factor from equation 1.34 as a function of the interloper fraction f. In this figure, for $f = 0.1$ the factor ω_g/ω_{obs} ranges between 1.1 and 1.2 only, but for higher contamination rates it rapidly becomes very uncertain, e.g. for $f = 0.4$ it ranges between 0.5 and 2.8. To avoid this systematic uncertainity we prefer to keep our galaxy samples as clean as possible, specially when the typical interlopers are galaxies.

1.4 The Halo Masses

To obtain physical properties of the dark matter halos from the statistical quantities determined using the techniques just presented, we need to rely on the theoretical predictions for the evolution of inhomogeneities within ΛCDM cosmology. The first complication that arises is the fact that we do not have a robust theory, from first principles, of how galaxies populate their hosting dark matter halos. This issue is not addressed in the present work, because we are concerned with the measurement of the two-point correlation function at scales large compared to the radius of the dark matter halos. At these scales the complex physical processes involving baryonic matter and light, which have an important role in the formation of the galaxies, are not relevant, and only the effects of gravity are noticeable. This means that the model we need to interpret our measurements simply contains dark matter and the
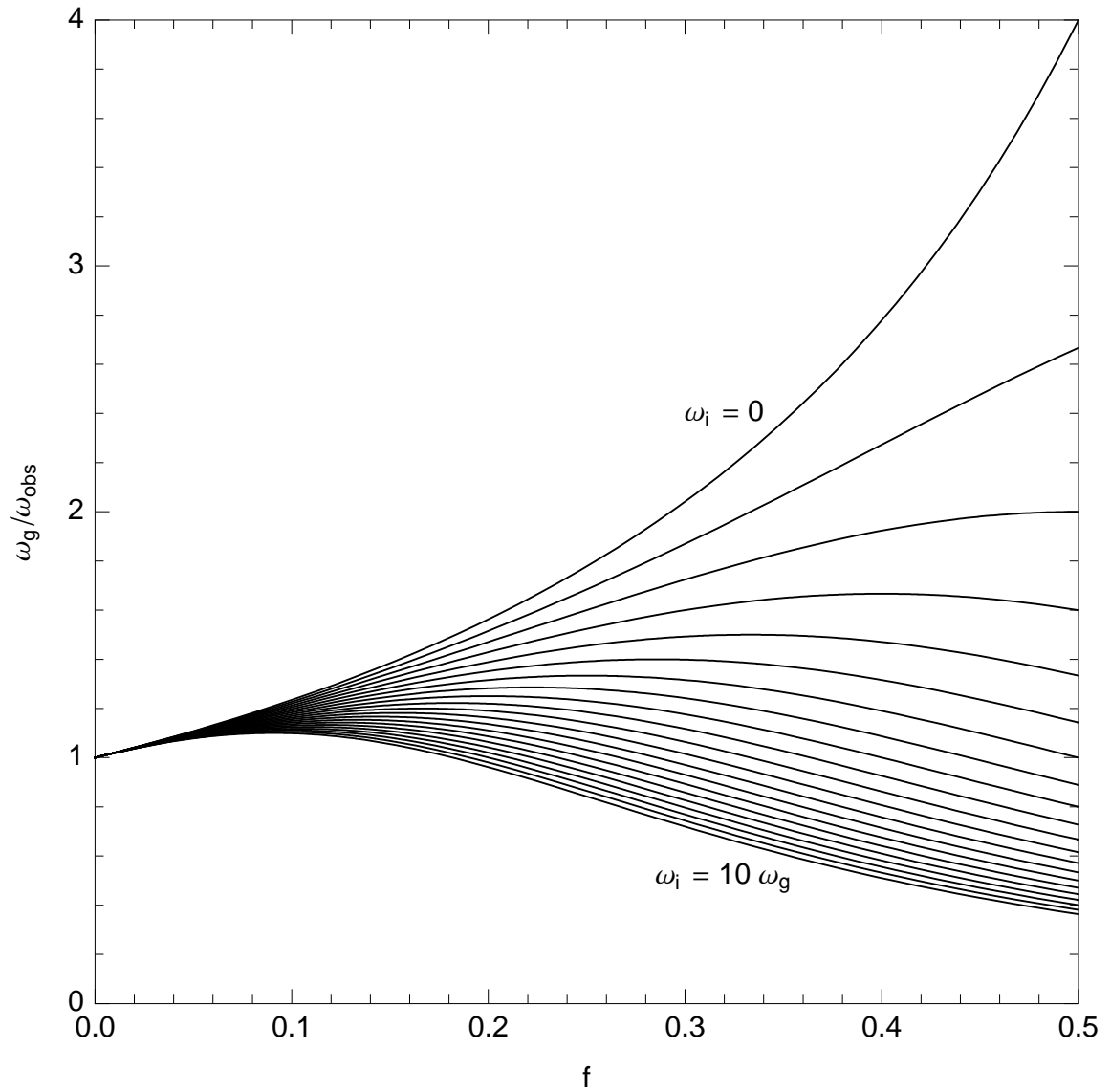
Figure 1.5 Correction factor for the observed angular clustering as a function of contamination fraction. Curve is evaluated for several values of the relative interloper angular clustering ω_i/ω_g, ranging from 0 to 10.

geometry of the universe. Such a model has been studied for decades, and is at present fairly well understood.

In this work we rely on semi-analytical prescriptions based on the model of Press & Schechter (1974) (PS) and several extensions derived from it. PS presented an analytical formula for the mass function of dark matter halos, in which at any given cosmic time, they identify the halos with the regions of the initial density field that have just collapsed at that moment, where the criteria for determining whether a halo has collapsed is based on a spherical infall model. A very powerful reinterpretation of the PS theory is the "excursion set" approach by Bond et al. (1991), in which the identification of collapsed halos is performed in terms of the "trajectories" of the matter field overdensity at a fixed position, as a function of a resolution variable. This interpretation allows not only the recovery of the PS mass function in an elegant way, but also to deduce the conditional mass function for the ancestors and descendants of a halo of a given mass, among other possible extensions of the original PS theory.

1.4.1 The unconditional and conditional mass functions

To understand the ideas mentioned above more quantitatively, let's deduce the unconditional mass function in the simplest approximation of PS, but using the excursion set approach. This will set the ground to easily deduce the conditional mass function later on. The argument in this sections follows the notation in Lacey & Cole (1993) for compatibility with more modern literature, but further details can be found in the original paper by Bond et al. (1991).

Let's consider the constrast field $\delta(\boldsymbol{x})$ and its Fourier decomposition $\delta_{\boldsymbol{k}}$:

$$\delta(\boldsymbol{x}) = \frac{\rho(\boldsymbol{x}) - \bar{\rho}}{\bar{\rho}} = \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}} \exp(i\boldsymbol{k} \cdot \boldsymbol{x})$$

where $\rho(\boldsymbol{x})$ is the density field of the universe and the coordinates $\boldsymbol{x}$ are in comoving units. We will consider a gaussian field, for which the Fourier amplitudes $\delta_{\boldsymbol{k}}$ are

independent random variables with random phases. To define the resolution variable, we will consider also a spherically symmetric window function $W_M(r)$, of radius R and enclosing a mass $M \sim \bar{\rho} R^3$. With this window we will define the smoothed field as the convolution:

$$\delta(M, \boldsymbol{x}) = \int W_M(|\boldsymbol{x} - \boldsymbol{y}|)\delta(\boldsymbol{y})d^3\boldsymbol{y} \tag{1.35}$$

and if we call $\hat{W}_M(k)$ the Fourier transform of the window function:

$$\hat{W}_M(\boldsymbol{k}) = \int W_M(\boldsymbol{x}) \exp(-i\boldsymbol{k} \cdot \boldsymbol{x})d^3\boldsymbol{x} \tag{1.36}$$

Now we can write eq. 1.35 in term of Fourier components:

$$\delta(M, \boldsymbol{x}) = \sum_{\boldsymbol{k}} \delta_{\boldsymbol{k}}\hat{W}_M(\boldsymbol{k}) \exp(i\boldsymbol{k} \cdot \boldsymbol{x}) \tag{1.37}$$

The purpose of defining this smoothed field is to identify halos: at a fixed position $\boldsymbol{x}$, we will look for the maximum enclosed mass M whose density contrast reaches a certain level that we will define as critical for collapse. This ensures that the identified halo is not enclosed in a larger structure that has already collapsed (the "cloud-in-cloud" problem). The "trajectories" mentioned before are the fluctuations of $\delta(M)$ as a function of M for fixed $\boldsymbol{x}$. This is only an analogy, of course, and the space where this trajectories are defined has little to do with the real space where the matter field is defined. To continue this analysis, though, it is more convenient to plot these trajectories as function of the variance at the scale that encloses M instead of M itself:

$$S(M) = \sigma^2(M) = \langle|\delta(M, \boldsymbol{x})|^2\rangle = \sum_{\boldsymbol{k}} \langle|\delta_{\boldsymbol{k}}|^2\rangle\hat{W}_M^2(\boldsymbol{k}) \tag{1.38}$$

Notice that all the possible trajectories will start at $(\delta, S) = (0, 0)$, since S is a monotonically decreasing function of M, and when $M \to \infty$, $S \to 0$. To proceed further, we need to specify the window function W_M. Many choices are available for this, it could be a gaussian, a three dimensional top-hat or any other function that has a "radius" R and is properly normalized. The simplest choice that yields the PS

analytical result is the "sharp k-space filter" (basically, a top-hat in Fourier space):

$$\hat{W}_M(\boldsymbol{k}) = \begin{cases} 1 & |\boldsymbol{k}| \leq k_s(M) \\ 0 & |\boldsymbol{k}| > k_s(M) \end{cases} \tag{1.39}$$

This case simplifies the calculation because when we increase the resolution $S \to S + dS$, we are including new Fourier modes that are completely uncorrelated with the previous set of modes. This can be regarded as particles being released from the origin and following random walks, diffusing away, as the "time" coordinate S increases. Again, the analogy of S with a time coordinate has nothing to with cosmic time, it only has to do with the role of this coordinate in the equation governing this problem. If we denote the density of all trajectories at S in the interval $\delta \to \delta + d\delta$ as $Q(\delta, S)$, then we can write directly:

$$Q(\delta, S) = \frac{1}{\sqrt{2\pi S}} \exp\left(-\frac{\delta^2}{2S}\right) \tag{1.40}$$

Given this simple scheme, we can easily check that the trajectories described this distribution follow a diffusion equation:

$$\frac{\partial Q}{\partial S} = \frac{1}{2}\frac{\partial^2 Q}{\partial \delta^2} \tag{1.41}$$

This is analogous to have a one dimensional heat problem: we have a rod with a source of heat at position $\delta = 0$, and we have the solution for the problem without any special boundary conditions. δ is taking the role of the "spatial" coordinate and S the role of the "time" coordinate. As "time" S passes, the rod gets hotter indefinitely in both directions. Notice that the sharp shape of the filter in k-space makes the steps of the trajectories independent. With other filter shape this is not true, and we would not be able to reduce the problem to a simple random walk and solve the problem analytically with a diffusion solution. For a more detailed analysis of the effect of using more general filters, see Bond et al. (1991).

Now we need to count the trajectories that cross a certain level $\delta = \delta_c$, that we will identify later with the collapse of the halo. The appropiate choice of δ_c and its evolution with cosmic time will be discussed in sec. 1.4.2. In the classic PS approach where the criteria for collapse is obtained from a simple spherical infall model, this

threshold δ_c does not depend on mass, and therefore it is a constant line in the (δ, S) plane. Also, as we will later see, $\delta_c(t)$ decreases with cosmic time, so as the universe evolves, this barrier goes down.

Within the random walk view of the problem, the easiest thing to do is first calculate the distribution of particles that *never* reach the critical level, or "barrier" δ_c, and then deduce the probability for the particles that do get "absorbed" by the barrier. This way we can take advantage of the solution for this random-walk problem discovered by Chandrasekhar (1943):

$$Q(\delta, S, \delta_c) = \frac{1}{\sqrt{2\pi S}} \left\{ \exp\left(-\frac{\delta^2}{2S}\right) - \exp\left(-\frac{(\delta - 2\delta_c)^2}{2S}\right) \right\} \tag{1.42}$$

which is analogous to putting a refrigerator that keeps the temperature fixed at 0 at the position of the barrier. This solution comes from the fact that the trajectories that start at the barrier are equally likely to wander above it as below it (see figure 1.6). Now we can calculate the probability of a trajectory to be absorbed by the barrier in the interval S to $S + dS$, since it must equal the reduction in the number of trajectories that survive below the barrier:

$$f_S(S, \delta_c) = -\frac{\partial}{\partial S} \int_{-\infty}^{\delta_c} Q d\delta \tag{1.43}$$

$$= -\int_{-\infty}^{\delta_c} \frac{\partial Q}{\partial S} d\delta \tag{1.44}$$

$$= -\frac{1}{2} \frac{\partial Q}{\partial \delta}\Big|_{-\infty}^{\delta_c} \tag{1.45}$$

where in the second line we have replaced the diffusion equation inside the integral. Substituting the solution we have in eq. 1.42, we obtain:

$$f_S(S, \delta_c) dS = \frac{\delta_c}{\sqrt{2\pi S^3}} \exp\left(-\frac{\delta_c^2}{2S}\right) dS \tag{1.46}$$

This is the fraction of mass associated with halos having variance between S and $S + dS$, also known as the first crossings distribution. The comoving density of halos
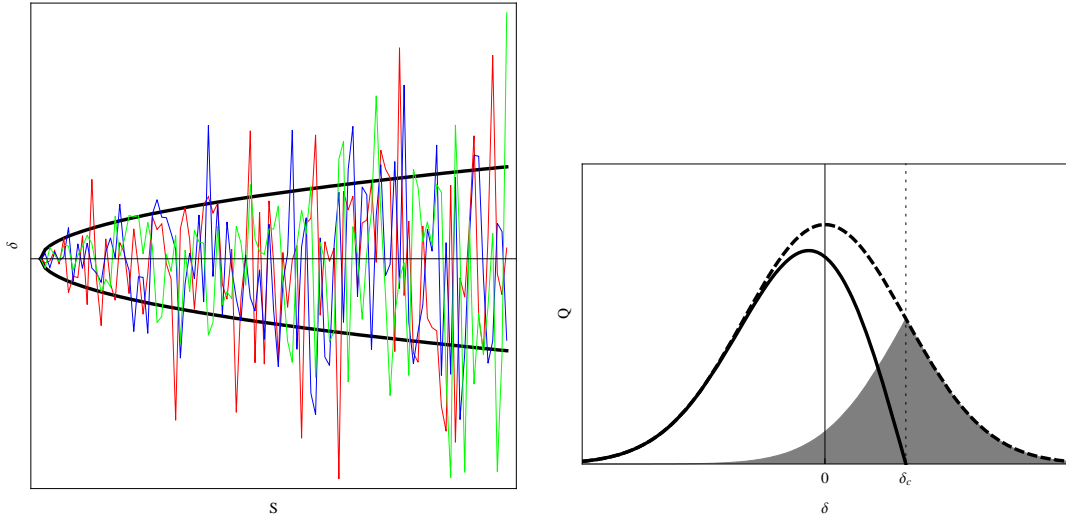
Figure 1.6 *Left*: Three examples of the random walks mentioned in the text, in color. The black curve is $\sqrt{S}$, i.e. the standard deviation in the distribution from eq. 1.40. *Right*: Solutions for the diffusion equation, for a fixed value of S and δ_c. The free random walk distribution from eq. 1.40 is plotted dashed, while the solution with a barrier, eq. 1.42 is plotted solid. The gray area below the dashed curve corresponds to the difference between the dashed and solid lines.

of mass M present at cosmological time t, the *unconditional mass function* is:

$$
n(M,t)dM \;=\; \frac{\bar{\rho}}{M}f_S(S,\delta_c(t))\left|\frac{dS}{dM}\right|dM \tag{1.47}
$$

$$
=\; \left(\frac{2}{\pi}\right)^{1/2}\frac{\bar{\rho}\delta_c(t)}{M^2\sigma(M)}\left|\frac{d\ln\sigma}{d\ln M}\right|\exp\left(-\frac{\delta_c^2(t)}{2\sigma^2(M)}\right)dM \tag{1.48}
$$

where finally we have replaced the expression in 1.46 and the variance $S = \sigma(M)$ to obtain the expression first derived by Press & Schechter. Notice that in all this treatment we have not defined the time at which the threshold δ_c and the variance $\sigma(M)$ is evaluated. The usual choice is to use the prediction of linear growth of structures to evolve δ to the present time and calculate σ at present time also, using the appropiate transfer function.

The advantage of considering these $\delta(S)$ trajectories is that they allow us to deduce the merger history of each particle. If we want to follow the evolution of the mass of the halo containing the particle, we just need to consider the upper envelope
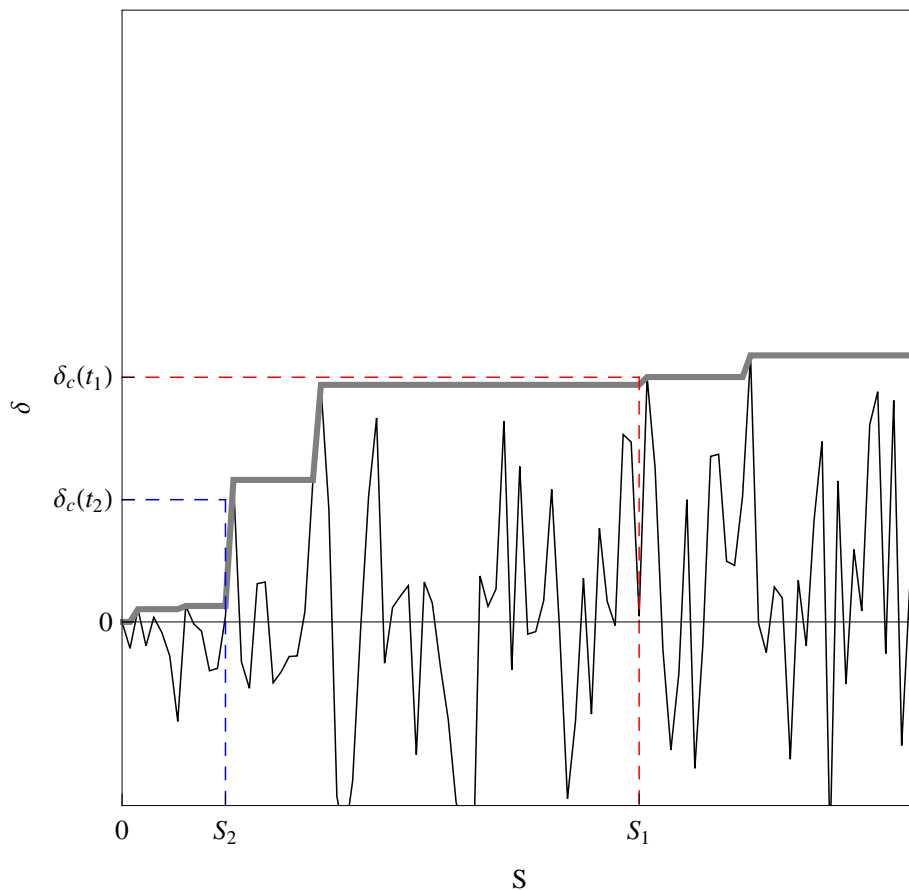
Figure 1.7 A random walk (black, thin) with its upper overdensity envelope (gray, thick). The thick gray line represents the merger history of the halo, and big steps in this path represent major mergers. Position of the selected coordinates (δ_1, S_1) and (δ_2, S_2) for the problem with two barriers are shown in dashed lines.

of the trajectory described by the particle, i.e. the curve described by $\mathrm{Max}_{\tilde{S} \leq S} \delta(\tilde{S})$. Figure 1.7 shows the merger history derived from the random walk for one particle. Mass increase corresponds to the decrease of S when we decrease the barrier δ_c as cosmic time advances. Major mergers correspond to big steps of the stairs, while the accretion of smaller halos correspond to little steps.

Notice that since the smoothed field $\delta(M, \boldsymbol{x})$ is calculated over the *initial* density field everything depends on cosmic time only through $\delta_c(t)$. For this reason, to calculate the conditional mass function, we only need to reconsider the random walk problem with two constant barriers in δ. Let's place two barriers, at $\delta_c(t_1)$ and $\delta_c(t_2)$

$(t_1 < t_2)$, and let's consider random walks that start from $(S_2, \delta_c(t_2))$. Figure 1.8 shows that paths of three of these random walks before they cross the second barrier $\delta_c(t_1)$. The distribution of S_1's where the random walks reach the second barrier is what we are looking for, since this will yield the mass distribution of halos at cosmic time t_1 that will end up in a halo of mass M_2 at a later time t_2.

We can see now that this new problem is only a trivial variable change over the solution we already had in eq. 1.46. It corresponds to a shift in the origin from $(0,0)$ to $(S_2, \delta_c(t_2))$:

$$
\begin{aligned}
f_{S_1}(S_1, \delta_1 | S_2, \delta_2) dS_1 &= f_S(S_1 - S_2, \delta_1 - \delta_2) dS_1 & (1.49) \\
&= \frac{\delta_1 - \delta_2}{\sqrt{2\pi(S_1 - S_2)^3}} \exp\left(-\frac{(\delta_1 - \delta_2)^2}{2(S_1 - S_2)}\right) dS_1 & (1.50)
\end{aligned}
$$

where we have simplified the notation by defining $\delta_X = \delta_c(t_X)$. The conditional mass function for the ancestors of a given halo can then be obtained from:

$$
n_1(M_1, z_1 | M_2, z_2) dM_1 = \frac{M_2}{M_1} f_{S_1}(S_1(M_1), \delta_1 | S_2(M_2), \delta_2) \left|\frac{\partial S_1}{\partial M_1}\right| dM_1, \qquad (1.51)
$$

This expression is naturally only valid for $S_1 > S_2$ and $\delta_1 > \delta_2$. However, we are also interested in finding the possible descendants of a given halo, so we also need the distribution $f_{S_2}(S_2, \delta_2 | S_1, \delta_1)$, which can be obtained simply from Bayes' theorem:

$$
\begin{aligned}
f_{S_2}(S_2, \delta_2 | S_1, \delta_1) dS_2 &= \frac{f_{S_1}(S_1, \delta_1 | S_2, \delta_2) f_{S_2}(S_2, \delta_2)}{f_{S_1}(S_1, \delta_1)} dS_2 \\
&= \frac{1}{\sqrt{2\pi}} \left(\frac{S_1}{S_2(S_1 - S_2)}\right)^{3/2} \frac{\delta_2(\delta_1 - \delta_2)}{\delta 1} \\
&\quad \times \exp\left(-\frac{(\delta_2 S_1 - \delta_1 S_2)^2}{2 S_1 S_2 (S_1 - S_2)}\right) dS_2 \qquad (1.52)
\end{aligned}
$$

in terms of mass, which is the variable of interest, the conditional mass function for the descendants would then be:

$$
n_2(M_2, z_2 | M_1, z_1) dM_2 = \frac{M_1}{M_2} f_{S_2}(S_2(M_2), \delta_2 | S_1(M_1), \delta_1) \left|\frac{\partial S_2}{\partial M_2}\right| dM_2, \qquad (1.53)
$$
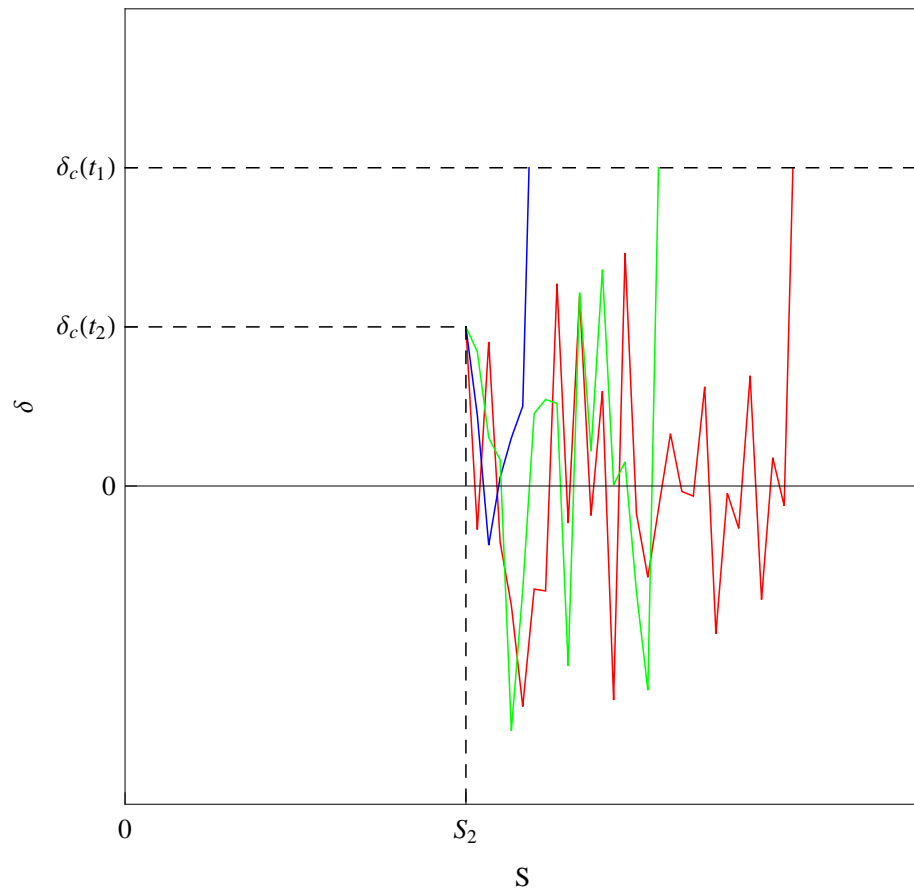
Figure 1.8 Three random walks that start from $(S_2, \delta_c(t_2))$ and reach the threshold barrier $\delta_c(t_1)$, in color.

1.4.2   The spherical collapse of a halo

We will now examine the condition for the collapse of a halo under the simple spherical infall model that we mentioned above Peebles (1967). Let's consider an isolated spherical perturbation over an homogeneous, Einstein-de Sitter universe. This models neglects the cosmological constant, but at early times in the universe that source of energy has little importance anyway. Since we are interested in inhomogeneities that will collapse as time goes by, inside the halo we have a Friedmann equation with positive curvature:

$$\left(\frac{\dot{r}}{r}\right)^2 - \frac{8\pi G\rho}{3} + \frac{\kappa}{r^2} = 0, \tag{1.54}$$

where $r(t)$ is the radius of the halo (taking the place of the scale factor in a closed universe), $\rho(t)$ is the matter density of the halo and the curvature $\kappa = 1$. The background cosmology is governed by an analog equation, but with $\kappa = 0$:

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{8\pi G\bar{\rho}}{3} = 0, \tag{1.55}$$

where now a is the scale factor of the universe, and $\bar{\rho}(t)$ is the mean density of the universe at time t. Equation 1.54 has a simple parametric solution:

$$\frac{r(\eta)}{r_m} = \frac{1 - \cos(\eta)}{2} \tag{1.56}$$

$$\frac{t(\eta)}{t_m} = \frac{\eta - \sin(\eta)}{\pi}, \tag{1.57}$$

The η parameter goes from 0 at the beginning of the expansion, to π at maximum expansion and back to full collapse at $\eta = 2\pi$. On the other hand, equation 1.55 has two solutions, $a(t) \propto t^{2/3}$ and $a(t) \propto t^{-1}$. We take only the first one, assuming there has already been enough time for the decaying mode to vanish. (see figure 1.9)

Notice that when approaching $\eta = 2\pi$ the above model for the halo is no longer realistic, since pressure support due to the baryonic matter stops the collapse and the halo virializes into a stable radius. This is not a problem, however, since we don't need to accurately describe the full evolution of the halo. We just need to define a threshold to identify the perturbations that will collapse. The linear approximation

to the evolution of the initial fluctuations will be enough, as long as we can estimate this threshold.

To do so, we can separate the cosmological expansion from the linear approximation of the halo collapse. For algebraic convenience, let's define a variable $b(t) = (t/t_m)^{2/3}$ and expand the solution for the radius of the halo in terms of that variable. A second order Taylor expansion in b yields:

$$\frac{r(b)}{r_m} \approx b(\eta = 0) + \left.\frac{dr}{db}\right|_{\eta=0} b + \frac{1}{2}\left.\frac{d^2r}{db^2}\right|_{\eta=0} b^2. \tag{1.58}$$

Using eq. 1.57 we can write b in the more convenient η variable:

$$b(\eta) = \frac{(\eta - \sin(\eta))^{2/3}}{\pi^{2/3}}, \tag{1.59}$$

and with that, the two derivatives we need:

$$\left.\frac{dr}{db}\right|_{\eta=0} = \left.\frac{dr/d\eta}{db/d\eta}\right|_{\eta=0} = \frac{(3\pi)^{2/3}}{2^{4/3}}, \tag{1.60}$$

$$\begin{aligned}\left.\frac{d^2r}{db^2}\right|_{\eta=0} &= \left.\frac{d}{d\eta}\left(\frac{dr}{db}\right)\right|_{\eta=0} \\ &= \left.\frac{1}{(db/d\eta)^2}\left(\frac{d^2r}{d\eta^2} - \frac{dr/d\eta}{db/d\eta}\frac{d^2r}{d\eta^2}\right)\right|_{\eta=0} = -\frac{(3\pi)^{4/3}}{10 \cdot 2^{2/3}}.\end{aligned} \tag{1.61}$$

Replacing 1.60 and 1.61 into 1.58, we find:

$$\frac{r(b)}{r_m} = \frac{(3\pi)^{2/3}b}{2^{4/3}} - \frac{(3\pi)^{4/3}b^2}{10 \cdot 2^{5/3}}, \tag{1.62}$$

and returning to the time variable, we finally get:

$$\frac{r_L(t)}{r_m} = \frac{1}{4}\left(\frac{6\pi t}{t_m}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6\pi t}{t_m}\right)^{2/3}\right]. \tag{1.63}$$

In this last expression we can readily identify the expansion factor of the Einstein-de Sitter universe, $a(t)/a(t_m) = (1/4)(6\pi t/t_m)^{2/3}$, which appears leading the expression, but also inside the factor in square brackets. The numerical constant in a is arbitrary
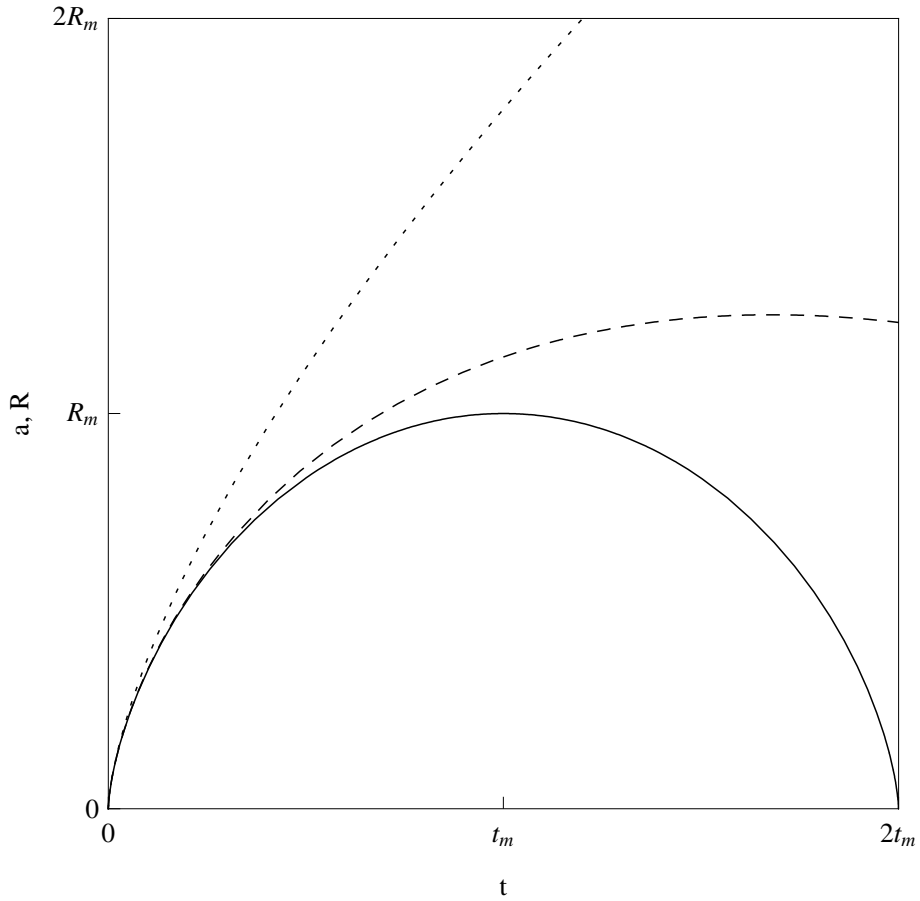
Figure 1.9 Radius of the spherical halo as a function of time (solid) plotted together with the scale factor of an Einstein-de Sitter universe (dotted) and the evolution of size of the halo predicted by linear perturbation theory (dashed).

for this cosmology, so we are free to choose it to match the size of the halo just for convenience. The second factor, the one in square brackets, will give the linear approximation to the growth of this structure. Figure 1.9 shows the cosmological expansion factor together with the aproximation in eq. 1.63 and the exact solution from eq. 1.56 for the size of the halo.

The figure 1.10 (left) shows the exact and approximate densities of the halo against the expansion factor of the universe a. These densities are given relative to the density of the halo at maximum size (when $\eta = \pi$), and are calculated from $\rho/\rho_m = (r/r_m)^{-3}$. For reference, also the density of the universe of shown, which in the case of an Einstein-de Sitter universe scales simply as $\bar{\rho}(t) \propto t^{-2}$. The density

as calculated from the exact solution naturally diverges at $\eta = 2\pi$, since this model collapses to a singularity, and the density contrast of the halo is given by:

$$\delta(t) = (r(t)/r_m)^{-3}/(a(t)/a_m)^{-3} - 1 \tag{1.64}$$

The prediction at the collapse point from linear theory is, however, finite. This allows us to determine a useful threshold contrast for our purposes. Up to linear order in $a(t)$ we obtain:

$$
\begin{aligned}
\delta_L &= \frac{\rho_L}{\bar{\rho}} - 1 = \frac{(r_L(t)/r_m)^{-3}}{(a(t)/a_m)^{-3}} - 1 \\
&= \left(1 - \frac{1}{20}\left(\frac{6\pi t}{t_m}\right)^{2/3}\right)^{-3} - 1 \approx \frac{3}{20}\left(\frac{6\pi t}{t_m}\right)^{2/3}
\end{aligned}
\tag{1.65}
$$

The right part of figure 1.10 shows both δ and the linear prediction δ_L against the cosmological expansion $a(t)$. Notice that at $\eta = 2\pi/3$ the evolution of the halo becomes completely non-linear, i.e. $\delta \sim 1$, and that point the linear prediction already separates significantly from the non-linear model. At $\eta = 2\pi$, i.e. $t = 2t_m$, linear theory predicts that the contrast is $\delta_L = 1.686$, and this is the value we will consider as our threshold in the case of spherical collapse. Actually, this value itself is telling us that we are outside the small fluctuations regime, but again that issue is not relevant for our purposes.

1.4.3   The mass-bias relation

The other key ingredient necessary for our analysis is the bias factor of the dark matter halos that host our galaxies. This quantity is defined as:

$$b = \frac{\delta_h(R)}{\delta_m(R)} \tag{1.66}$$

where δ_h is the density contrast of the dark halos, and δ_m is the density contrast of the matter field. Both quantities depend on the radius of the sphere being considered, or equivalently, on the mass M enclosed. This makes b in general a function of the
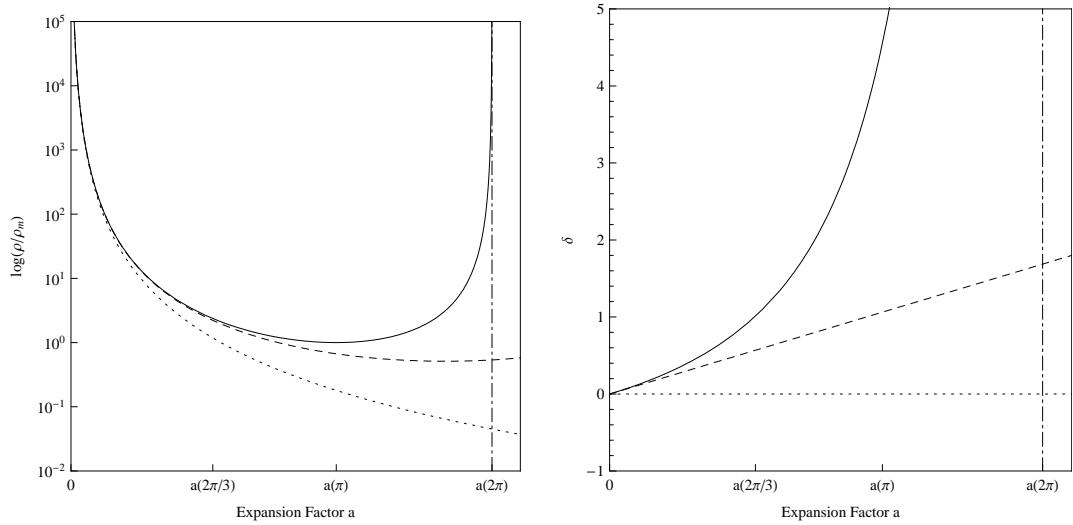
Figure 1.10 *Left*: Density of the full non-linear solution of the halo (solid), for the linear approximation (dashed), and for the universe (dotted) plotted against the cosmological expasion factor. The vertical dot-dashed line marks the moment of collapse, at $\eta = 2\pi$. *Right*: Evolution of the halo constrast (solid) and of the linear approximation to it (dashed) as calculated in eq. 1.65. At $\eta = 2\pi$ the non-linear calculation diverges, while the linear approximation reaches 1.686.

scale R, but for large scales, it becomes a constant. Since we are interested in this regime, we will treat b as a constant number. Within this approximation, we can see from the definition in equation 1.6 that the relation between the autocorrelation function of a certain halo population and the autocorrelation function of the mass field is very simple in this approximation:

$$\xi_h(r) = b^2 \xi_m(r) \qquad (1.67)$$

Mo & White (1996) related the bias to the random walk model used in the excursion set approach, by arguing that the bias relation was connected to the crossing of two barriers. Basically, we will consider adding a small perturbation to a region and obtaining the linear response of the halo number density to it. Let's consider halos with radius R_2 and overdensity δ_2 which by the target redshift z_1 had mass M_1 and overdensity δ_1 ($z_1 > z_2$). At this redshift the halo would have a radius $R(R_2, \delta_2, z_1)$ and a non-linear overdensity $\delta(\delta_2, z_1)$. If we compare the average number of those halos given by the ancestor's conditional mass function (eq. 1.51) with the total mean halo abundance at that epoch given by unconditional mass function times the

volume of these spheres (eq. 1.48), we obtain:

$$\delta_h(M_1, z_1 | M_2, z_2) = \frac{n_1(M_1, z_1 | M_2, z_2)}{n(M_1, z_1)V} - 1 \qquad (1.68)$$

where $V = 4\pi R^3 / 3$. Replacing the first crossings distribution in this expression, we get:

$$\delta_h(M_1, z_1 | M_2, z_2) = \frac{M_2}{\bar{\rho}V} \frac{f_1(S_1, \delta_1 | S_2, \delta_2)}{f(S_1, \delta_1)} - 1 \qquad (1.69)$$

and since $M_2 / \bar{\rho}V = 1 + \delta$, we can rewrite this as:

$$
\begin{aligned}
\delta_h(M_1, z_1 | M_2, z_2) &= (1 + \delta)\frac{f_1(S_1, \delta_1 | S_2, \delta_2)}{f(S_1, \delta_1)} - 1 \\
&= (1 + \delta)\frac{f(S_1 - S_2, \delta_1 - \delta_2)}{f(S_1, \delta_1)} - 1.
\end{aligned} \qquad (1.70)
$$

The approximation mentioned above corresponds to the limit $S_2 \approx 0$ and $\delta_1 \gg \delta_2$, we can make a Taylor expansion of this expression on δ_2:

$$\delta_h(M_1, z_1 | M_2, z_2) \approx (1 + \delta)\left(1 - \frac{1}{f(S_1, \delta_1)}\frac{\partial f(S_1, \delta_1)}{\partial \delta}\delta_2\right) - 1 \qquad (1.71)$$

Since $\delta_2 \approx \delta$, we can further approximate and find an expression for the average bias for halos of mass M_1 at z_1:

$$
\begin{aligned}
\delta_h(M_1, z_1 | M_2, z_2) &\approx \left(1 - \frac{1}{f(S_1, \delta_1)}\frac{\partial f(S_1, \delta_1)}{\partial \delta}\right)\delta \\
\implies b(M_1, z_1) &= \frac{\delta_h}{\delta} = \left(1 - \frac{1}{f(S_1, \delta_1)}\frac{\partial f(S_1, \delta_1)}{\partial \delta}\right)
\end{aligned} \qquad (1.72)
$$

Using Press & Schechter's mass function gives the following mass-bias relation:

$$b(M_1, z_1) = 1 + \frac{\delta_1(z_1)^2 / S_1(M_1) - 1}{\delta_1(z_1)}. \qquad (1.73)$$

Notice that for this equation to be valid δ does not need to be small, and we can use it even at high redshifts.

It is worth mentioning that all these formulas are also encountered in terms of

the rescaled variable $\nu(M, z) = \delta_c(z)/\sqrt{S(M)}$ in the modern literature:

$$f(\nu)d\nu = \sqrt{\frac{2}{\pi}} \exp\left(-\frac{\nu^2}{2}\right) d\nu, \qquad (1.74)$$

$$n(M, z)dM = \frac{\bar{\rho}}{M} f(\nu) \left|\frac{d\nu}{dM}\right| dM, \qquad (1.75)$$

$$b(M, z) = 1 + \frac{\nu(M, z)^2 - 1}{\delta_c(z)}, \qquad (1.76)$$

which naturally results appealing when analyzing the first crossings distribution, since that change turns it into a one variable function.

1.4.4 The ellipsoidal collapse improvement

Now that we have examined the mass functions and bias of halos in the spherical collapse approximation, where a complete analytical solution is possible, we will consider departures from it that will ultimately yield the expressions that we will use in the present work. Sheth & Tormen (1999) showed that a notorious improvement in the reproduction of the mass function in numerical simulations could be achieved by considering a "moving" barrier instead of a constant one in the random walk problem, so that now the threshold the trajectories have to cross is no longer horizontal. This is motivated by the dynamics of the collapse of an ellipsoidal rather than simply a spherical perturbation. The first crossings distribution is slightly different than the one we obtained for the spherical collapse case:

$$f(\nu)d\nu = A\sqrt{\frac{2a}{\pi}}(1 + (a\nu^2)^{-q})e^{-a\nu^2/2} \qquad (1.77)$$

where constants are $a = 0.707$, $q = 0.3$ and $A = 0.322$. The first two constants are fitted to cosmological numerical simulations (Sheth & Tormen, 1999 used the GIF simulations, Kauffmann et al., 1999), while the last constant is only a normalization obtained by requiring the distribution to integrate to unity. The bias associated with this new approximation can be calculated replacing this last expression in eq. 1.72

to yield:

$$b(M, z) = 1 + \frac{1}{\delta_c(z)} \left(a\nu^2 - 1 + \frac{2p}{1 + (a\nu^2)^p} \right) \tag{1.78}$$

Notice that in the derivation of the conditional mass function we used the fact that in the spherical collapse approximation the barrier is one of constant height. Since this is no longer the case, applying the same method for obtaining the conditional mass function is only an approximation. The same authors later provided a further refinement of their formulas fitting a more general barrier function with three parameters (Sheth et al., 2001, Sheth & Tormen, 2002):

$$B(S, z) = \sqrt{a}\delta_{sc}(z)[1 + b(a\delta_{sc}(z)^2/S)^{-c}], \tag{1.79}$$

where again $a = 0.707$, but we also have the constants b $= 0.5$ and c $= 0.6$. In terms of this barrier, the authors present the following approximation in terms of an expansion series in B(S):

$$f(S, z) = |T(S, z)| \exp\left(-\frac{B(S, z)^2}{2S}\right) \frac{dS}{\sqrt{2\pi}S^{3/2}} \tag{1.80}$$

$$T(S, z) = \sum_{n=0}^{5} \frac{(-S)^n}{n!} \frac{\partial^n B(S, z)}{\partial S^n}. \tag{1.81}$$

Beware of the slightly different definitions of ν in these papers, and in the different variables used for the same constants. Thanks to this parametrization, the the method used to obtain the conditional mass function is analog as the one we stated before, but is bit more complicated to compute. Now the replacements necessary are: $S \to S_1 - S_2$ and $B(S) \to B(S_1) - B(S_2)$. The approximation this yields for the conditional mass function is the following:

$$f(S_1, z_1|S_2, z_2) = |T(S_1, z_1|S_2, z_2)| \exp\left(-\frac{[B(S_1, z_1) - B(S_2, z_2)]^2}{2(S_1 - S_2)}\right)$$
$$\times \frac{dS_1}{\sqrt{2\pi}(S_1 - S_2)^{3/2}} \tag{1.82}$$

$$T(S_1, z_1|S_2, z_2) = \sum_{n=0}^{5} \frac{(S_2 - S_1)^n}{n!} \frac{\partial^n [B(S_1, z_1) - B(S_2, z_2)]}{\partial S_1^n}. \tag{1.83}$$
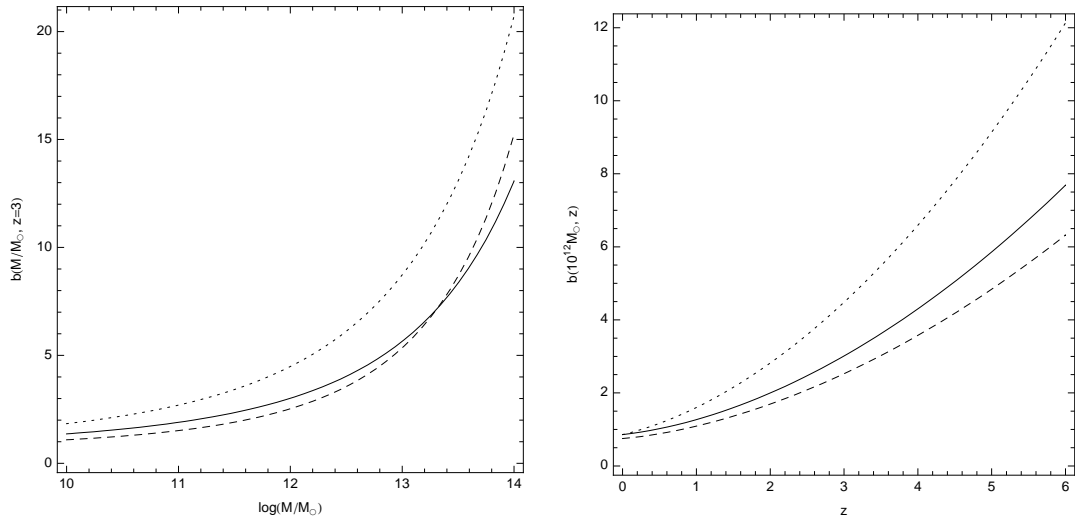
Figure 1.11 *Left*: Bias-mass relation at redshift z=3, given for three cosmological models: ΛCDM (solid), the fiducial model adopted in this work; SCDM (dashed), the old standard cold dark matter model, defined by the parameters $\Omega_m = 1.0$, $\Omega_\Lambda = 0.0$ and $h = 0.5$; OCDM (dotted), an open model with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.0$ and $h = 0.7$. *Right*: Bias for a 10^{12} M$_\odot$ halo as a function of redshift, for the same three cosmological models.

Naturally, this gets even more lengthy when we use Bayes' theorem to compute the descendants conditional mass function, but it not extremely necessary unless one wants to use it for halos of very low mass or at very high redshifts, where the barrier gets crossed at high S values (see figure 2 and 4 of Sheth & Tormen, 2002). The authors also give the more accurate version of the bias-mass relation in Sheth et al. (2001), which is given by:

$$b(M, z) = 1 + \frac{1}{\delta_c(z)} \left(a\nu^2 + b(a\nu^2)^{1-c} - \frac{(a\nu^2)^c / \sqrt{a}}{(a\nu^2)^c + b(1-c)(1-c/2)} \right) \qquad (1.84)$$

but again, when evaluated, the numerical difference with the result in eq. 1.78 is rather small. Figure 1.11 depicts this relation for fixed redshift and fixed mass, and for three different cosmological models.

1.5 Connection to observations and cosmological quantities

The first quantity derived from our measurements that we need to compute is the effective bias factor for our galaxies. Here we adopt the following expression to estimate it:

$$b_{\rm g} = \frac{\sigma_{8,\rm g}}{\sigma_{8,\rm dm}} \tag{1.85}$$

where $\sigma_{8,\rm X}^2$ is the variance in spheres of radius $8h_{100}^{-1}$ Mpc. The variance of the galaxy field can be computed directly from the autocorrelation function (Peebles, 1980):

$$\sigma_{8,\rm g}^2 = \frac{72(r_0/8h_{100}^{-1}Mpc)^\gamma}{(3-\gamma)(4-\gamma)(6-\gamma)2^\gamma} \tag{1.86}$$

while the variance in the matter field is simply replaced with the σ_8 parameter measured by WMAP, appropriately scaled with the cosmological growth factor for a ΛCDM cosmology (Carroll et al., 1992):

$$\sigma_{8,dm}(z) = D(z)\sigma_8 \tag{1.87}$$

$$D(z) = \frac{g(z)}{g(0)(1+z)} \tag{1.88}$$

$$g(z) \approx \frac{5}{2}\Omega_m(z)\left[\Omega_m^{4/7}(z) - \Omega_\Lambda(z) \right.$$
$$\left. +(1+\Omega_m(z)/2)(1+\Omega_\Lambda(z)/70)\right]^{-1} \tag{1.89}$$

where these cosmological quantities are related to their present-day value by:

$$\Omega_m(z) = \frac{\Omega_{m,0}(1+z)^3}{E^2(z)}, \qquad \Omega_\Lambda(z) = \frac{\Omega_{\Lambda,0}}{E^2(z)} \tag{1.90}$$

$$E(z) = \sqrt{\Omega_{\Lambda,0} + (1-\Omega_{\Lambda,0}-\Omega_{m,0})(1+z)^2 + \Omega_{m,0}(1+z)^3}. \tag{1.91}$$

Another quantity of importance is the virial radius of a dark matter halo. Following common practice, we calculate it as the radius enclosing a density 200 times the mean density of the universe:

$$r_{200} = \left(\frac{GM}{100\Omega_m(z)H^2(z)}\right)^{1/3} \tag{1.92}$$

For the calculation of the variance $\sigma(M)^2$, we need to compute the integral (Dodelson, 2003):

$$\sigma(R)^2 = C \int_0^\infty k^2 P(k) \tilde{W}^2(kR) dk \tag{1.93}$$

where the radius $R(M) = (3M/4\pi\bar{\rho}_0)^{1/3}$ (with $\bar{\rho}_0$ the current mean density of the universe), the power spectrum is assumed to have $n = 1$ and be matter dominated, i.e. $P(k) \propto k T^2(k)$, the Fourier transform of the top-hat window function $\tilde{W}(x) = 3(\sin(x) - x\cos(x))/x^3$ and the transfer function is taken from the fit by Bardeen et al. (1986):

$$
\begin{aligned}
T(q = k/[\Omega_{m,0}h^2 Mpc^{-1}]) &= \frac{\ln(1 + 2.34q)}{2.34q}(1 + 3.89q \\
&\quad + (16.1q)^2 + (5.46q)^3 + (6.71q)^4)^{-1/4}. \tag{1.94}
\end{aligned}
$$

Finally, the constant C is obtained by demanding that $\sigma(R = 8h^{-1}Mpc) = \sigma_8$.

To connect the measured bias factor with theory, we need to average the theoretical value $b(M, z)$ over the whole population being probed. To do so, we assume the simplest halo occupation distribution (HOD) model possible: a one to one galaxy-halo correspondance above a threshold mass, and no galaxies below the threshold. This makes the relation between the observed bias factor and the quantity calculated in (1.78) fairly simple:

$$b_{\text{eff}}(M_{\text{min}}, z) = \frac{\int_{M_{min}}^\infty b(M, z) n(M, z) dM}{\int_{M_{min}}^\infty n(M, z) dM}. \tag{1.95}$$

Constraining more complicated HOD models requires a much higher signal to noise that the one available in our galaxy samples, and a more refined treatment of the correlation function which is outside the scope of the present work. Given this estimation for M_{min}, we will calculate the median mass of such a distribution and the number density for such halos, using the appropiate expressions given in the previous sections. To estimate the evolution of the halo population we will follow the median of the descendants conditional distribution (median of M_2) at each redshift z_2, then we calculate the corresponding threshold masses M_{min} that correspond to distributions with these median values, and finally we calculate b_{eff} for each threshold

mass. This way we build the bias tracks we will see in the next chapters and that allow for a direct comparison with published bias values in the literature.

Chapter 2

Observations: The MUSYC survey

2.1 Optical photometry and catalogs

The MUSYC survey was optimized to study galaxies at $z \sim 3$, with imaging depths down to the spectroscopic limit, U, B, V, $R \sim 26$ (Gawiser et al., 2006b). The survey comprises four fields, which were chosen because of their low reddening, low H I column density and low 100μm dust emission, in order to be best suited for satellite coverage with *HST*, *Spitzer*, *Chandra* and *XMM-Newton*. As their objective is mainly extragalactic studies, the four fields have also been chosen to have high galactic latitudes ($|b| > 30$) and few bright foreground sources. Figure 2.1 shows UBR color images of the four fields. The field coordinates, sizes and other general properties are shown in table 2.1, while the limiting magnitudes achieved on each optical filter[1] are shown in table 2.2. Filter transmission curves are shown in figure 2.2. The main MUSYC catalog is based on the sources detected on the combined BVR image, and aperture photometry is performed on each filter at those positions (see Gawiser et al., 2006a). Additionally, sources detected only in the narrow band image were added to the main BVR catalog, to allow the selection of emission line objects using

---

[1] A list of the characteristics for each of the filters used, for both CTIO and KPNO, can be found at http://www.noao.edu/kpno/mosaic/filters/index.html

Table 2.1. Summary of MUSYC fields

Field Name	Area (arcmin^2)	Equatorial R.A.	Coordinates Dec.	Galactic Coordinates	$E(B-V)$ (mag)	$N(\text{H \textsc{i}})$ (10^{20}cm^{-2})	# Obj
ECDF-S	31.́6×31.́6	03 32 29.0	-27 48 47	(224, -54)	0.01	0.9	84410
EHDF-S	34.́3×32.́9	22 32 35.6	-60 47 12	(328, -49)	0.03	1.6	62968
SDSS 1030	36.́0×32.́0	10 30 27.1	+05 24 55	(239, 50)	0.02	2.3	69619
CW 1255	34.́3×34.́3	12 55 40.0	+01 07 00	(306, 64)	0.02	1.6	60344

Note. — Right ascension is given in hours, minutes, and seconds, and declination in degrees, arcminutes and arcseconds, at J2000.0 epoch. Reddening values and H \textsc{i} column densities from Burstein & Heiles (1978). Number of objects are the number of detections on the BVR image.

Table 2.2. Optical depths

Field	U	B	V	R	I	z	NB5000	BVR
ECDF-S	26.0	26.9	26.4	26.4	24.6	23.6	25.5	27.1
EHDF-S	26.0	26.1	26.0	25.8	24.7	23.6	24.1	26.3
SDSS 1030	25.7	26.0	26.2	26.0	25.4	23.7	24.8	26.5
CW 1255	tbd	tbd	tbd	tbd	tbd	tbd	tbd	tbd

Note. — Bandpasses depths are for 5-σ point source detection, in AB magnitudes. The depths in the CW 1255 field have not been determined at the time of this publication yet.

narrowband excesses. Images were taken with the MOSAIC-I camera at Kitt Peak National Observatory and with the MOSAIC-II camera at Cerro Tololo International Observatory, between 2002 and 2005. These instruments yield a scale on the images of 0.́′267/pixel. Details on the survey design and imaging data reduction process, as well as details on the construction of the photometric catalogs can be found in Gawiser et al. (2006a).

2.2   X-ray ECDF-S dataset

The multiple Chandra observations of this field have produced four X-ray catalogs: Giacconi et al. (2002), Alexander et al. (2003), Lehmer et al. (2005), and Virani
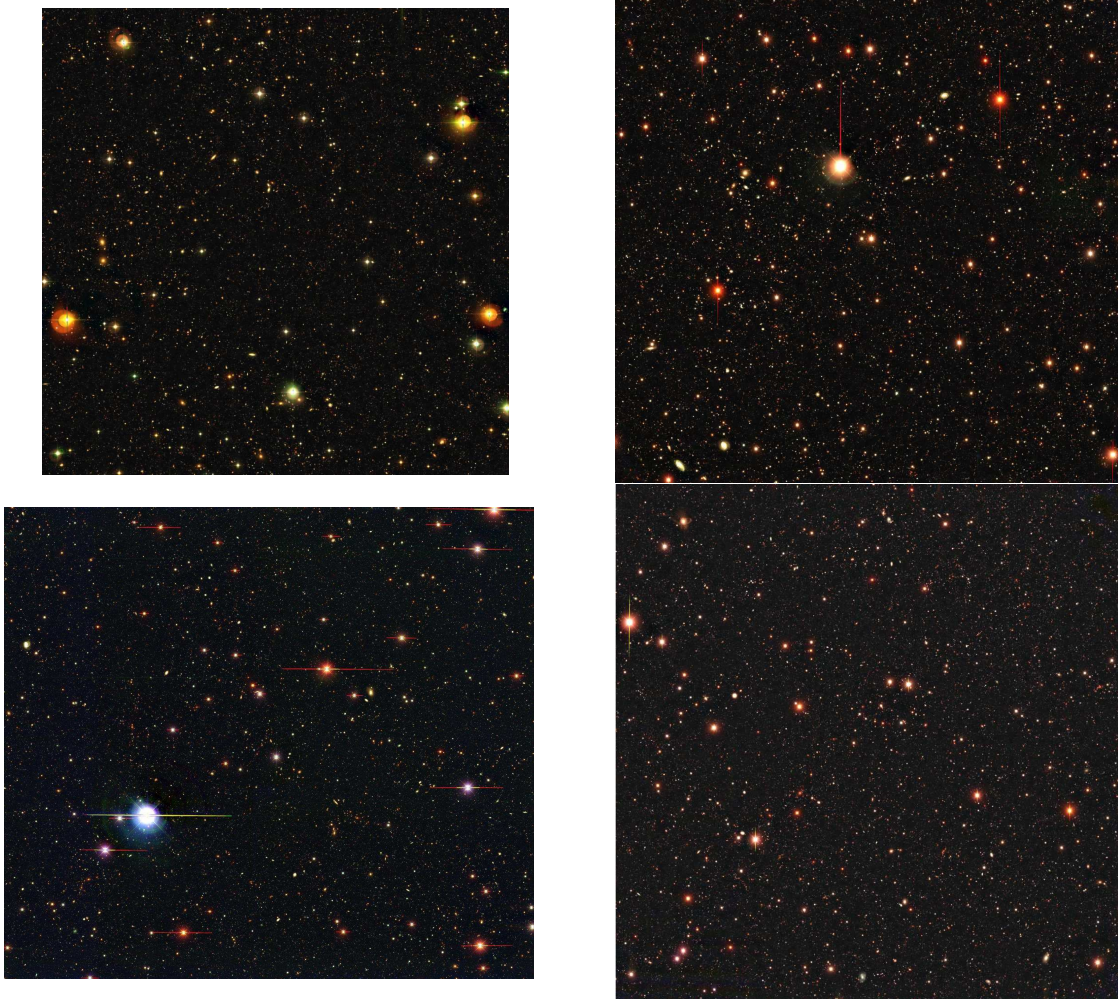
Figure 2.1 Optical UBR color images of the E-CDFS (top-left), E-HDFS (top-right), SDSS 1030 (bottom-left) and CW 1255 fields (bottom-right). Images are scaled to the physical sizes of the fields.
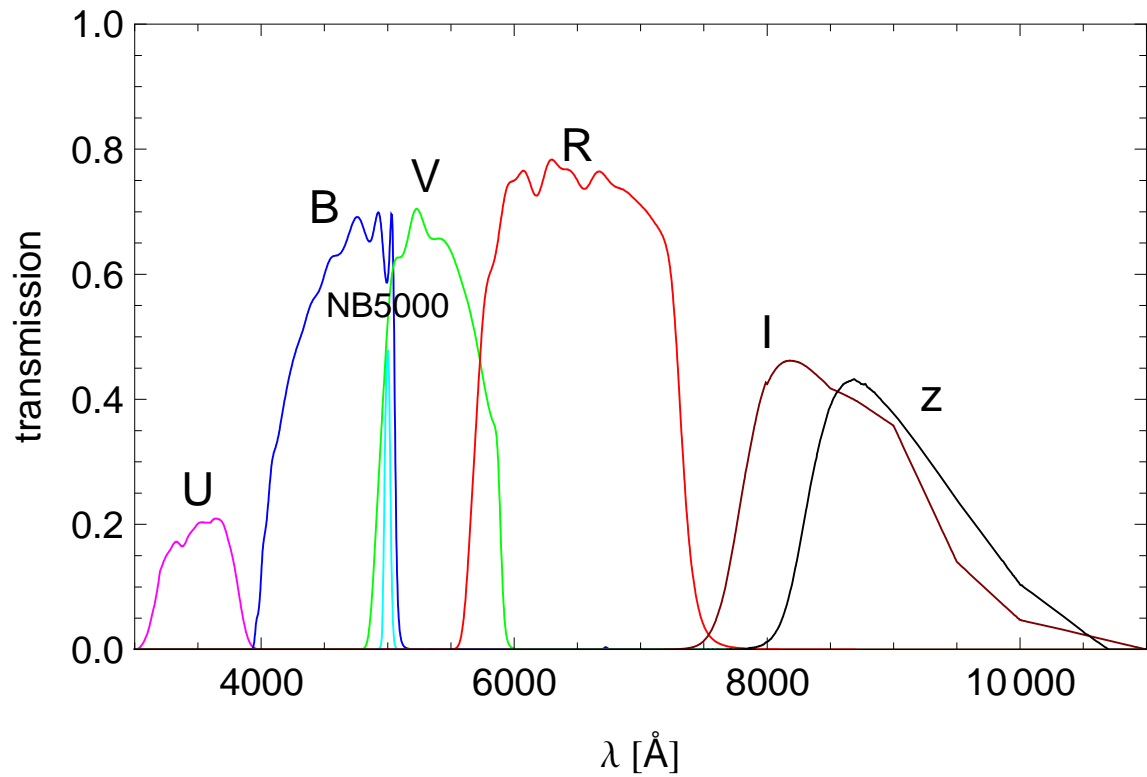
Figure 2.2 Transmission curves for the set of optical MUSYC filters used.

et al. (2006). The first two comprise 1 Ms of exposures inside the central region (CDF-S *proper*), covering an area of $\approx$0.1 deg^2 (PI R. Giacconi). The Alexander et al. catalog has reported flux limits of 5.2×10^{-17} erg cm^{-2} s^{-1} and 2.8×10^{-16} erg cm^{-2} s^{-1} in the soft (0.5-2 keV) and hard (2-8 keV) bands, respectively. The last two catalogs come from the four $\approx$250 ks pointings that cover an area of $\approx$0.3 deg^2 around the former field (PI N.Brandt). Figure 2.3 shows these pointings in reference to the optical observations. Limiting X-ray fluxes in the extended region are 1.1×10^{-16} erg cm^{-2} s^{-1} in the soft and 6.7×10^{-16} erg cm^{-2} s^{-1} in the hard bands, correspondingly. Notice that the individual observations in the central CDF-S do not have equal roll angles, so the shape of the final stacked X-ray image is not square.

2.3 Spectroscopic observations

The MUSYC spectroscopic follow-up program included in this work was carried out with the Inamori-Magellan Areal Camera & Spectrograph (IMACS)[2] at the Baade Magellan telescope. The instrument is installed in the Gregorian Nasmyth port 2 of the telescope. It has two cameras, called f/2 and f/4 ("short" and "long"), which differ little in wavelength range coverage, but rather in the size of the field of view and in scale. At the time MUSYC data was taken, both cameras shared the same detector, a 8K×8K CCD mosaic that was exchanged between the two. As of march 2008, a second CCD was installed, and both remain fixed in their respective cameras. The f/2 camera has a field of view of 27.2×27.2 arcmin^2 and a scale of 0.2″/pixel, while the f/4 camera has field of 15.46×15.46 arcmin^2 and a scale of 0.111″/pixel. The MUSYC spectroscopic data were obtained with the 300 line/mm grism in the f/2 camera, a configuration that yields a dispersion of 1.341 Å/pixel and a wavelength coverage going from $\sim$4000 Å to 9000 Å. In the observing runs between may 2004 and february 2005, we used slitlets of 1.2″, which gave a resolution of R=625 ($\Delta v = 480$ km/s or $\Delta\lambda = 8$ Å at 5000 Å). In the run from october 2003

---

[2]The user manual of the instrument, together with a observing cookbook can be found inside the webpage from Las Campanas Observatory: http://www.lco.cl/lco/telescopes-information/magellan/magellan/instruments/imacs/user-manual-1/user-manual
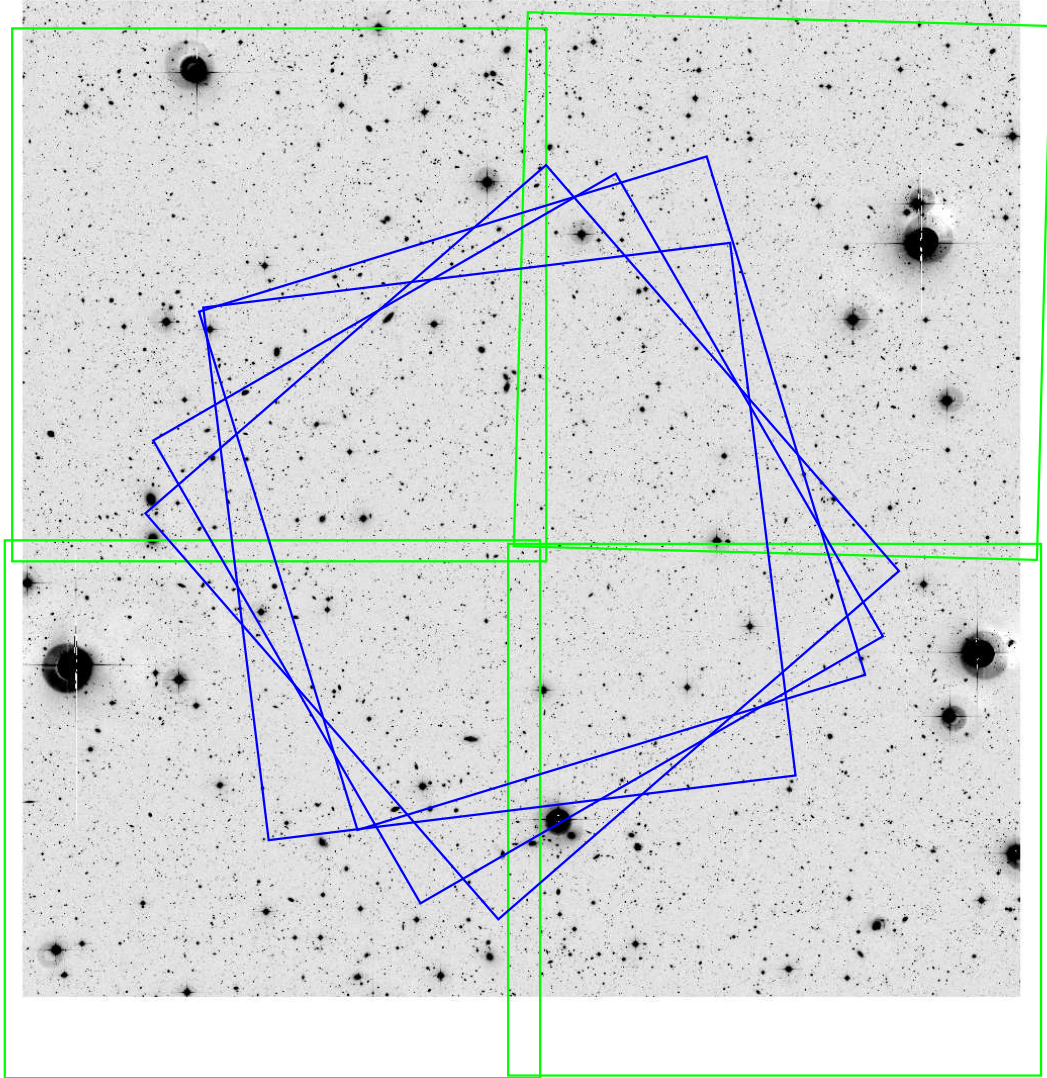
Figure 2.3 Chandra pointings for the E-CDFS (green) and CDF-S *proper* (blue) overlaid on the MUSYC BVR optical image.

and the runs between september 2005 and february 2007, we used slitlets of 1.0″, which gave a resolution of R=746 ($\Delta v = 400$ km/s or $\Delta \lambda = 6.7$ Å at 5000 Å). These are lower limits to the resolutions achieved, because even though seeing was >1″for most observing runs, some nights benefited from sub-arcsec seeing: notably on november 2005 seeing was consistently near 0.6″, yielding a resolution of 4Å. The broad wavelength coverage provided by IMACS allows the detection of Lyα and C IV 1550, the two most prominent ultraviolet lines at redshifts $2.3 < z < 4.0$. In starforming galaxies, the Lyα, which depends greatly on the dust content and geometry of the interstellar medium, may show up in emission or absorption. On the other hand, the C IV 1550, which is predominantly formed in massive stars, is always seen in absorption, in a P-Cygni type of profile. In AGN on the other hand, both lines can be seen in emission in the wavelength range available in these spectra, and come from the active nucleus. For more details on the spectroscopic observations, see Lira et al. (in prep). The selection of LBG, AGN and LAE candidates for spectroscopic follow-up is explained in §3.2, §4.2 and §5.2, respectively. In this chapter we will show only some examples of these spectra, more details will be given in each of the chapters. So far, the MUSYC spectroscopic follow-up program has yielded over 280 successful identifications of $z > 2$ galaxies.

2.3.1   Design of IMACS multislit masks

The design of multislit masks for IMACS (as well as for the LDSS-3 spectrograph installed in the twin Magellan telescope) must be done using the `maskgen` program.[3] The first input from the user to the program has to do with the creation of a suitable object catalog. This is a ASCII text file which contains first, a list of possible alignment stars, and then a list of objects where slits could be assigned. An example of such a file is given in printout 1. Of course, a real catalog file will have many, many entries, thousands even. The first character of each line tells the software what to do with the information that follows. The alignment stars are preceded

---

[3]Installation instructions and a manual for this program can be found at http://users.ociw.edu/clardy/imacs/maskmaking. Notice that the package itself contains an update script that should always be used to make sure the mask will be designed according to the latest updates of the instrument.

by an asterisk, and the objects of interest by an "at" sign (@). The first column id an identifier, useful for the user to track down which objects are finally chosen to be in the mask. In MUSYC observations, this number was chosen to be the ID from the optical catalog tied with the R magnitude of the object, just to help the observers. Second and third column are the coordinates of the objects. It is very important that the *relative* astrometry between the alignment stars and the objects targeted for spectroscopy is precise to a sub-arcsecond level. The absolute astrometric accuracy of the catalog is much less important, since the ability to put objects in the slits at the telescope depends on the relative position of the objects with respect to the alignment stars. The typical recommendation is to include stars with R magnitude between 17 and 19 in the catalog, from which 10-15 should be chosen for the alignment of an IMACS mask. The fourth column in the catalog is a priority of the object, the smaller the number the higher the priority. This priority makes it possible to have small set of objects together with sets of more numerous objects in the same mask. The program will try to lay down slits for the less numerous objects first, and then fill the remaining space with "filling objects". In the case of the galaxies that concern this investigation, AGN were always given the highest priority, then LAEs, then bright LBG and finally the dimmest LBGs. The absolute priority of each of these galaxies varied from mask to mask, since other MUSYC science projects added objects into the same multislit masks, with sample sizes and science priorities that changed in the course of different observing runs. Since the mask generation program takes objects randomly within each priority level of the input list of objects, and the mask aimpoints were selected to uniformly cover each of the fields, the spectroscopic database resulted to have a fairly uniform coverage in R.A. and Dec.

When this file has been generated, the user must run the `intgui` program that helps creating the *observation file*. In this stage the instrument setup is entered, the coordinates for the aimpoint of the telescope, the path to the object catalog, the slit and alignment box sizes, the number of alignment stars and the guide stars.[4] Another important parameter is the slit position angle. In single slit observations

---

[4]Note that to be able to use the *Skywin* feature for guide star selection, the user must install additional files not included in the standard package and set up their path in the GMAPCATS enviroment variable.

this is usually aligned with the parallactic angle of the field at the time of observation, but since the mask must be kept aligned to the alignment stars, this value is set to a constant for this kind of observations. For the MUSYC multislit masks we have chosen the closest multiple of 90° from the average parallactic angle the field had at the observing night. After the observation file is generated, the user must run the `maskgen` program, which generates the `.SMF` file. This is the file to be sent to the observatory. Another very useful utility is the `obscat` program, which given a set of `.SMF` files generates the *observing catalog*. This last file must be given to the telescope operator during the night and contains the name, coordinates and rotation angle for the mask. Although it can be made manually, this avoids mistakes, specially since the conversion between the slit position angle in the sky and the instrumental rotator angle is non-trivial.

```
# This is a comment line
#Obj ID         RA(hours)    Dec.(deg)    Priority
*2750_18.9      3.52406      -28.0488
*2989_17.9      3.5387827    -28.04911
*4273_18.0      3.5595193    -28.03761
*4406_17.2      3.531364     -28.04206
@115302_24.3    3.5302417    -27.674381   1
@115553_26.2    3.5286222    -27.66769    1
@965_23.5       3.5460133    -28.05662    2
@2412_24.7      3.53166      -28.04656    2
@5976_24.9      3.5341173    -28.02365    2
@6736_24.7      3.5546807    -28.01811    2
@20464_25.8     3.528084     -27.93304    3
@22861_26.3     3.5469933    -27.91792    3
@26866_25.7     3.5416093    -27.89164    3
@32202_25.7     3.5353013    -27.858      3
@32437_25.7     3.5441013    -27.85596    3
@32967_26.1     3.53251      -27.85232    3
@49117_26.2     3.5405367    -27.75398    3
@54242_25.8     3.5251793    -27.72367    3
```

Printout 1: Example of a object catalog file.

2.3.2 IMACS Data reduction

The reduction of the IMACS spectra was done using the Carnegie Observatories System for MultiObject Spectroscopy (COSMOS) data reduction package[5]. This software creates the master bias and flatfield frames, applies them to the science frames, calibrates the spectra in wavelength, subtracts the sky and performs the extraction of the individual spectra from the science images. We followed the standard cookbook for IMACS data reduction available in the software's webpage.

Most of the user's input within this software was needed at the stage of the creation of the "spectral map" file, which was the first step in the reduction. This file contains the solution for the mapping between (x,y) coordinates in the images and the spatial and wavelength direction in each of the spectra. It was created in two stages: the first one was the correction of the mask alignment from the one expected by default for the instrument to the actual alignment of the observed mask, which was done with the `align-mask` program. This alignment correction is described in terms of an average linear shift, rotation and an expansion/contraction of the entire frame, taking a comparison lamp as a reference. One of the inputs of the alignment program is a small list of strong and well isolated lines, which guide the program in finding the actual positions of the slits in the mask. Table 2.3 shows the initial set of lines we used. For reducing most masks, we typically started with the parameters `SEARCHBOX` $= 10$ and `MAGFACTOR` $= 20$. It is also possible to apply a shift and a rotation manually, by means of the `adjust-offset` task, something we needed to do in a few exceptional cases.

The second stage was the creation and refinement of the spectral map on a slit-by-slit basis. From the alignment file, a rough spectral map solution was first created using the `map-spectra`, and then refined with the `adjust-map` programs. For the first iteration in the refinement it is important to also use a set of well isolated lines in order to prevent the task from misidentifying lines and thus generate wrong solutions or not even converge. We used the same set of initial lines we used with

---

[5]This package and its documentation, including the cookbook for data reduction, can be found at the software repository from Carnegie Observatories: http://www.ociw.edu/Code/cosmos

`align-mask`. When the task has already found a reasonably close spectral map, it is possible to continue iterating with a more dense set of lines, and in general we used the list provided with the COSMOS program itself at this step. In that case, it may also be necessary to decrease the size of the search box in the `adjust-map` parameters, to prevent the program from confusing lines when a more dense list of lines is used. For each one of these iterations, we used the `spectral-map` program from COSMOS and IRAF's `tvmark` task to check the goodness of the spectral map solution, as is advised in the cookbook.

Although the default parameters for all the programs worked usually fairly well, for some masks the number of iterations and the degree of the polynomial needed to be changed, since the fitting process depends on the design of the multislit mask. Depending of the layout of the slits, sometimes it was more difficult to achieve a uniform convergence in the wavelength solution for all the slits at the same time and starting from the same alignment file. This occurs in general because there are mild distortions of the image that increase towards the borders of the image, so the slits near the edge are typically more curved. Whenever some of the slits can only be fit starting from a different alignment file, or using a different set of comparison lines, a simple trick that helps in building the spectral map file for the mask is to use the fact that the slits are fitted independently. Printout 2 shows the typical format of the `.map`. Every slit will fall on one or more chips, and several coefficients describe the fit perfomed on each chip. It is not mandatory that all the slits appear in a given `.map` file and thus, if the `adjust-map` task is not able to converge on the solution of a given slit, that slit can be removed from the `.map` file. That way it can be adjusted separately from a different initial alignment file, and/or using a different set of wavelength calibration lines. To obtain a final and complete `.map` file with the final solution for all the slits, it is straightforward to copy/paste the slit entries among `.map` files. Although rarely, this trick helped us with some mask reductions. Additionally, since the shifts due to flexure of the instrument between observations during the night are small (and even on different nights!) instead of starting the process to obtain a `.map` file from scratch with the `align-mask` program, we usually started from a previous `.map` file and went straight into the iterations with the `adjust-map` task, using the new comparison lamps instead. Modern versions of COSMOS have become more sophisticated and intelligent and the new versions of

the programs run fairly well without complicated tricks or modifying very much the default parameters most of the time (beyond the basic file path and instrument configuration, which are naturally mandatory).

```
Xdispersion =  N
Fit orders = N N N N N
Scale ~       N.NNNNN
Lambda  =  NNNN.N  NNNN.N
Dewar = XXXXXXXX

SLIT 1 YYYYYYYY N NNN.NNN   NNN.NNN
LENGTH = NNN.NNNN POS = NNN.NNN   NNN.NNN
CHIP 1 NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN
...
CHIP 2 NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN
...
SLIT 2 YYYYYYYY N NNN.NNN   NNN.NNN
LENGTH = NNN.NNNN POS = NNN.NNN   NNN.NNN
CHIP 1 NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN
...
CHIP 2 NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN NNN.NNN
...
```

Printout 2: Format of the .map files. File continues until every slit in the mask is listed, listing in each entry every chips the slit may lay on.

After the spectral map file is obtained, the rest of the pipeline is fairly straightforward to execute and requires very little human intervention. The rest of the cookbook can be turned easily into a script after initializing the parameters of the `Sflats`, `biasflat`, `subsky`, `extract-2dspec` and `sumspec`. The creation of the master bias frame was accomplished using the IRAF routine `imcombine`, and this last task was also used in replacement of `sumspec` program in some older versions of COSMOS, to avoid a bug that prevented the correct removal of cosmic rays. Printout 3 show the parameters used in these COSMOS programs. The extraction of the 1D spectrum from the 2D image was perfomed using the IRAF task `apall`. As a default we used the parameters listed in printout 4, but we refined the sizes of the apertures

and background regions, as well as the orders of the background subtraction and of the tracing function, in an spectrum by spectrum basis.

Although for identifying most of the spectra a flux calibration was not necessary, that step did simplify many times the identification and differentiation of objects without emission lines, such as LBGs with Lyα in absorption or late type stars, a typical interloper of the LBG selection. Flux calibration of the spectra was performed using the IRAF task `calibrate` and an average sensitivity function for the IMACS spectrograph. This provides only a rough flux calibration, but for this kind of observations, there are significant intrinsic uncertainties: most of the objects in each of the multislit masks do not show up in the individual science exposures, and several hours of integration are required to even detect them. This means that each final image has a been observed under a range of airmasses and seeing conditions. Furthermore, for early IMACS runs, the instrument lacked an atmospheric dispersion corrector (ADC), and since in multislit observations the slits orientations cannot follow the movement of the paralactic angle, there might have been some non-monocromatic flux loss in those spectra. The influence of this issue in the present investigation is, however, very small, and we do not attempt a further refinement of this reduction step.

2.3.3 Spectroscopic identifications

The main feature present in Lyman Break Galaxies (LBG) and Lyman Alpha Emitters (LAE) spectra that allowed for an identification and redshift determination was the presence of the Lyα line in emission and a decrement in the continuum level going from redward to bluewards of that line. AGN typically show, on the other hand, Lyα together with C IV 1550 in emission. Figures 2.4, 2.5 and 2.6 show some examples of the typical spectra of confirmed sources and interlopers found in the MUSYC spectroscopic catalog for these three set of sources. LBGs can show their Lyα line in emission or in absorption, while LAEs, due to their selection method, naturally show this line only in emission. The AGN identified in the MUSYC spectra have been mostly of the broad line type, although also some narrow line AGN have been

```
map-spectra:                          subsky:
------------                          -------
SAGORDER = 5                          MINLAMBDA = 4000.0
LAMBDA0 = 4000.0                      MAXLAMBDA = 9000.0
MINPIXL = 200                         EXCLUDE = 5
LAMBDA1 = 9000.0                      DELTAKNOT = 1.0
DISORDER = 5                          SPLINEORDER = 3
TILTORDER = 3                         MEDBOX = 30
SAGITORDER = 3                        SIGLIMIT = 3.0
SLENORDER = 3                         NOISE = 4.9
                                      GAIN = 0.9
adjust-map:
-----------                           extract-2dspec:
SEARCH_HEIGHT = 5                     ---------------
SEARCH_WIDTH = 5                      DELTALAM = 1.0
SLIT_WIDTH = 4                        DELTASLIT = 0.2
ORD_DISP = 5                          TELSCALE = 3.0
ORD_SAG = 1                           SEARCH = 4
MINLINES = 5                          MINLAMBDA = 4000.0
MAGFACTOR = 100                       MAXLAMBDA = 9000.0

Sflats:                               sumspec:
-------                               --------
MINLAMBDA = 4000.0                    CLEAN = "Y"
MAXLAMBDA = 9000.0                    SIGLIMIT = 3.0
TELSCALE = 3.0
CLEAN = "Y"
SIGLIMIT = 3.0
NOISE = 4.9
GAIN = 0.9
ORDER = 5
FIT_MODE = "MEDIAN"
MED_BOX = 100
```

Printout 3: Summary of the parameters used in the reduction of masks between 2005-2007, with 1.0″ slitlets. Some parameters were different for observing runs with other specifications, and furthermore, newer versions of COSMOS have additional parameters for the same programs.

```
PACKAGE = apextract
   TASK = apall

input   =                 List of input images
(output =              ) List of output spectra
(apertur=              ) Apertures
(format =  multispec) Extracted spectra format
(referen=              ) List of aperture reference images
(profile=              ) List of aperture profile images

(interac=          yes) Run task interactively?
(find   =          yes) Find apertures?
(recente=          yes) Recenter apertures?
(resize =           no) Resize apertures?
(edit   =          yes) Edit apertures?
(trace  =          yes) Trace apertures?
(fittrac=          yes) Fit the traced points interactively?
(extract=          yes) Extract spectra?
(extras =          yes) Extract sky, sigma, etc.?
(review =          yes) Review extractions?

(line   =         1110) Dispersion line
(nsum   =          100) Number of dispersion lines to sum or median

                    # DEFAULT APERTURE PARAMETERS

(lower  =          -5.) Lower aperture limit relative to center
(upper  =           5.) Upper aperture limit relative to center
(apidtab=              ) Aperture ID table (optional)

                    # DEFAULT BACKGROUND PARAMETERS

(b_funct=   chebyshev) Background function
(b_order=           2) Background function order
(b_sampl=-25:-5,5:25) Background sample regions
(b_naver=           1) Background average or median
(b_niter=           2) Background rejection iterations
(b_low_r=          2.) Background lower rejection sigma
(b_high_=          2.) Background upper rejection sigma
(b_grow =          1.) Background rejection growing radius
```

Printout 4: Default parameters used in the IRAF task apall.

```
                          # APERTURE CENTERING PARAMETERS

(width  =              5.) Profile centering width
(radius =             10.) Profile centering radius
(thresho=              0.) Detection threshold for profile centering

                          # AUTOMATIC FINDING AND ORDERING PARAMETERS

nfind   =               1  Number of apertures to be found automatically
(minsep =              5.) Minimum separation between spectra
(maxsep =           1000.) Maximum separation between spectra
(order  = increasing) Order of apertures

                          # RECENTERING PARAMETERS

(aprecen=                ) Apertures for recentering calculation
(npeaks =          INDEF) Select brightest peaks
(shift  =            yes) Use average shift instead of recentering?

                          # RESIZING PARAMETERS

(llimit =          INDEF) Lower aperture limit relative to center
(ulimit =          INDEF) Upper aperture limit relative to center
(ylevel =            0.1) Fraction of peak or intensity for automatic
                          width
(peak   =            yes) Is ylevel a fraction of the peak?
(bkg    =            yes) Subtract background in automatic width?
(r_grow =             0.) Grow limits by this factor
(avglimi=             no) Average limits over all apertures?
```

Printout 4: Default parameters used in the IRAF task apall. (cont)

```
                    # TRACING PARAMETERS

(t_nsum =           10) Number of dispersion lines to sum
(t_step =           50) Tracing step
(t_nlost=            3) Number of consecutive times profile is lost
                        before quitting
(t_funct=     legendre) Trace fitting function
(t_order=            2) Trace fitting function order
(t_sampl=            *) Trace sample regions
(t_naver=            1) Trace average or median
(t_niter=            0) Trace rejection iterations
(t_low_r=           3.) Trace lower rejection sigma
(t_high_=           3.) Trace upper rejection sigma
(t_grow =           0.) Trace rejection growing radius

                    # EXTRACTION PARAMETERS

(backgro=          fit) Background to subtract
(skybox =            1) Box car smoothing length for sky
(weights=     variance) Extraction weights (none|variance)
(pfit   =        fit1d) Profile fitting type (fit1d|fit2d)
(clean  =          yes) Detect and replace bad pixels?
(saturat=      65535.) Saturation level
(readnoi=          2.7) Read out noise sigma (photons)
(gain   =         1.29) Photon gain (photons/data number)
(lsigma =           2.) Lower rejection threshold
(usigma =           2.) Upper rejection threshold
(nsubaps=            1) Number of subapertures per aperture
```

Printout 4: Default parameters used in the IRAF task apall. (cont)

Table 2.3.   Initial set of wavelength calibration lines

λ (Å)	Element
4471.48	He I
5015.68	He I
5944.83	Ne I
6334.43	Ne I
6506.53	Ne I
6598.95	Ne I
6717.04	Ne I
6929.47	Ne I
6965.43	Ar I
7245.17	Ne I
7383.98	Ar I
7635.10	Ar I
7724.21	Ar I
7948.18	Ar I
8264.52	Ar I
8521.44	Ar I
8667.94	Ar I

detected too. Figure 2.5 shows a narrow line AGN with velocity widths of <1000 Km/s, showing that resolution in the MUSYC spectra was enough for distinguishing those cases. The broad line AGN example in figure 2.5 also shows an associated absorption system, an interesting feature present in several of the AGN MUSYC spectra. More details on the spectral features identified in these spectra and in the measurement of redshifts and their uncertanties can be found in §3.3, §4.3 and §5.3, respectively.
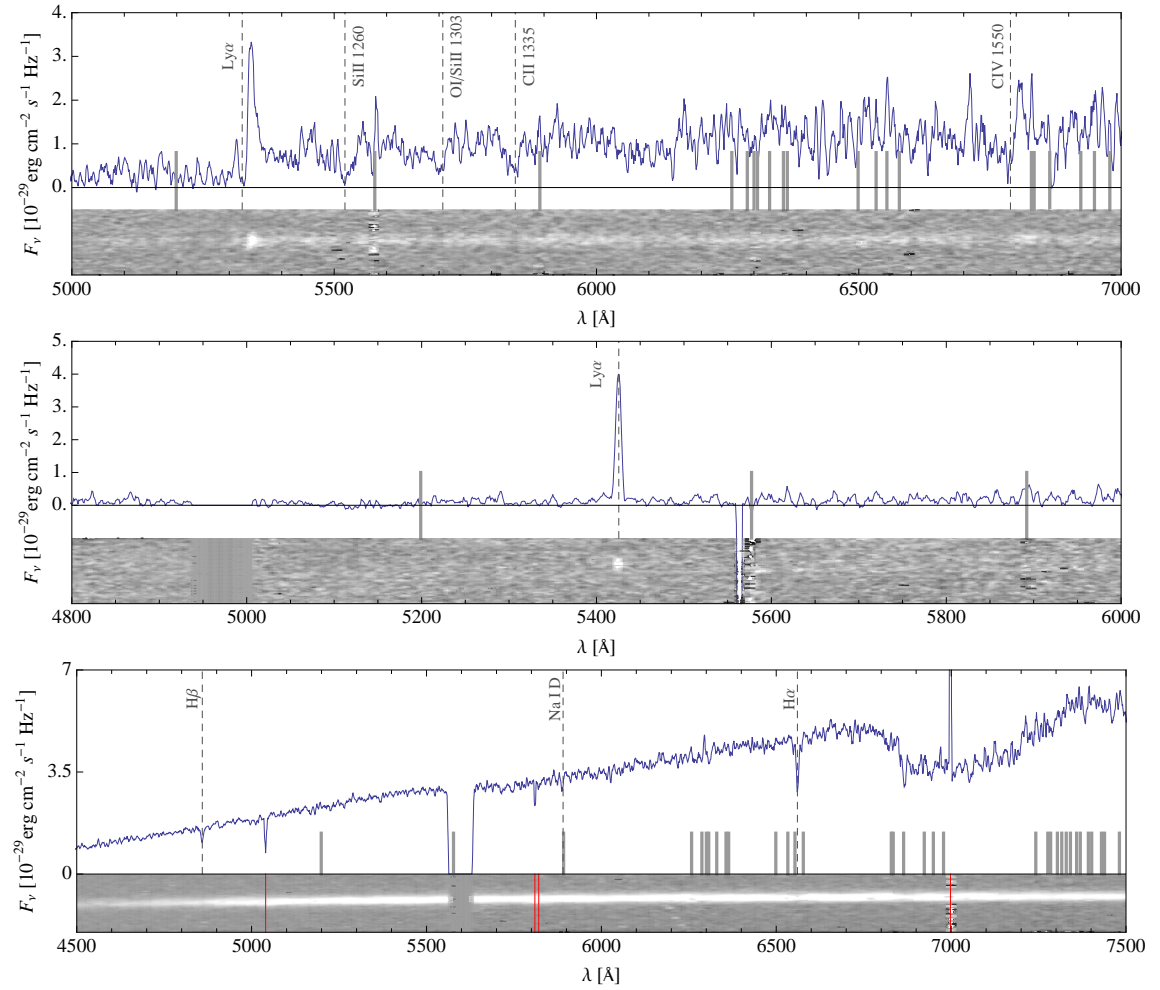
Figure 2.4 Examples of spectra in the MUSYC database, in the observed frame. The image at the bottom of each plot shows the respective 2D image. Gray lines mark the expected location of sky lines, dashed lines mark the location of identified spectral lines. From top to bottom: LBG found in E-CDFS, ID 12485, redshift z=3.38, magnitude R=22.8; LBG in E-CDFS, ID 53532, z=3.46, R=25.2; M star in HDFS, ID 46948, R=19.2, with a radial velocity of -108 Km/s. This last one is a typical interloper in the LBG selection technique.
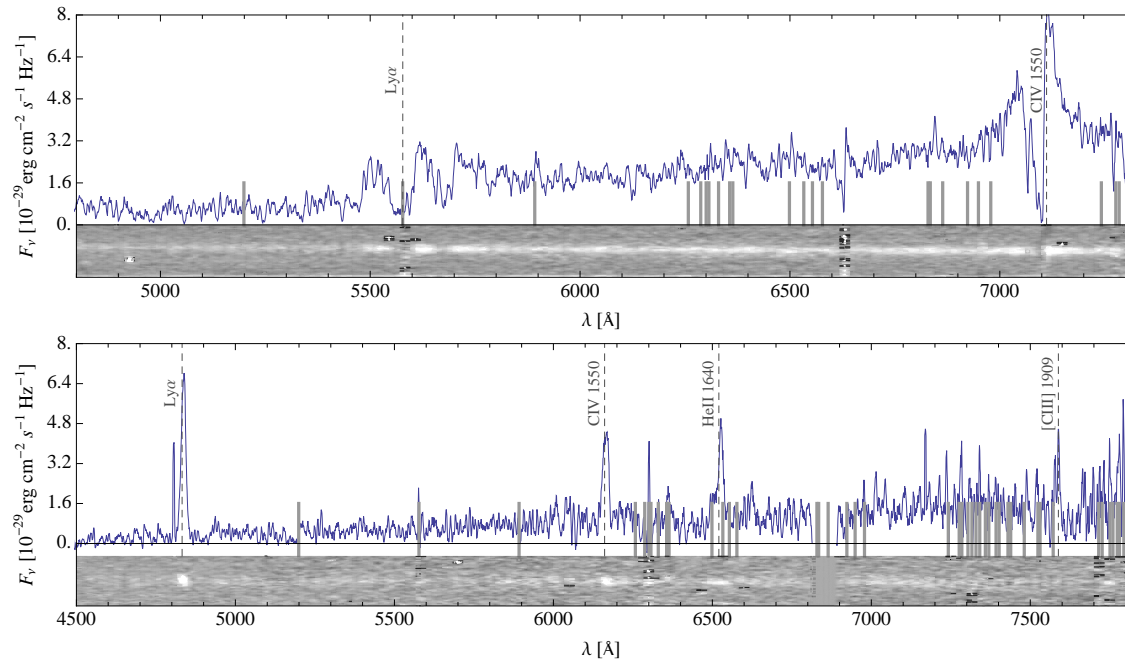
Figure 2.5 More examples of MUSYC spectra. From top to bottom: Broad line AGN in E-CDFS, ID 30393, z=3.6, R=22.4; Narrow line AGN in SDSS 1030, ID 6927, z=2.975, R=23.5.
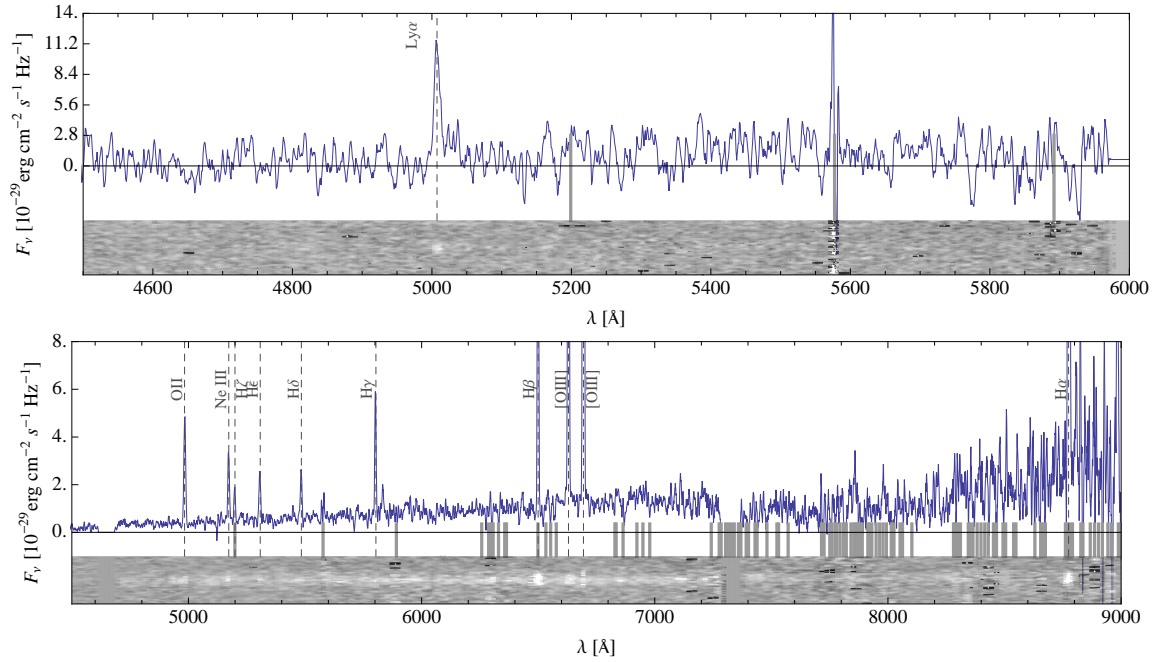
Figure 2.6 More examples of MUSYC spectra. Top: LAE found in SDSS 1030, ID 45706, redshift z=3.119, magnitude R=25.2. Bottom: Emission Line Galaxy in CW 1255, ID 36177, z=0.337, R=22.2, a typical interloper in the LAE samples.

Chapter 3

The Clustering of Lyman-Break Galaxies

3.1 Introduction

Since the successful spectroscopic confirmation of a large population of normal, star-forming galaxies at high redshift by Steidel et al. (1996a,b), increasingly larger and more homogeneous samples have been gathered. This opened the possibility to measure the clustering strength of galaxies in the young universe, given the efficiency of the Lyman-Break selection: the rate of low-redshift interlopers encountered is small and the redshift range that this method to targets is reasonably restricted, and hence well suited for clustering studies. Previous to this breakthrough, all known samples with a significant number of high redshift galaxies had active nuclei, detected by their radio emission or optical colors, and were therefore not good representatives of the typical galaxy in the young universe (see, e.g. McCarthy, 1993, and references therein). In contrast, the Lyman-Break technique selects mostly blue, starforming, but otherwise normal galaxies.

Ever since their discovery, Lyman-Break Galaxies (LBG) have been found to

cluster significantly. Giavalisco et al. (1994) found a significant overdensity of LBGs around the damped Lyα absorber of quasar Q0000-2619. Steidel et al. (1998) found a large spike in the redshift distribution of spectroscopically confirmed LBGs in their SSA22a field and suggested that these galaxies might be hosted by fairly massive ($\sim 10^{12}$ M$_\odot$) dark matter halos. From a count-in-cells analysis in a spectroscopic sample of 268 LBGs in an area of 0.1 deg^2, Adelberger et al. (1998) gave an estimate for the correlation length r_0 between 7-10 Mpc, and from the galaxy number density they observed, suggested that these galaxies have a nearly one-to-one correspondence with their hosting halos.

In a sample of 871 candidates at $z \sim 3$, with a spectroscopic database of 376 confirmed ones and within an total area of 0.3 deg^2, Giavalisco et al. (1998) were able to measure the correlation function of LBGs, and obtained a correlation length of $r_0 \sim 7$ Mpc. Afterwards, Giavalisco & Dickinson (2001) were able to confirm the previous results with greater accuracy and additionally showed that the LBG clustering strength scales with UV luminosity. They used an increased ground based sample of 1240 candidates over 0.4 deg^2 (with 547 spectroscopically confirmed), and a much deeper sample based on HST observations with 271 candidates over 0.003 deg^2 (28 spectroscopically confirmed). Ouchi et al. (2001, 2004) extended these studies to higher redshifts, using the data from the Subaru Deep Field, where they obtained samples of 2170 LBGs at $z \sim 4$ and 386 at $z \sim 5$, although with almost no spectroscopic confirmation and basing their analysis purely on photometric reshifts. Continuing this line of work, Ouchi et al. (2005) managed to probe the dependence of clustering strength with luminosity at $z \sim 4$, and to study the small scale departure of the correlation function from a power law.

Adelberger et al. (2005b) further refined the work of Giavalisco & Dickinson (2001) by increasing the total area covered to 0.8 deg^2, the photometric LBG sample to 4934 and the spectroscopic sample to 836, and obtained a fairly precise measurement of the spatial correlation length, $r_0 = 6 \pm 1$ Mpc (with $\gamma = 1.6$) and compared it to the populations of BM and BX galaxies at $1 < z < 3$ (see Adelberger et al., 2004 for more details on these other type of color selections). They conclude that all these galaxies end up in halos that typically host elliptical galaxies at present. More recently, Hildebrandt et al. (2007), using the dataset from the ESO Deep Public

Survey with 8826 LBG candidates over 2 deg^2, managed to increase the precision in this measurement, and obtained a value of $r_0 = 6.9\pm0.4$ Mpc (with $\gamma = 1.6$). Their additional statistical precision allowed them to more finely probe the variation of clustering with luminosity, and study the small-scale clustering signal in some detail. They had only photometric redshifts for their sample, however, and also found their sample to be more contaminated than the two authors mentioned above, therefore they may have suffered from additional systematic errors.

Since the redshift distribution and contamination fraction of a sample of LBG candidates depend on the shape of the filters used for the optical imaging as well as on the depth achieved in each filter, extrapolating the redshift distribution obtained in one survey to another, or using only photometric redshifts, can lead to unknown or complicated systematics which are difficult to constrain. On the other hand, these pencil beam, deep surveys cover little area when compared to low-redshift surveys, and have therefore, a significant contribution both from poisson noise and "cosmic variance" due to large scale structure in their clustering measurements. The MUSYC survey, with its combination of deep and wide (relative to high-redshift surveys) coverage, and with its fairly extensive spectroscopic database can, therefore, contribute significantly to the study of LBG clustering at $z \sim 3$.

3.2 Lyman-Break selection

The idea behind the Lyman-Break selection technique is to take advantage of the break caused by hydrogen absorption in the spectrum of galaxies at restframe 912Å. This break is present already in the spectra of early type stars, but gets enhanced by interstellar H I absorption inside the galaxy, and even more pronounced by the absorption of the intergalactic hydrogen during the travel of the light towards earth. Figure 3.1 shows a LBG template being redshifted and the relevant filter bandpasses. At redshift $\lesssim 3$ the Lyman-break is redshifted into the U band, giving the galaxy a red color in $U - V$, while maintaining a fairly blue $V - R$ color. When further redshifted, the break passes the U filter and the galaxy is no longer detected in that filter. At that point, the galaxy will approach $z \sim 4$ and could be selected in an

analogous way by using the $B - R$ and $R - I$ colors.

To select our LBG samples, we used color criteria similar to Steidel et al., 1996b, adapted to the optical bandpasses in MUSYC (Gawiser et al., 2006a). Fig. 3.2 shows the $U - V$ versus $V - R$ color-color plot for sources in each of the fields. Lyman break color selection ('UVR') corresponds to the region outlined in the upper left side, namely:

$$(U - V) \geq 1.2 \tag{3.1}$$
$$-1.0 \leq (V - R) \leq 0.6 \tag{3.2}$$
$$(U - V) - 3.8(V - R) \geq 1.2 \tag{3.3}$$
$$23.0 \leq R \leq 25.5 \tag{3.4}$$

For this color selection, U-band fluxes are required to be detected at 1-σ, and are otherwise set to an upper limit equal to their 1-σ error. This avoids interlopers (typically dwarf stars) with uncertain photometry at the cost of some incompleteness in the sample. The last constraint is a magnitude cut required to allow for spectroscopic confirmation. The diagonal cut in these diagrams avoids red dwarf stars and low-z galaxies that might satisfy the $U - V$ criteria due to heavy reddening. As reported by Adelberger et al. (2004), we also have found that the additional bright magnitude cut of $R \geq 23$ also contributes to reducing the number of interlopers, especially stars (see next section). In the full spectroscopic sample (without the bright magnitude cut), we found 9 interlopers and 122 LBGs, whereas applying this cut reduces the number of interlopers to 4 while reducing the nuber of LBGs only to 119. This translates into a reduction of the contamination rate from $\sim$7.4% to $\sim$3.4% in the spectroscopic sample. Figure 3.3 shows the sky distribution of these galaxies samples in the four MUSYC fields, and table 3.1 lists the number of objects in each sample. Table 3.2 lists the number of UVR candidates observed, the number of confirmed LBGs and the number of low-z interlopers found.
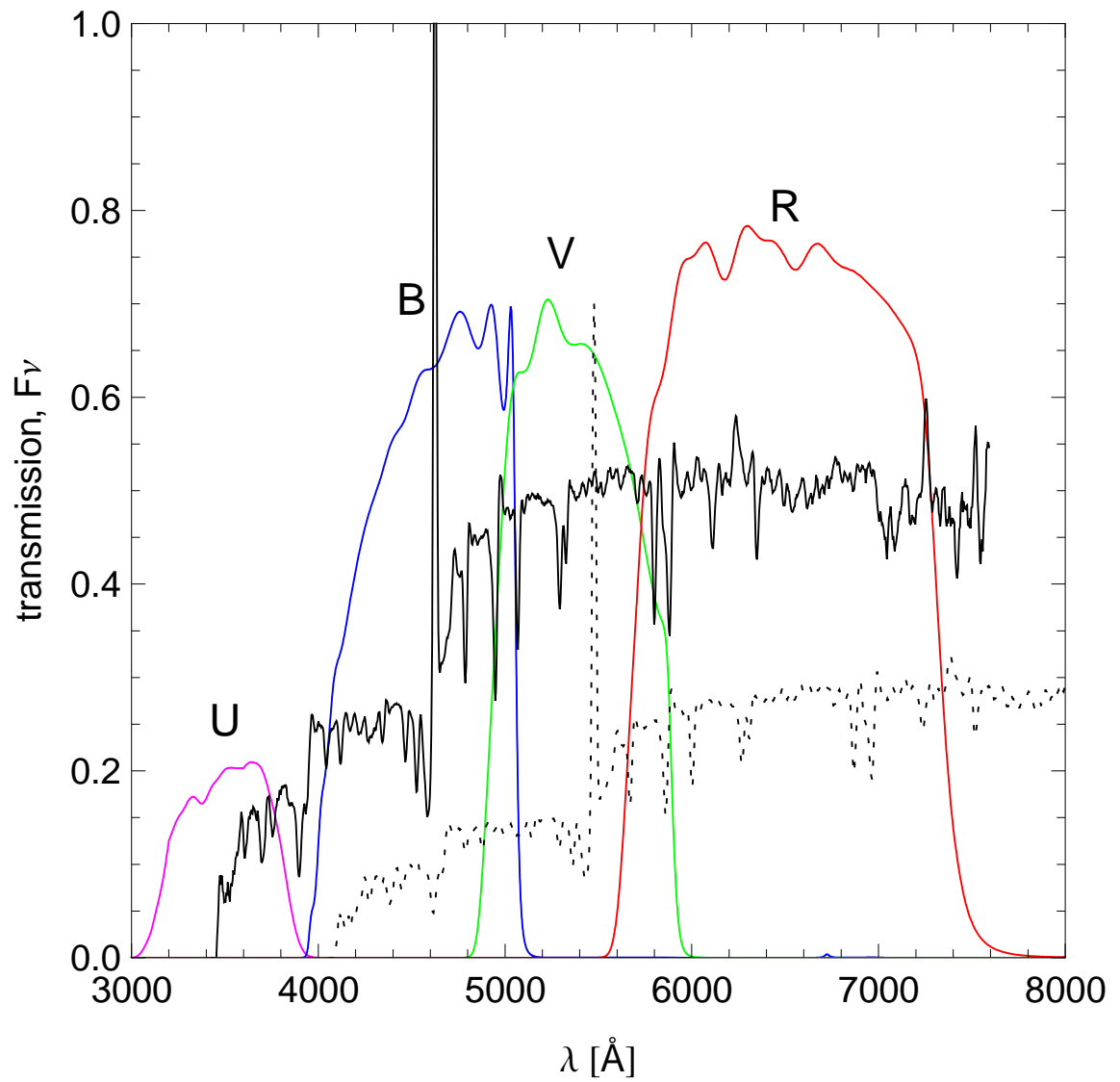
Figure 3.1 Total transmission curves for the U, B, V and R filters drawn over a LBG template at z=2.8 (solid) and at z=3.5 (dotted). The spectra are in units of F$_\nu$, but have been arbitrarily moved in the y-axis for clarity.
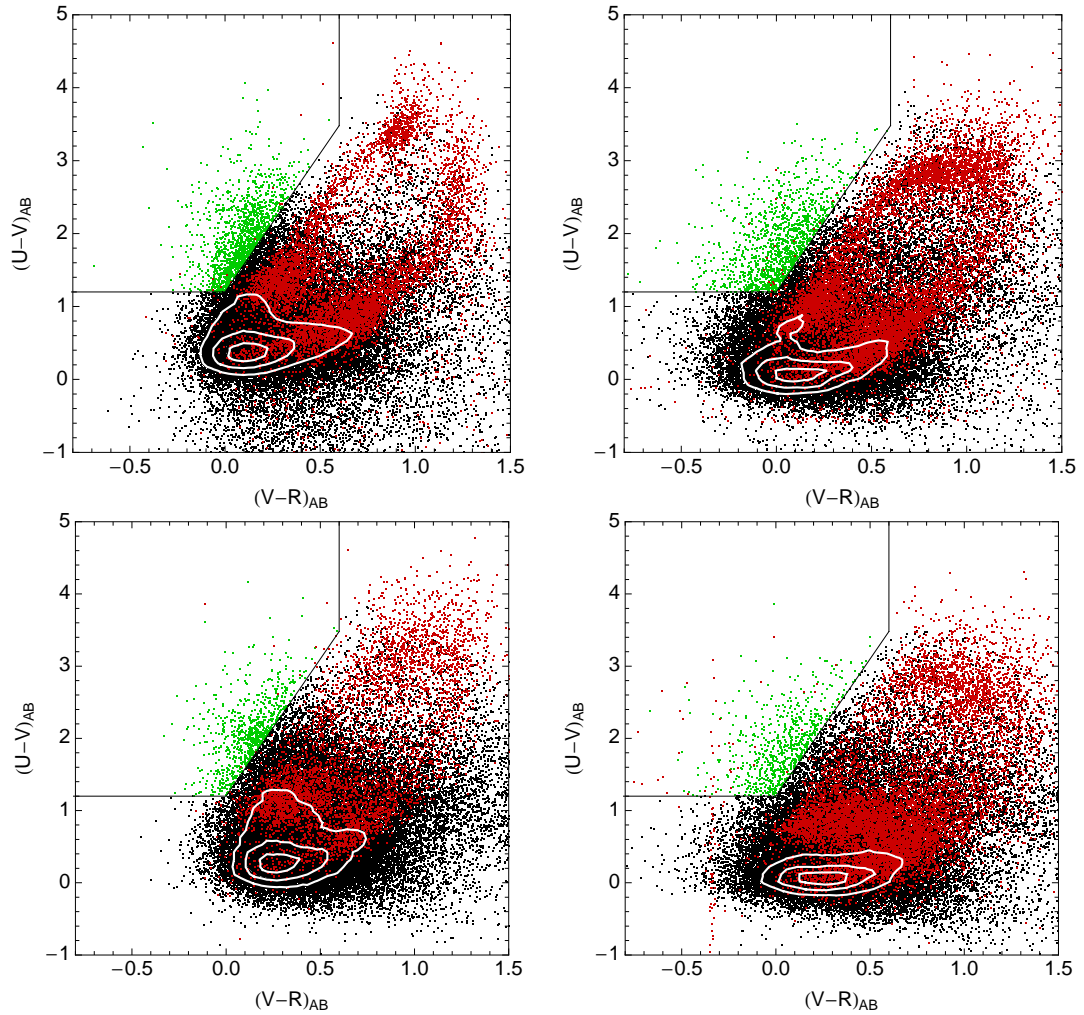
Figure 3.2 $U - V$ versus $V - R$ color-color plot for all the sources brighter than $R = 25.5$ in ECDF-S (upper left), EHDF-S (upper right) SDSS1030 (lower left) and the CW1255 (lower right) fields. Sources brighter than $R = 23.0$ are marked red, and the locus of dwarf stars can be clearly seen below the diagonal color cut of the Lyman-Break selection. The UVR selected candidates are marked in green. In the densest region of the plot, contours are drawn for clarity at 50, 70 and 90% of the maximum source density in each diagram.
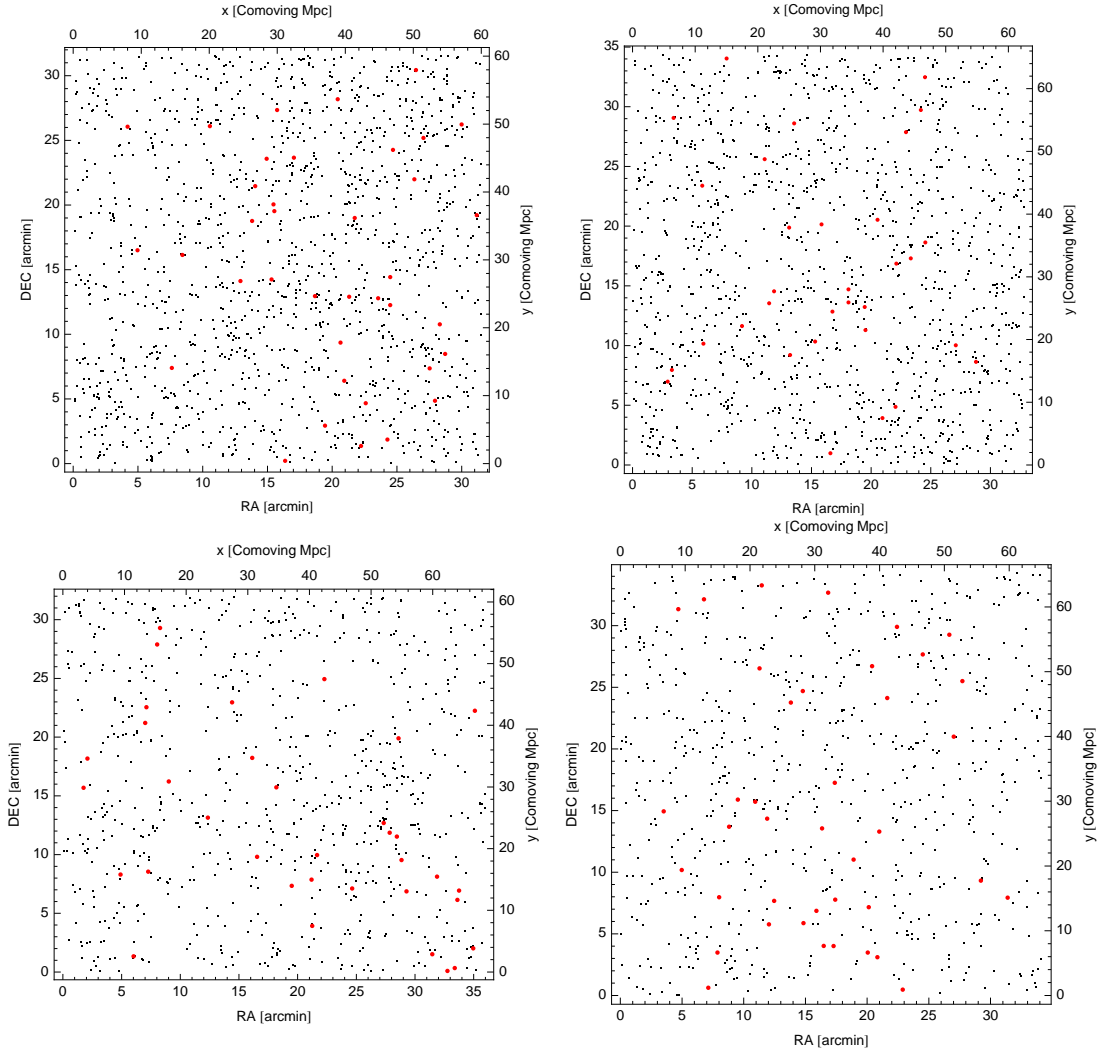
Figure 3.3 R.A.-Dec. distribution of the four UVR samples selected in ECDF-S (upper left), EHDF-S (upper right) SDSS1030 (lower left) and the CW1255 (lower right) fields. Black points are all LBG candidates, while red circles correspond to spectroscopically confirmed LBGs. The gray regions in the SDSS1030 field correspond to manually excised regions that contain bright stars that caused bleeding in the R.A. direction and an increase in the surrounding background level. The axes in comoving Mpc were calculated at the mean redshift $z = 3.2$.

Table 3.1. Field depths and photometric LBG samples

Field	area [arcmin^2]	U depth [mag]	V depth [mag]	R depth [mag]	LBGs
ECDF-S	997.9	26.0	26.4	26.4	1388
EHDF-S	1137.0	26.0	26.0	25.8	1276
SDSS 1030	1154.4	25.7	26.2	26.0	774
CW 1255	1173.4	tbd	tbd	tbd	709
4 Fields	4462.7	–	–	–	4147

Note. — Bandpasses depths are for 5-σ point source detection, in AB magnitudes. The depths in CW 1255 have not been determined at the time of this publication.

Table 3.2. Spectroscopic LBG subsamples

Field	Observed	Confirmed	Interlopers
ECDF-S	190	29	3
EHDF-S	199	30	0
SDSS 1030	128	29	0
CW 1255	224	31	1
4 Fields	741	119	4

3.3 The spectroscopic subsample

To assign slits to LBG candidates in the IMACS masks, the candidates were only separated by R magnitude: a few $R < 23$ exceptional sources were assigned higher priority, a second group having $23 < R < 24$ was assigned a lower priority and a third group, having $24 < R < 25.5$ was assigned the lowest priority. Since many other MUSYC samples were also created for different science projects, the absolute priority in each multislit mask for each of the UVR selected samples varied from mask to mask. From the spectroscopic follow-up database described in §2.3, we have spectroscopically confirmed over 100 LBGs. The main feature present in the spectra that allowed for an identification and redshift determination was the presence of the Lyα line in emission and a decrement in the spectrum from redward to bluewards of that line. This decrement is due to the Lyα forest absorption and at this redshift is very noticeable, even in fairly noisy spectra. The Lyα line is not always present in emission, however. See Shapley et al. (2003) for a high signal-to-noise sequence of examples showing this line from high equivalent width emission to almost complete absorption. Additionally, for sources with a bright, high signal-to-noise continuum, it was also possible to identify several interstellar absorption lines in their spectra. As can be seen in the example of the bright LBG source in figure 2.4 of the previous chapter, sometimes the C IV 1550 line appears with a P-Cygni profile, showing absorption as well as some emission. For the Lyman-Break selection at z=3, typical interlopers are M dwarf stars, which can be readily differentiated from LBG with Lyα in absorption due to their characteristic continuum shape, especially after flux calibrating the spectrum. To determine the central wavelength of emission or absorption lines we used the option 'e' in the `splot` IRAF task. The list of spectroscopically confirmed LBG can be found in appendix A, and their positions in each of the fields are plotted in figure 3.3. As can be seen in those plots, thanks to the wide field of view of the IMACS spectrograph and the fact that this sample was obtained in multiple observing runs, the spectrocopic sample is uniformly distributed over the field. Figure 3.4 shows the redshift histogram of all UVR selected sources that yield a successful identification. The gray distribution in that plot, after removal of low-redshift interlopers, has a mean and standard deviation of $\bar{z} = 3.23 \pm 0.02$ and $\sigma_z = 0.22 \pm 0.01$. Naturally, one might question whether this distribution faithfully
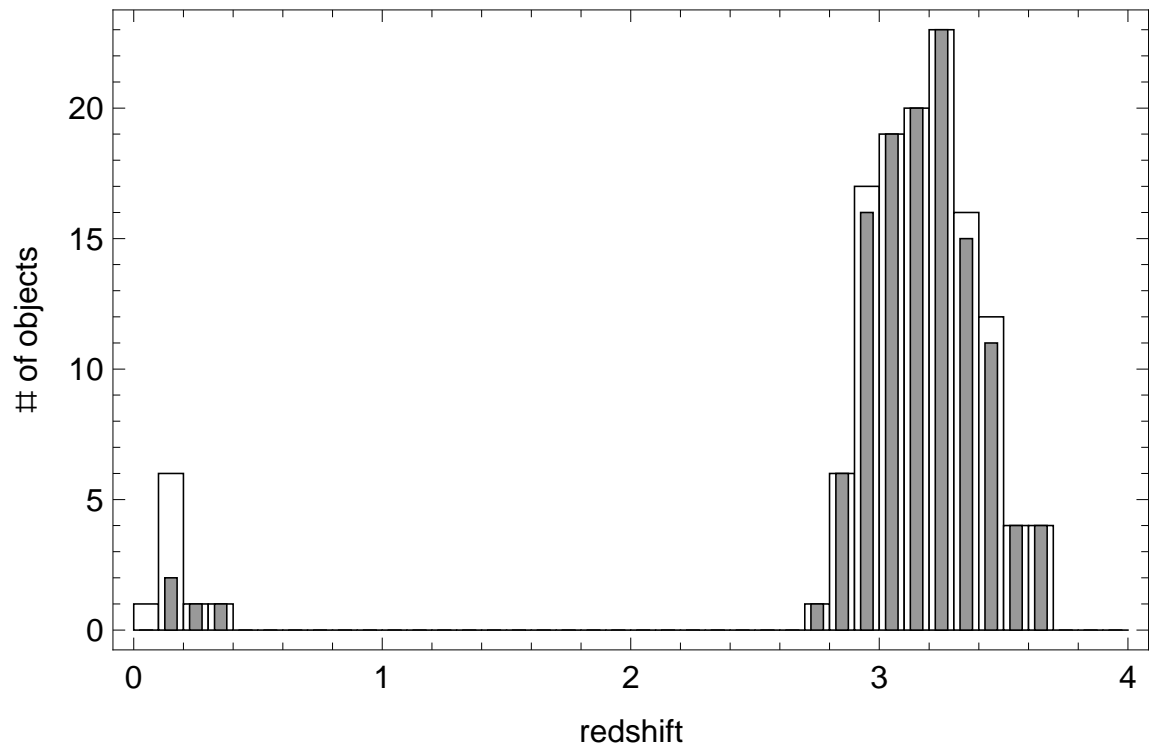
Figure 3.4 Redshift histogram of confirmed LBG, including identified interlopers. White bars are all UVR selected sources without the bright magnitude cut $R > 23$, while gray bars are the sources that additionally comply with that requirement. As can be seen to the left of the histogram, the number of interlopers, especially stars, are greatly reduced.

represents the overall population of LBGs that we have targeted. We are going to ponder this issue in the rest of the section.

3.3.1 Constraining biases in the spectroscopic sample

The first possible bias that might affect our determination of the redshift distribution is the redshift measurement of the individual galaxies. Whenever possible, the redshift for LBGs was determined from the absorption lines, since the Lyα line often appears redshifted with respect to the stellar systemic velocity, as well as skewed to the red (Franx et al., 1997). However, since for most spectra we only see a Lyα emission line, there is some uncertainty in the systemic redshift determination. In

a sample of 323 LBG spectra showing low-ionization insterstellar absorption lines, Shapley et al. (2003) measured the velocity offset between those lines and the Lyα emission. The distribution of their measured offsets had a mean $\Delta v = 650$ Km/s and a standard deviation $\sigma_{\Delta v} \sim 300$ Km/s. In the observed frame, this translates into a bias in the redshift determination of $\Delta z = (1+z)\Delta v/c = 0.008$ and a dispersion $\sigma_z = 0.004$. The uncertainty added into the deprojection of the angular correlation function is, however, immaterial. To see this explicitly, we take eq. 1.28 for the autocorrelation case and for an order of magnitude estimation, replace the comoving radial distance distribution of our LBG sample $p(r)$ with a simple top-hat of center R_c and width W,

$$p(r) = \begin{cases} \frac{1}{W}, & |r - R_c| \leq \frac{W}{2} \\ 0, & |r - R_c| > \frac{W}{2} \end{cases}. \tag{3.5}$$

Since the width of the radial distance distribution is fairly small compared to the mean distance, we can aproximate:

$$\begin{aligned} A &= r_0^\gamma B\left(\frac{1}{2}, \frac{\gamma - 1}{2}\right) \int_0^\infty d\bar{r}\; \bar{r}^{1-\gamma} p^2(\bar{r}) \\ &\approx r_0^\gamma B\left(\frac{1}{2}, \frac{\gamma - 1}{2}\right) \times W \times R_c^{1-\gamma} \times \frac{1}{W^2} \\ \Longrightarrow \ln(A) &= \gamma \ln(r_0) + \ln(B) + (1 - \gamma)\ln(R_c) - \ln(W). \end{aligned} \tag{3.6}$$

From this last relation we can propagate the relative uncertainty in the clustering length r_0 due to ΔR_c and ΔW:

$$\frac{\Delta r_{0,R_c}}{r_0} = \frac{\gamma - 1}{\gamma}\frac{\Delta R_c}{R_c} \quad \text{and} \quad \frac{\Delta r_{0,W}}{r_0} = \frac{1}{\gamma}\frac{\Delta W}{W}, \tag{3.7}$$

respectively. The comoving radial distance distribution for our LBG sample has a mean $R_c \sim 6600$ Mpc and a width of $W \sim 1100$ Mpc, and translating the redshift variance and dispersion mentioned above into comoving distances gives $\Delta R_c \approx 7$ Mpc and $\Delta W \approx 4$ Mpc. Evaluating eq. 3.7 with these numbers yields (using a fiducial slope of $\gamma=1.8$):

$$\frac{\Delta r_{0,R_c}}{r_0} \sim 5 \times 10^{-4} \quad \text{and} \quad \frac{\Delta r_{0,W}}{r_0} \sim 2 \times 10^{-3}, \tag{3.8}$$

values that are completely negligible when compared with the pure random errors in the determination of the redshift distribution. This means that neither the shift in the mean nor the broadening in the LBG redshift distribution due to the individual biases and errors are a relevant issue.

Another possible systematic is a bias in the sample of LBGs with redshift against faint sources. As can be seen in figure 3.5, the spectroscopically confirmed LBG have indeed a different apparent magnitude distribution, and sources with $R > 24.5$ are underrepresented. This is not a significant bias, however, since the redshift distribution of bright sources does not differ much from the distribution of faint sources. The sample means and standard deviations for the two distributions in this last plot are: $\bar{z}_{\mathrm{bright}} = 3.18 \pm 0.03$, $\sigma_{\mathrm{bright}} = 0.21 \pm 0.02$ and $\bar{z}_{\mathrm{dim}} = 3.21 \pm 0.02$, $\sigma_{\mathrm{dim}} = 0.19 \pm 0.02$, respectively. The difference between these two distributions is fairly small, barely distinguishable within sample errors. A K-S test on these two samples yields a p-value of 0.63, and thus we cannot reject the hypothesis that these samples come from the same distribution. In spite of that, we used them both in the deprojection step to give an upper and lower limit due to this uncertainity in our measurements. This mild dependence of the redshift distribution on apparent magnitude is due to the modest change in luminosity distance from the front to the back of the survey, relative to the intrinsic spread in UV luminosities of our sources. We calculated the absolute magnitude at $\lambda \sim 1550$Å for our LBG sample from the total R magnitude, ignoring K-correction:

$$M_{1550} = R - 5\log(D_L(z)) + 5 + 2.5\log(1 + z) \qquad (3.9)$$

where $D_L = (1 + z)D_c$ is the luminosity distance. The UV spectra of LBG redwards of the Lyα line are fairly constant in flux density (F_ν), as can be seen in fig. 3.7, justifying neglecting the K-correction in the formula above. Fig. 3.6 shows the relation between the absolute and apparent UV magnitude computed this way. The dispersion in that relation (generated by the spread in redshift distribution) is fairly small when compared to the range of absolute magnitudes spanned by our LBG sample, and thus, little redshift information is lost at the apparent magnitude faint end.
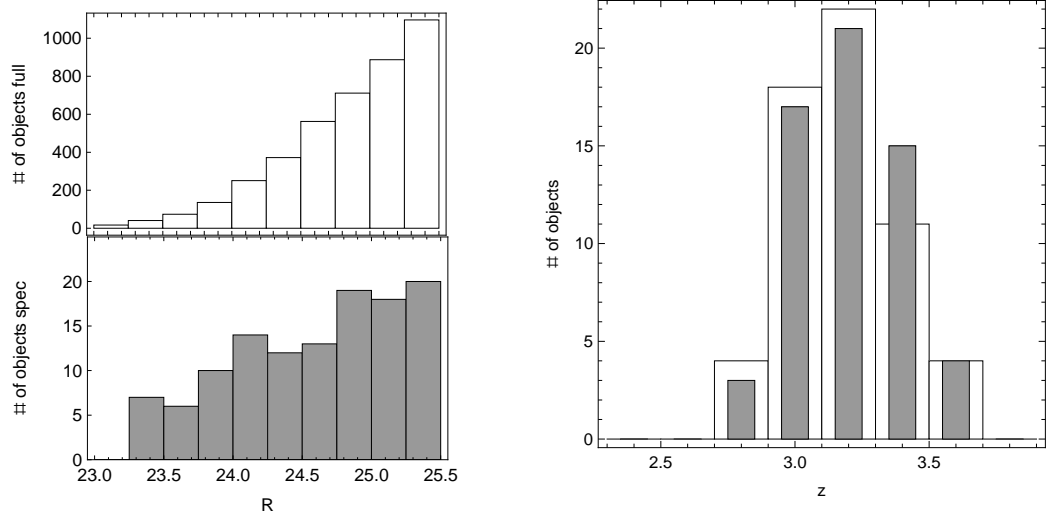
Figure 3.5 *Left:* Apparent magnitude of the spectroscopic LBG sample (black bars) versus the full catalog of LBG candidates (white bars). Around $R \sim 24.7$ and above the spectroscopic sample seems to underrepresent the full photometric selection. *Right:* Redshift histograms of bright (white bars) and dim (gray bars) spectroscopically confirmed LBGs. The spectroscopic sample is split at R=24.7, this makes the bright sample total 59 sources and the dim sample 60.
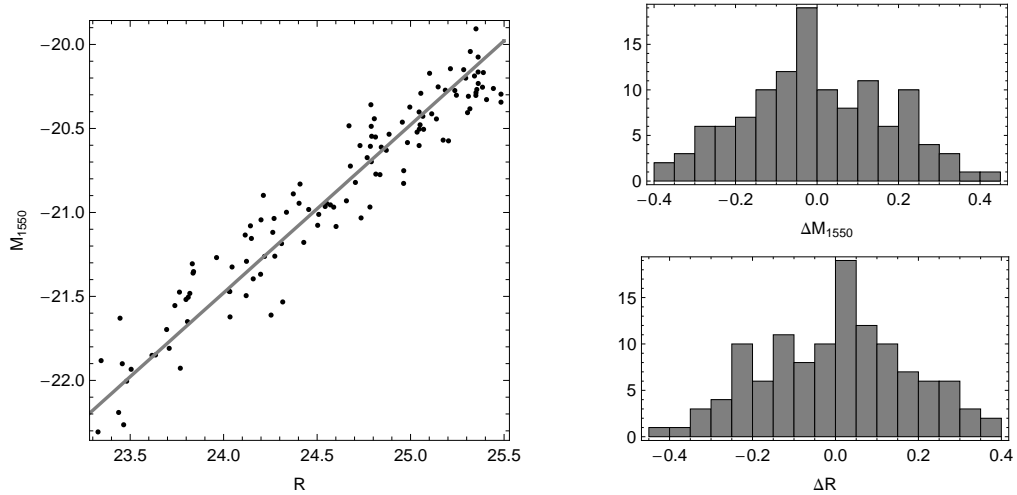
Figure 3.6 *Left:* Absolute magnitude M_{1550} versus apparent magnitude R of spectroscopically confirmed LBG. Gray line is the best linear fit, corresponding to constant z=2.9. *Right:* Dispersions in the X and Y axis of the previous plot with respect to the linear fit. Both show a standard deviation $\sigma_R = \sigma_{M_{1550}} = 0.17$ mag. From the dispersion of $\sigma_z = 0.2$ in the redshift distribution alone one expects a dispersion in M_{1550} of 0.11 mag.

A third possible bias in the spectroscopic sample comes from the fact that emission lines are much easier to identify than absorption lines. Shapley et al. (2003) generated 4 high signal-to-noise LBG template spectra dividing their sample by Lyα equivalent width (EW) into four quartiles, $\sim$200 spectra each. From these templates Lucia Guaita (priv. comm.) calculated the $U - V$ and $V - R$ colors for the MUSYC bandpasses, as a function of redshift. The effect of the intergalactic medium (IGM) in the empirical templates was first removed, dividing by the average intergalactic absorption over the redshift distribution of Shapley et al.'s sample, given in Adelberger et al. (2005b). Then the templates were redshifted, the IGM absorption at each particular redshift was reapplied, and finally the colors for the MUSYC filters were computed. To approximate the IGM absorption function, the analytical expression from Madau (1995) was used. Figure 3.7 shows these four templates and the expected track in the UVR color-color diagram, together with the UVR colors of confirmed LBG in the MUSYC spectroscopic catalog. Due to the Lyα line in emission, spectra that resemble the blue, orange and green templates are much easier to spectroscopically identify than a spectrum that has the shape of the red template.

It is important to notice that, although the redshift distribution reaches up to $z \sim 3.8$, the $U - V$ color of the spectroscopically confirmed LBG does not follow, in general, the theoretical tracks up to such red $U - V$ colors, at the top of the diagram. This is due to the way we are handling the dropouts in the U filter, which gives every non-detection in U a fairly conservative estimate of the U flux, and therefore, a fairly blue $U - V$ color.

Another important thing to notice is that the difference between the color-color tracks of the four templates is mostly a shift in the $V - R$ color, and it is the upper EW quartile, the blue track from the spectrum showing the greatest Lyα emission line, the one that most prominently separates from the rest of the LBGs. This makes the bias against absorption line LBGs less important, since the the red track separates relatively little from the orange and green tracks, which contain 50% of Shapley et al.'s LBG sample. It is interesting to estimate the effect of the systematic color difference between the blue curve and the rest in the redshift distribution determination. We separated the spectroscopic sample by $V - R$ color in two sets
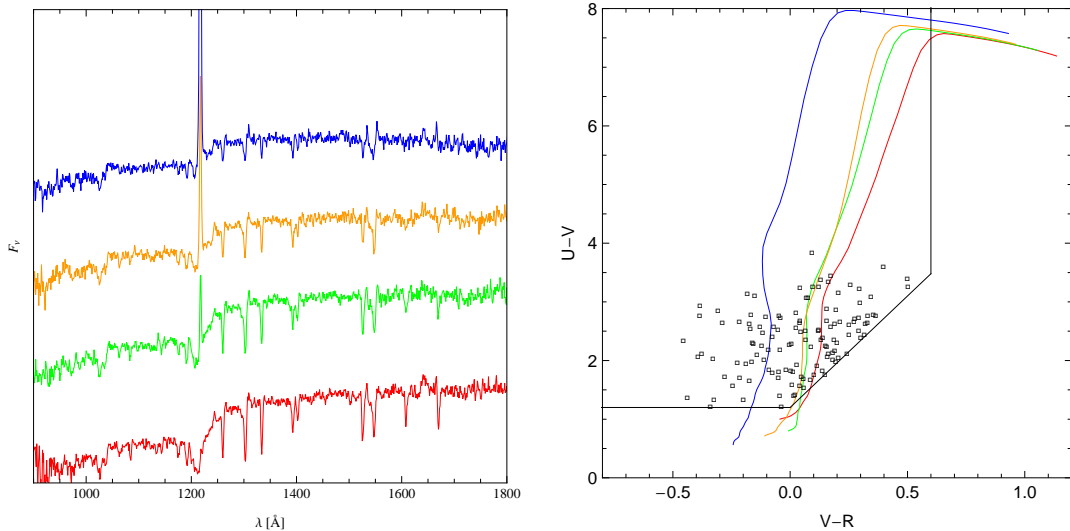
Figure 3.7 *Left:* Shapley et al.'s templates for the four quartiles of Lyα EW in their sample. Each spectrum is the average of 200 individual LBG spectra. *Right:* Color evolution in the UVR diagram for each of the templates shown to the left, color coded in the same way, from z=2.2 to 3.8. The colors of spectroscopically confirmed LBG are marked in open boxes. The Lyman-break selection region is plotted in thick black lines.

and compared the resulting distributions. Figure 3.8 shows these two samples. The sample means and standard deviations for these two are: $\bar{z}_{\mathrm{red}} = 3.24\pm0.03$, $\sigma_{\mathrm{red}} = 0.21\pm0.02$ and $\bar{z}_{\mathrm{blue}} = 3.16\pm0.02$, $\sigma_{\mathrm{blue}} = 0.18\pm0.02$, respectively. Although this difference is greater than the one obtained between the brightness separated samples, a K-S test on these new two samples yield a p-value of 0.08, rejecting the null hypothesis that these samples come from the same distribution only at the 8% (or 1.75σ) significance level. However, in this case neither subsample is representative of the entire LBG UVR color distribution, so we do not apply any correction to the redshift distribution based on color.

To better constrain the effect of color biases in our redshift distribution, we have simulated the Lyman-Break color selection itself, using the templates from Shapley et al. (2003) mentioned above. The UVR colors of the four templates were linearly interpolated to create a grid of LBG models forming a continuous distribution from the lowest Lyα EW (red template) to the highest Lyα EW (blue template), see figure 3.9. This is done for the four fields separately, taking into
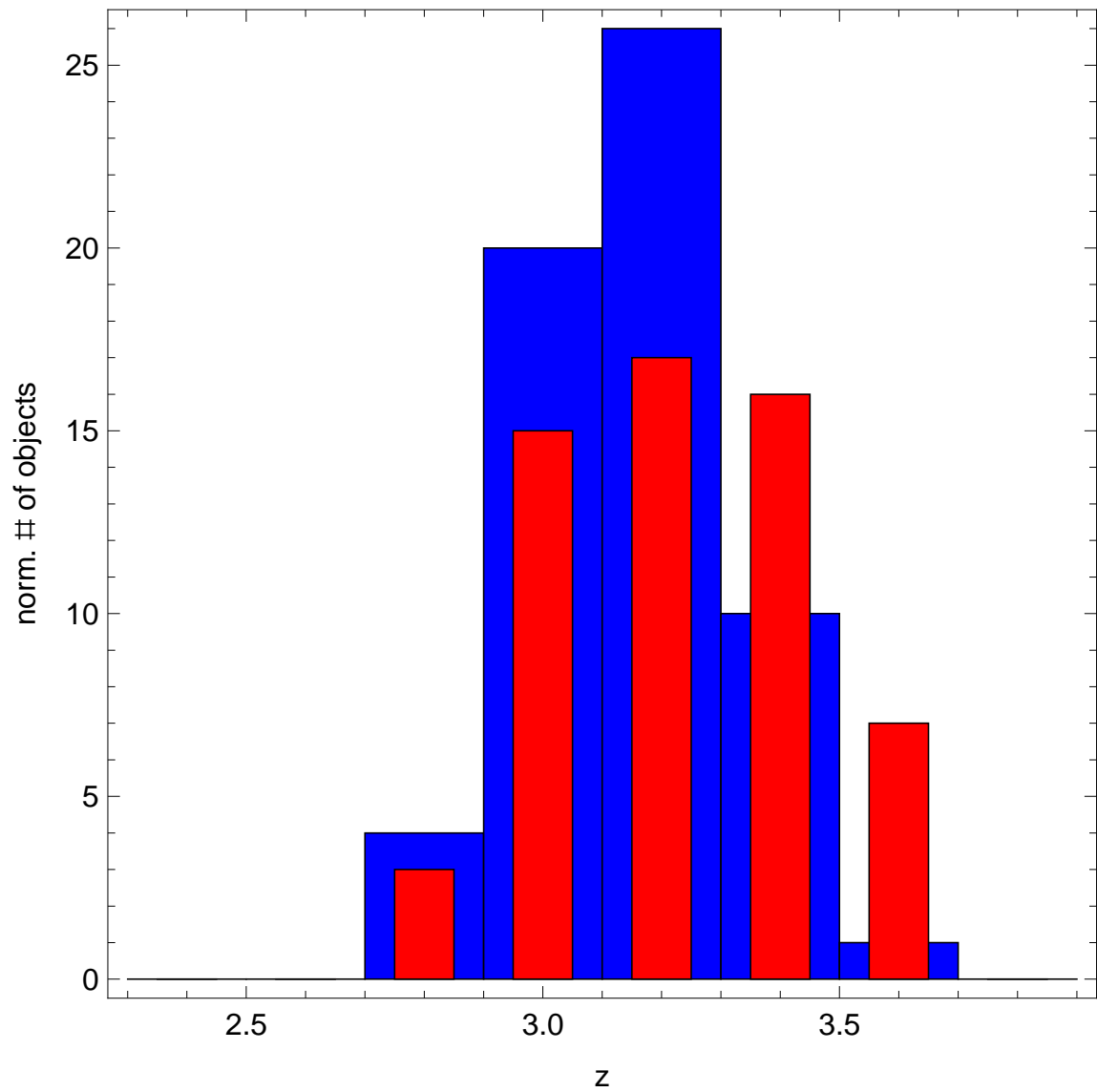
Figure 3.8 Redshift histograms of spectroscopically confirmed LBG with $V-R > 0.05$ (red bars) and $V - R < 0.05$ (blue bars).

account the slight difference in filter transmission curves. To populate this grid, we took advantage of the fairly small effect luminosity distance has on apparent magnitude within the redshift range of interest (see figure 3.6). We assigned to our simulated sample of sources a redshift between 1.5 and 4.0, following a uniform distribution. Independently, we also assigned them an R magnitude following the distribution observed for the photometric catalog of LBGs (figure 3.5), and for the EW "coordinate', we also independently generated a uniform distribution within the EW range we wanted to probe. From the EW-z grids mentioned above, and choosing randomly between the four fields, we then calculated the $U - V$ and $V - R$ colors expected for each of the sources, and the apparent magnitudes in U and V. We then added the typical noise to U, V and R that is found in the optical catalog and recomputed the colors. Finally, we selected our simulated LBG following the same color criteria applied to the optical catalog. The most dramatic difference that EW can make in the selection can be seen by comparing the redshift distribution of highest Lyα EW sources with the distribution of sources with Lyα in absorption. Left panel in figure 3.10 shows the range expected to appear between those two set of sources. This situation is far from what would be expected for our spectroscopic LBG sample, where even sources with restframe EW of a few Å have been confirmed. In this case, the most realistic simulation of our sample would be to generate the redshift distribution from an EW ranging between the blue and green templates from figure 3.7, as shown in the right panel of figure 3.10. The difference between this distribution and the one obtained by considering all four templates is negligible, though, due to the small color differences between the track of the red, green and orange templates mentioned before.

3.3.2 Survey volume and LBG number density determination

To calculate the total comoving volume covered by our LBG survey, we integrated the comoving volume element over the redshift distribution:

$$V = \Omega \int_0^\infty \frac{dV}{d\Omega dz} q(z) dz \tag{3.10}$$
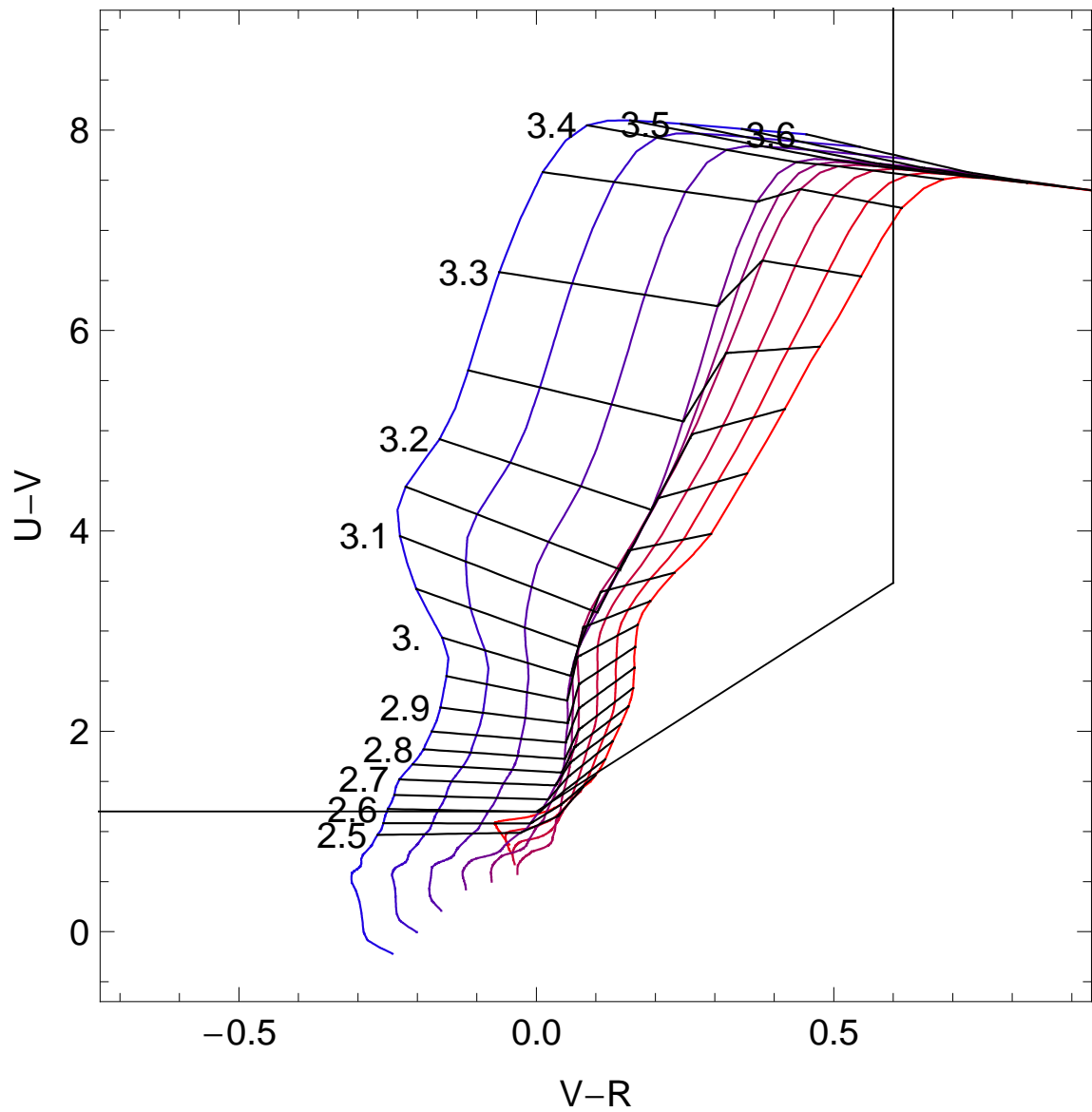
Figure 3.9 EW-z grid of interpolated LBG template colors in the ECDFS. Color curves corresponds to constant EW, ranging from red (Lyα absorption) to blue (highest EW Lyα emission). Black curves have constant redshift, with the redshift tagged at the left extreme of the curve. This figure is analogous for the other three fields, with some slight shifts due to the difference in filter sets.
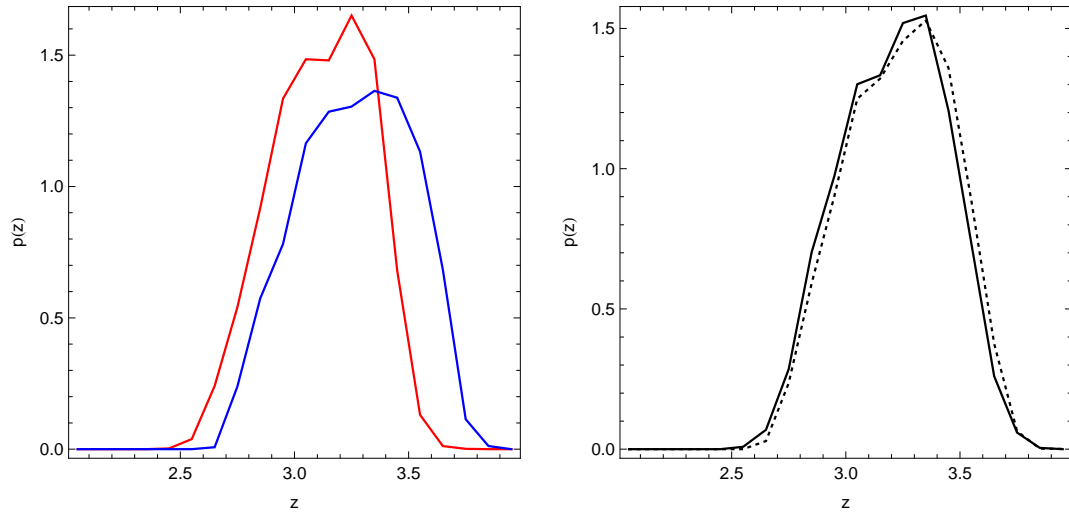
Figure 3.10 Simulated redshift distributions. *Left:* The red curve correspond to the distribution generated only from the first quartile LBG template, with Lyα in pure absorption. The blue curve corresponds to the distribution generated from the last quartile template, with the highest EW. These are extreme cases, showing the biggest possible difference expected for this set of LBG templates. *Right:* A more realistic modelling of the situation in our dataset is to include the three LBG templates with some Lyα in emission (dotted line) and compare it to the distribution obtained from the full set of templates (solid line). In this case, the distributions are practically indistinguishable.

Table 3.3.   Field volumes and LBG densities

Field	area [arcmin^2]	number of LBGs	Volume[1] [$\times 10^6$ Mpc^3]	n_{LBG}[2] [$\times 10^{-4}$ Mpc^{-3}]
ECDF-S	997.9	1388	3.22	4.3±0.1
EHDF-S	1137.0	1276	3.67	3.5±0.1
SDSS 1030	1154.4	774	3.73	2.1±0.1
CW 1255	1173.4	709	3.79	1.9±0.1
4 Fields	4462.7	4147	14.40	2.88±0.05

[1] Cosmological volume covered by the field, in comoving Mpc.

[2] Comoving number density of LBGs. Error bars come from considering poisson fluctuations in the observed number of LBGs.

where $q(z)$ is the redshift distribution, Ω is the survey solid angle and the volume element is given by:

$$\frac{dV}{d\Omega dz} = \frac{c}{H_0} \frac{D_c^2(z)}{E(z)} \tag{3.11}$$

the comoving distance $D_c(z)$ given by:

$$D_c(z) = \frac{c}{H_0} \int_0^z \frac{1}{E(s)} ds \tag{3.12}$$

and the factor $E(z)$ by eq. 1.91. Numerically evaluating this integral yields:

$$V = 3226.7 \left(\frac{\Omega}{\mathrm{arcmin}^2} \right) \mathrm{Mpc}^3 \tag{3.13}$$

formula that gives the numerical values in table 3.3.

3.4   Results

We fitted the angular autocorrelation functions between $60''$ and $1250''$, where the lower limit was set in a fairly conservative way, to ~ 2 comoving Mpc at $z \sim 3$.

This avoids the 1-halo term in the correlation function, where complications on how LBGs populate their halos begin to arise. The upper limit was set simply to avoid scales that are too large to be accurately sampled by our survey volume, and are therefore very noisy. Figure 3.11 shows the angular autocorrelation functions measured in each of the fields, and in the combined ω_{LS}. The best fit parameters in this combined correlation function are shown in figure 3.12, along with $\Delta\chi^2$ contours and the results of the jackknife realizations of the data. Although this plot shows a strong correlation between A and β, the deprojected quantities r_0 and γ show only a very weak correlation within the inner $\Delta\chi^2$ contours and in the core of the distribution of jackknife realizations, as can be seen in figure 3.13. We therefore neglect the covariance term between them when propagating errors in bias and mass. The 1-σ contours in the r_0 and γ plane and the jackknife estimates applied according to eq. 1.26 give consistent errors on these parameters, namely: $\Delta r_{0,A,\beta} = 0.55$ Mpc and $\Delta\gamma_{A,\beta} = 0.3$. We also need to consider the random errors coming from the Limber deprojection, which turn out to contribute very little to this, $\Delta r_{0,p} = 0.14$ Mpc. This gives a final estimate for the spatial correlation function of:

$$(r_0, \gamma) = (6.1 \pm 0.6 \text{Mpc}, 1.9 \pm 0.3) \tag{3.14}$$

Additionally, table 3.4 shows the best fit parameter A when β is kept fixed at 0.9 in the individual fields. This allows for a comparison of the angular clustering strength measured in each field.

Aproximating r_0 and γ with two uncorrelated gaussians given by the means and variances shown above, we estimate a bias value for our full LBG sample of 3.3 ± 0.3 at the mean of the spectroscopic redshift distribution, z=3.2. From eq. 1.95, we estimate a minimum halo mass for this halo population of $\log(M_{\min}/M_\odot) = 11.7 \pm 0.2$, which corresponds to a population with median $\log(M_{\med}/M_\odot) = 11.9 \pm 0.2$. The number density of halos having this minimum threshold mass is $n_{\rm h} = 1.2^{+1.1}_{-0.6} \times 10^{-3}$ Mpc^{-3}, and comparing this halo density with the observed LBG density quoted in table 3.3, we deduce an occupation fraction of 23^{+23}_{-11} %.

Figure 3.15 shows our LBG result in comparison to other literature clustering measurements. To connect these different populations we calculate the descendant
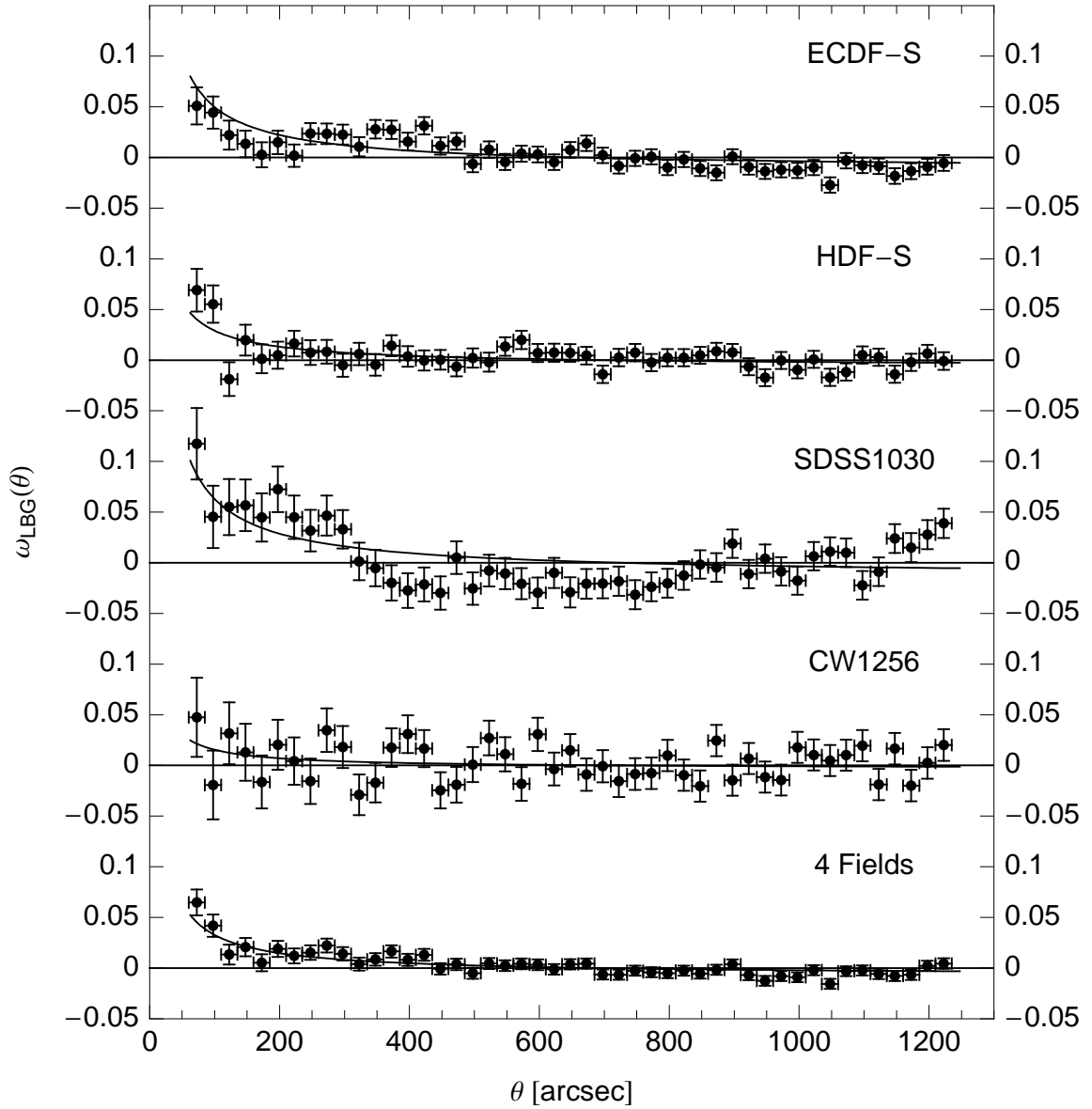
Figure 3.11 Angular autocorrelation function for the samples in each of the fields and for the combination. Solid line is the model for the best fit parameters for each correlation function.
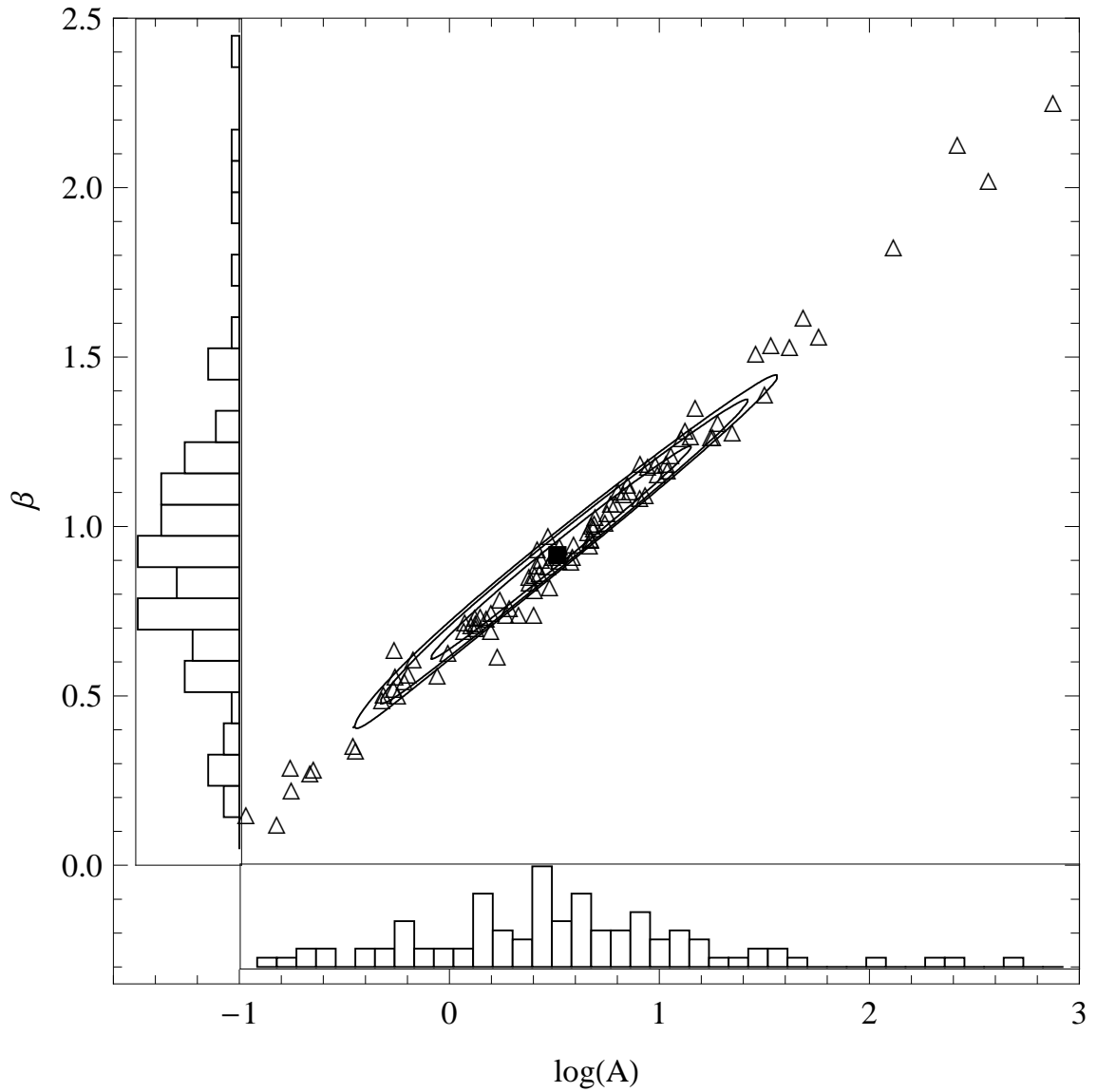
Figure 3.12 Confidence intervals for the best fit parameters of the angular correlation function A and β. The $\Delta\chi^2$ contours correspond to 1, 2 and 3 σ, while the open triangles show the jackknife realizations as computed with eq. (1.25). The best fit, $(\log A,\ \beta) = (0.52,\ 0.92)$ is plotted as a black square. Horizontal and vertical histograms show the distribution of log(A) and β values.
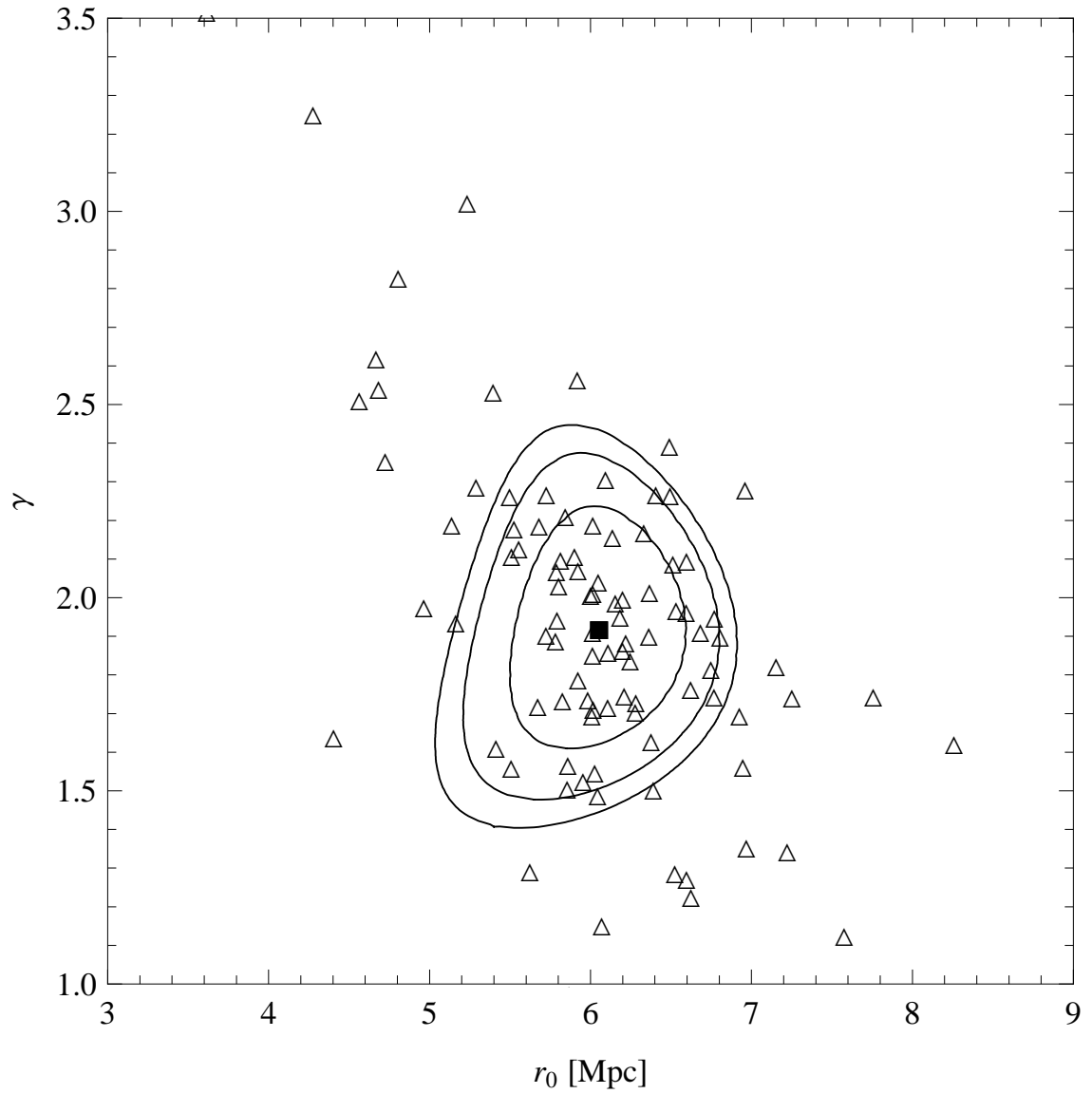
Figure 3.13 Deproyection of the $\Delta\chi^2$ contours and jackknife realizations shown in the previous plot using eq. 1.28. The best fit, $(r_0,\ \gamma) = (6.05,\ 1.92)$ is plotted as a black square. Notice that these contours do not include the uncertainties due to deprojection effects.

Table 3.4. Summary of best fit parameters for fixed slope fits

Field	A	reduced χ^2	ω_Ω
ECDF-S	3.9 ± 0.5	1.18	0.0115
EHDF-S	2.3 ± 0.6	1.10	0.0063
SDSS 1030	5.1 ± 1.0	1.15	0.0138
CW 1256	1.2 ± 1.0	0.90	0.0032
4 Fields	3.0 ± 0.3	0.99	0.0085

Note. — Fits are for a fixed slope β=0.9, which corresponds to the best fit slope in the combined 4-field data, when allowed to vary. Formal errors are 1-σ, $\Delta\chi^2$ confidence intervals for one free variable.

and ancestor conditional mass functions (CMFs) at every redshift and track the median and 1-σ limits of this distribution, as show in figure 3.14. For $z < 3.2$, the black curve corresponds to the median of the descendant CMF and for $z > 3.2$ it corresponds to the median of the ancestor CMF. The bias evolution expected for the dark matter halos hosting our sample of LBGs shows a wide variety of possible low-redshift descendants: the samples of BX galaxies of intermediate K magnitude (20.5<K<21.5) and BM galaxies targeted by Adelberger et al. (2005a) show a level of clustering consistent with being descendants of these LBG halos, as well as red galaxy samples at $z \sim 1$ from the DEEP2 survey (Coil et al., 2008). The upper 1-σ envelope in the LBG descendants black curve also suggests that the star-forming BzK selected galaxies analyzed by Blanc et al. (2008) could also be among the most massive descendants of our LBG sample. At the present time, i.e. $z = 0$, the median halo mass of the LBG descendants is $\log(M_{\mathrm{med}}) = 13.5\pm0.1$[1] and a 1-$\sigma$ dispersion of $\log(\Delta M) = 0.8$, consistent with a typical halo that hosts big elliptical galaxies or a group of galaxies. On the other hand, tracing the bias evolution, some of the high redshift galaxy populations, e.g. the LAE sample targeted by Kovač et al. (2007),

[1]Notice that the error quoted in M_{med} is purely due to random error propagation, while the dispersion corresponds to the 1-σ limits for the conditional mass distribution function, marked in fig. 3.14 with dashed lines.
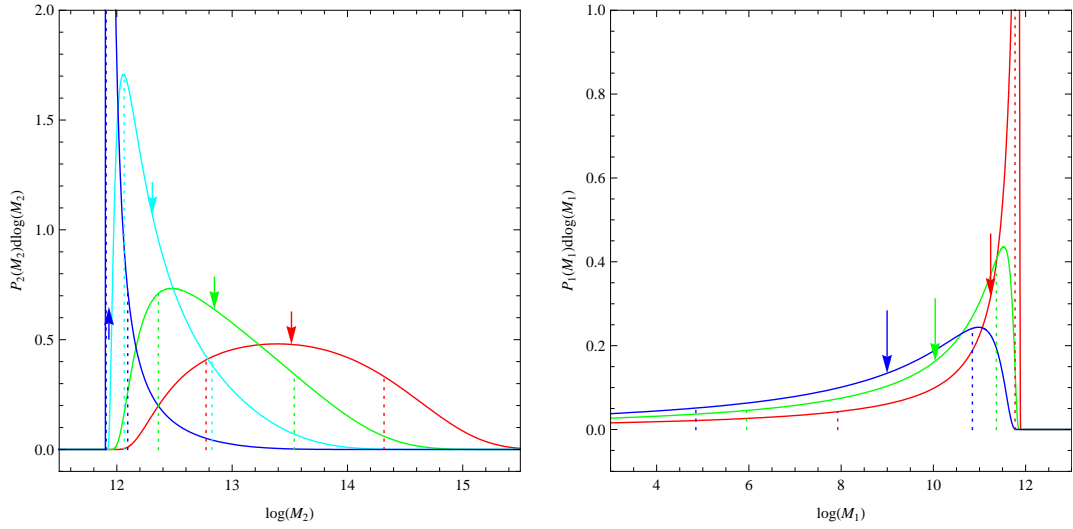
Figure 3.14 *Left:* Conditional probability distribution funtions of the median LBG halo descendants at $z = 3$ (blue), 2 (cyan), 1 (green) and 0 (red). Arrows point to the position of the median in the distributions, and dotted lines mark the position of 1-σ limits of this distribution, defined to enclose 68.3% of the probability around the median. *Right:* Conditional mass funtion of LBG ancestors at $z = 4$ (red), 5 (green) and 6 (blue).

could be among the possible ancestors of $z = 3$ LBGs. The green curve in figure 3.15 shows the bias evolution expected for these LAEs. It might seem surprising that the ancestors track for $z = 3$ LBGs in black does not coincide with the descendants track for $z = 4.5$ LAEs in green, but calculating the number densities of the halos hosting these LAEs helps in elucidating this: for a halo having a bias of ~ 3.9 at $z = 4.5$, $n_{\rm h} \sim 2 \times 10^{-3}$ Mpc^{-3}, which is similar to the number density of LBG halos. From $z = 4.5$ to 3 however, these halos also have a high probability of merging into halos more massive than the ones hosting LBGs, and less massive halos, which are also more abundant, contribute therefore significantly to the buildup of the set of LBG halos.
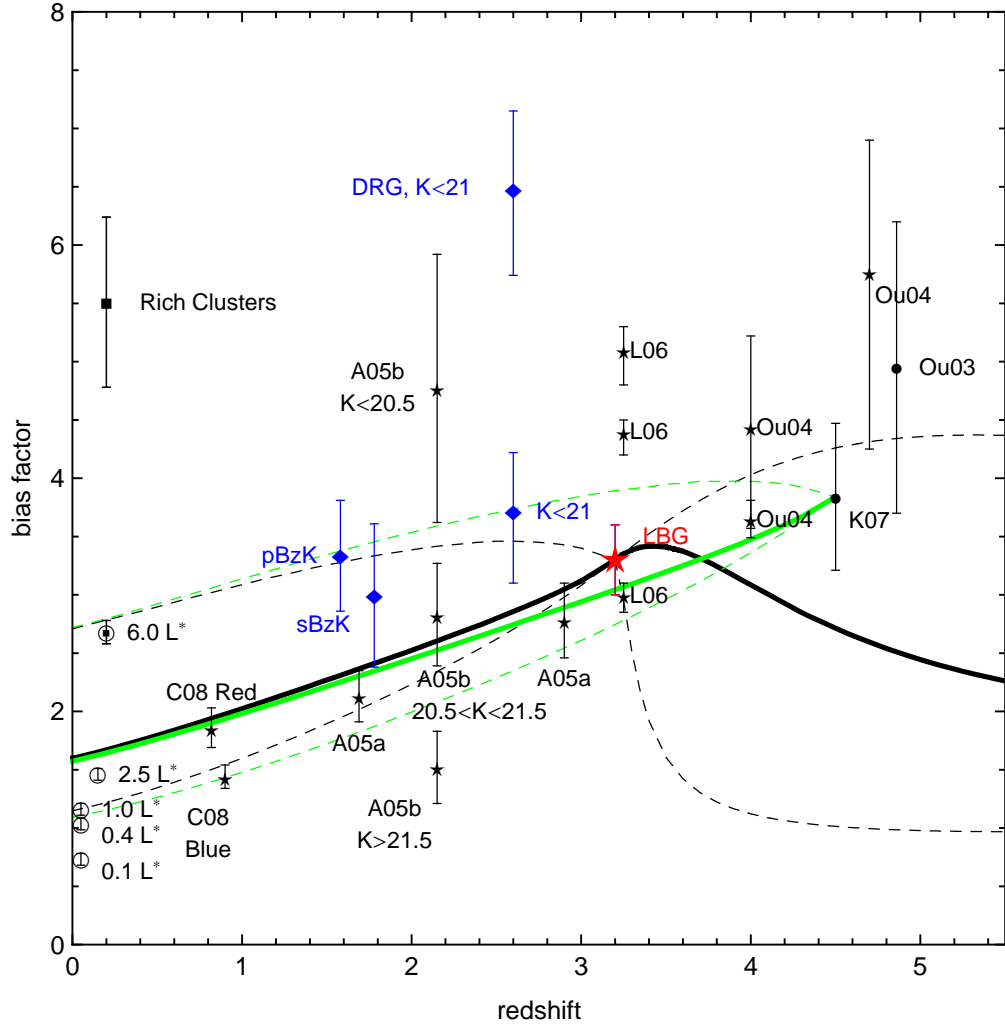
Figure 3.15 Bias evolution in redshift for our LBGs sample (red star) and literature results for a variety of galaxy populations. Stars are galaxies selected from their UV colors/magnitudes, diamonds are K selected, filled circles are narrow band selected and open circles are low-redshift samples. Labels correspond to: BM galaxies and LBG (A05a) from Adelberger et al. (2005b), BX galaxies divided by their K magnitude (A05b) from Adelberger et al. (2005a), LBGs at three limiting magnitudes (L06) from Lee et al. (2006), LBG at $z = 4$ and 5 (Ou04) from Ouchi et al. (2004), high-z LAE measurements (Ou03, K07) from Ouchi et al. (2003) and Kovač et al. (2007), for which bias evolution is also presented (green curves), $z \sim 1$ galaxies separated by color (C08) from Coil et al. (2008). Low redshift galaxies, labeled by their optical luminosity, come from Zehavi et al. (2005) and the single point for rich galaxy clusters from Bahcall et al. (2003). Other results derived from MUSYC data are in blue, namely K-selected galaxies and DRGs at $z = 2.6$ from Quadri et al. (2007), and both passive and star-forming, BzK-selected galaxies from Blanc et al. (2008).

3.5 Discussion

In this study we measured the spatial correlation function of LBGs selected as *U*-dropouts with an apparent magnitude $R < 25.5$, in the four fields of MUSYC, and found that they cluster as strongly as expected for dark matter halos with a typical mass of $\sim 8 \times 10^{11}$ M$_\odot$ at redshift 3.2. This is in good agreement with literature results at $z \sim 3$ from Adelberger et al. (2005b), Hildebrandt et al. (2007) and Lee et al. (2006) for samples with similar brightness. Our inferred occupation fraction of 10-40% is consistent within the errors with the result from Adelberger et al. (2005b), who obtained 20-60%. This last author also suggested, based on comparison with dark matter halos in the ΛCDM-GIF numerical simulation, that there would be an evolutionary sequence between their 3 galaxy samples, from LBG ($z = 2.9$) to BX ($z = 2.2$) and to BM ($z = 1.7$) galaxies, and then to absorption line galaxy samples from the DEEP survey, at $z = 1$. They also argue, based on their expectation for the clustering strength, that the descendants from their LBG population at present time should be found among elliptical galaxies. From our clustering measurement and the analysis based on the extended Press-Schechter formalism, we also find agreement with Adelberger's claim using our data. Additionally, we find some of the high-redshift LBG and LAE samples (Ouchi et al., 2004; Kovač et al., 2007) to be good candidates for the ancestors of our LBGs at $z = 3$, and that a significant portion of galaxies selected with the BzK technique (Blanc et al., 2008, and references therein) could be among the high-mass end of the descendants of the dark matter halos hosting our LBG sample.

Chapter 4

The clustering of Active Galactic Nuclei[1]

4.1 Introduction

There is a wealth of evidence that nuclear supermassive black holes play a significant role in the process of galaxy formation and evolution. This has become evident in the past few years with the discovery of correlations between the properties of the massive black holes, the stellar systems that host them, and their dark matter halos (Magorrian et al., 1998; Gebhardt et al., 2000; Ferrarese, 2002).

Luminous quasars have been studied in great detail, with systematic spectroscopic studies by the Sloan Digital Sky Survey and the 2dF Galaxy Redshift Survey. These objects are rare and represent the bright end of the Active Galactic Nuclei (AGN) luminosity function. In order to fully understand the link between the growth of supermassive black holes (SMBH) and galaxy evolution, we need to study young, high-redshift galaxies hosting AGN with more typical luminosities.

---

[1]A slightly shortened version of this chapter was published on Astrophysical Journal Letters, volume 673, page L13.

One of the basic properties of galaxy populations is their clustering strength, but there are few constraints on this quantity for fainter, high-redshift AGN. Furthermore, there are important disagreements in the literature: e.g. galaxy-AGN cross-correlation measurements at $z \sim 3$ by Adelberger & Steidel (2005b) (A05c), imply a bias factor of 3.9 ± 3.0 for luminous AGN (UV luminosities between $-30 < M_{1350} < -25$), while recent work by Shen et al. (2007) (S07) suggests a value of 9.1 ± 0.9 for sources with similar luminosities at the same redshift. For galaxies, e.g., measurements indicate a strong luminosity dependence in the clustering length (Giavalisco & Dickinson, 2001). On the other hand, for AGN there are a handful of claims that this is not the case (Croom et al., 2005; Myers et al., 2006; Adelberger & Steidel, 2005a). Determining the bias with accuracy puts important constraints on models of AGN formation and evolution, (e.g., Lidz et al., 2006; Hopkins et al., 2007) but requires spanning a broad range in luminosity and obscuration level. This is challenging using optical color selection plus spectroscopy, because the low fraction of AGN (3%) found among these candidates (Steidel et al., 2002) demands huge amounts of spectroscopy time. Using X-ray detection plus spectroscopy (e.g., Szokoly et al., 2004) provides a more efficient and unbiased selection of AGN, since their surface density is higher in X-rays than in optical images and obscuration effects are much less important. However, the unrestricted redshift range sampled by this method makes spectroscopic follow-up highly inefficient, and often prevents confirmation of dimmer, high-z AGN (R>24). Additionally, targeting a very broad range in redshift yields a low signal-to-noise measurement of the angular correlation function due to dilution, and adds the evolution term to the clustering parameters that need to be constrained. Hence it is difficult to obtain AGN samples suitable for clustering studies. There have been several measurements of the spatial correlation function of X-ray selected AGN at $z < 1$ (Mullis et al., 2004; Gilli et al., 2005; Basilakos et al., 2004; Basilakos et al., 2005; Yang et al., 2006; Miyaji et al., 2007), with samples ranging from 200-500 sources. At higher redshifts, the statistics are much poorer, and consist of purely optically selected AGN. In this work, we constrain the clustering strength of an AGN sample at $z \sim 3$ jointly selected by optical and X-ray photometry. We determine whether these sources cluster more or less than non-active galaxies at this redshift, and discuss their present-day descendants.

4.2 AGN and LBG samples

The full set of unique X-ray counterparts is taken from the MUSYC ECDF-S X-ray catalog (Cardamone et al., 2008), constructed joining the catalogs from Giacconi et al. (2002), Alexander et al. (2003), Virani et al. (2006) and Lehmer et al. (2005), using a likelihood procedure (Brusa et al., 2005) to match sources between the X-ray and optical (BVR-selected) catalogs.

Figure 4.1 shows the $U - V$ versus $V - R$ color-color plot for sources in the ECDF-S from MUSYC photometry. Lyman break color selection ('UVR') corresponds to the region outlined in the upper left side (Steidel et al., 1996b; Gawiser et al., 2006a). For this color selection, U-band fluxes are required to be detected at 1-σ, and are otherwise set to an upper limit equal to their 1-σ error. This avoids interlopers (typically dwarf stars) with uncertain photometry at the cost of incompleteness in the sample. Additionally, $R < 25.5$ is required to allow for spectroscopic confirmation. Sources presenting a drop in the U filter due to intergalactic absorption, along with a blue continuum between the V and R filters that rules out heavily reddened lower-z objects, are expected to lie at $2.7 < z < 3.7$. For XUVR selection of AGN we require an X-ray detection with counterpart UVR colors in the same region and we drop the R magnitude limit and the 1-σ U-band requirement, since the extra requirement of X-ray emission already rules out most dwarf stars. The upper limit in U flux set by the 1-σ error is a conservative approach, so for AGN selection we replaced dropout magnitudes with $U = 30$. This number is arbitrary, but it exact value is rather immaterial, as long as is set to anything above ~ 28, because it just moves sources up or down in the upper part of the UVR diagram. This procedure yields 1385 LBG and 59 AGN candidates. Unobscured AGN and LBG have somewhat similar UVR colors, as can be seen in the color-color track in figure 4.1, since these colors are mainly determined by the intergalactic hydrogen absorption. Furthermore, obscured AGN are the dominant population among AGN, and since the SED of the first is dominated by their host galaxies (Treister et al., 2004, Treister & Urry, 2005), we expect the redshift distribution of our AGN sample to differ mildly from the LBG distribution. Notice that in the case of significant obscuration, we don't expect to select AGN hosted by very red galaxies, since they

Table 4.1. Spectroscopic XUVR targets in ECDF-S

Sample	R mag range	Candidates	Observed	Identified	$z \sim 3$ AGN	Interlopers
bright xuvr	R<25.5	39	31	13	10	3
faint xuvr	R>25.5	20	11	1	1	0

will probably not be bright enough in the restframe UV. Figure 4.2 shows the position of these sources in the ECDF-S field. This kind of selection has been previously applied to target AGN at this redshift (Nandra et al., 2005), and even estimates for the luminosity function of these sources have been made (Aird et al., 2008).

4.3   Spectroscopic confirmations

From the spectra obtained over the entire 4-field survey so far, 131 LBG and 30 AGN were identified at the target redshift ∼3, and in the ECDF-S in particular, 31 LBG and 11 AGN have been confirmed. Table 4.1 lists the number of spectroscopically targeted and confirmed AGN in ECDF-S, separated by R magnitude. LBGs are listed in table 3.2 of the previous chapter. While AGN in the ECDF-S were directly targeted using a joint X-ray and optical selection, in the rest of the MUSYC fields AGN have been discovered serendipitously among the UVR candidates. Figure 4.3 shows the redshift histograms of the confirmed objects, which implies X-ray luminosities between 10^{43} and 10^{45} ergs/s for the entire set of AGN candidates (see figure 4.4). The Lyman-Break (UVR) selection shows a fraction of low-redshift interloper less than 10%. In the spectra of the XUVR-targeted AGN in ECDF-S, we did not observe any interlopers. To compare our results to clustering measurements from optically selected AGN, we also calculate the UV luminosity of our samples from the R magnitude, see figure 4.5. The list of spectroscopically confirmed AGN can be found in appendix A.
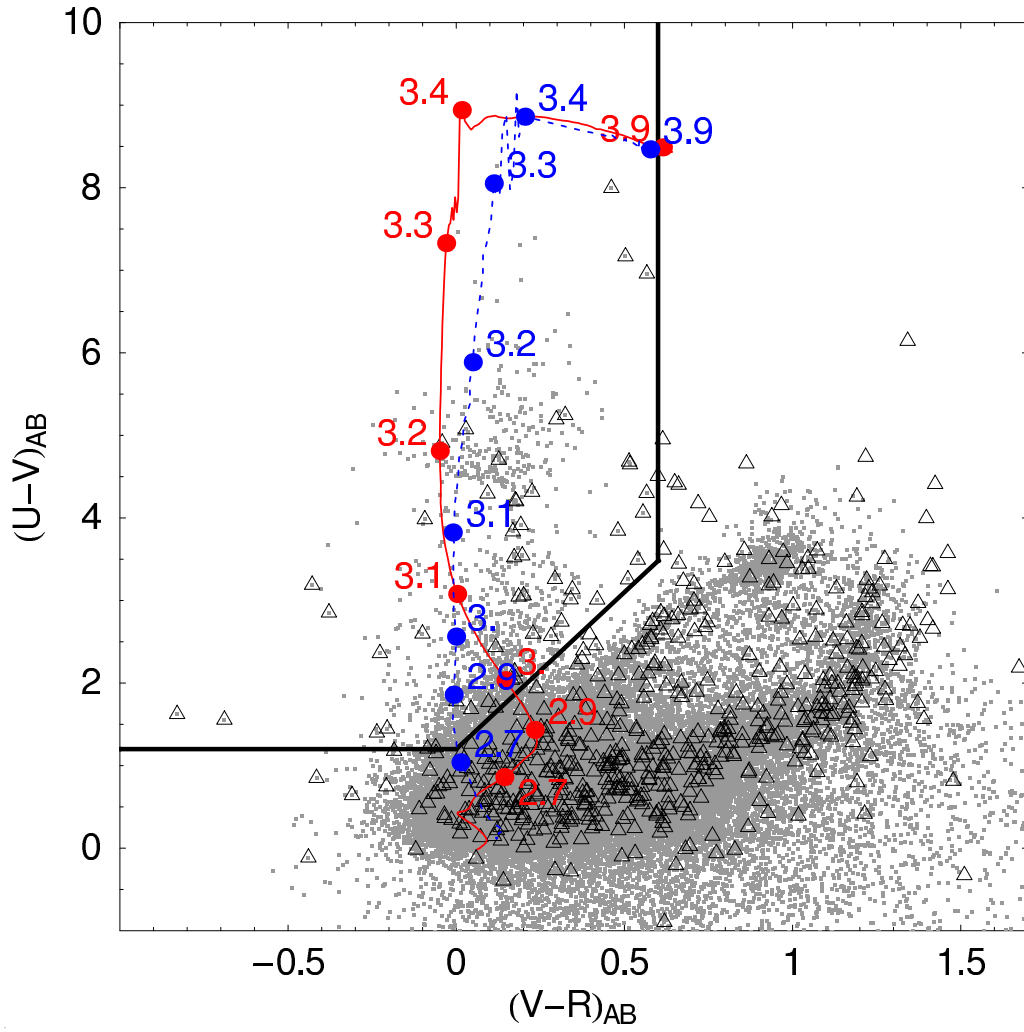
Figure 4.1 $U - V$ versus $V - R$ color-color plot for all the sources brighter than $R =$ 25.5 in the MUSYC ECDF-S optical catalog (gray points) and X-ray counterparts (open triangles). LBG and $z \sim 3$ AGN selection are inside the region marked in solid lines. The color-color tracks of a template AGN and a LBG as they are redshifted from z=2.0 to 3.9 are shown in the red solid and blue dotted curves, respectively.
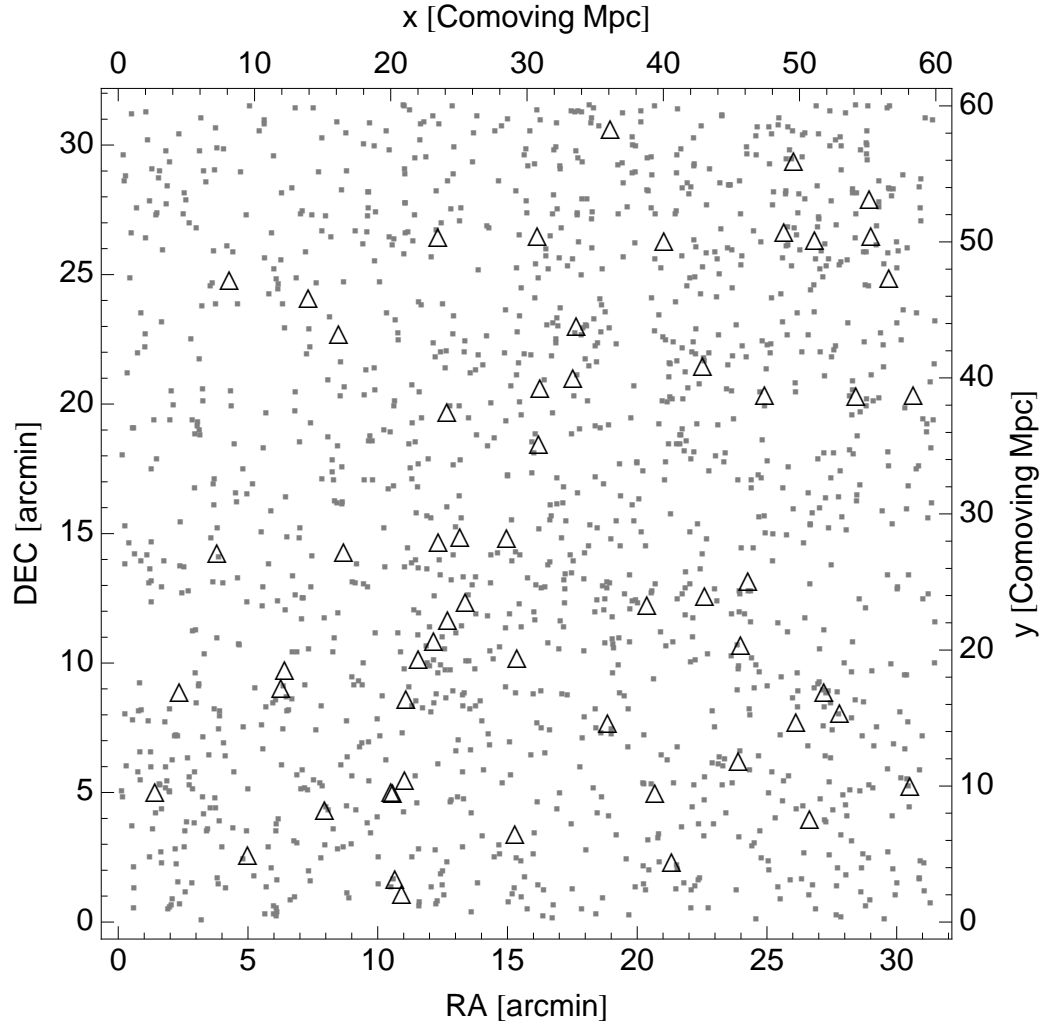
Figure 4.2 R.A. and Dec. plot of the selected AGN (open black triangles) and LBGs (gray dots) at $z \sim 3$ in ECDF-S. The X and Y axes show the projection in comoving Mpc to the mean redshift of both samples, $z = 3.2$.
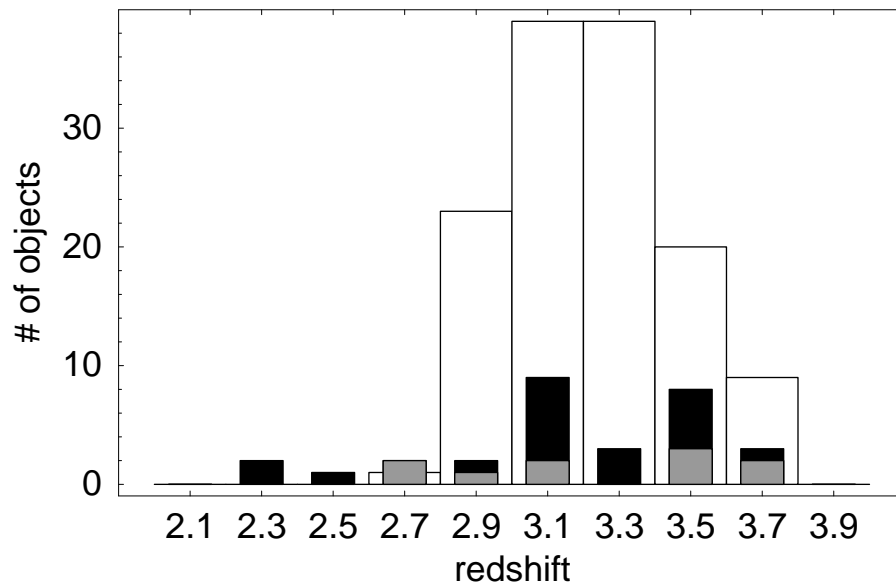
Figure 4.3 Redshift distribution of all spectroscopically confirmed LBGs (white bars) and $z \sim 3$ AGN (black bars) in MUSYC. The subset of AGN in ECDF-S is in gray bars. These distributions have mean and standard deviations: $\bar{z}_{AGN} = 3.18$, $\sigma_{z,AGN} = 0.39$, $\bar{z}_{LBG} = 3.23$, $\sigma_{z,LBG} = 0.22$.
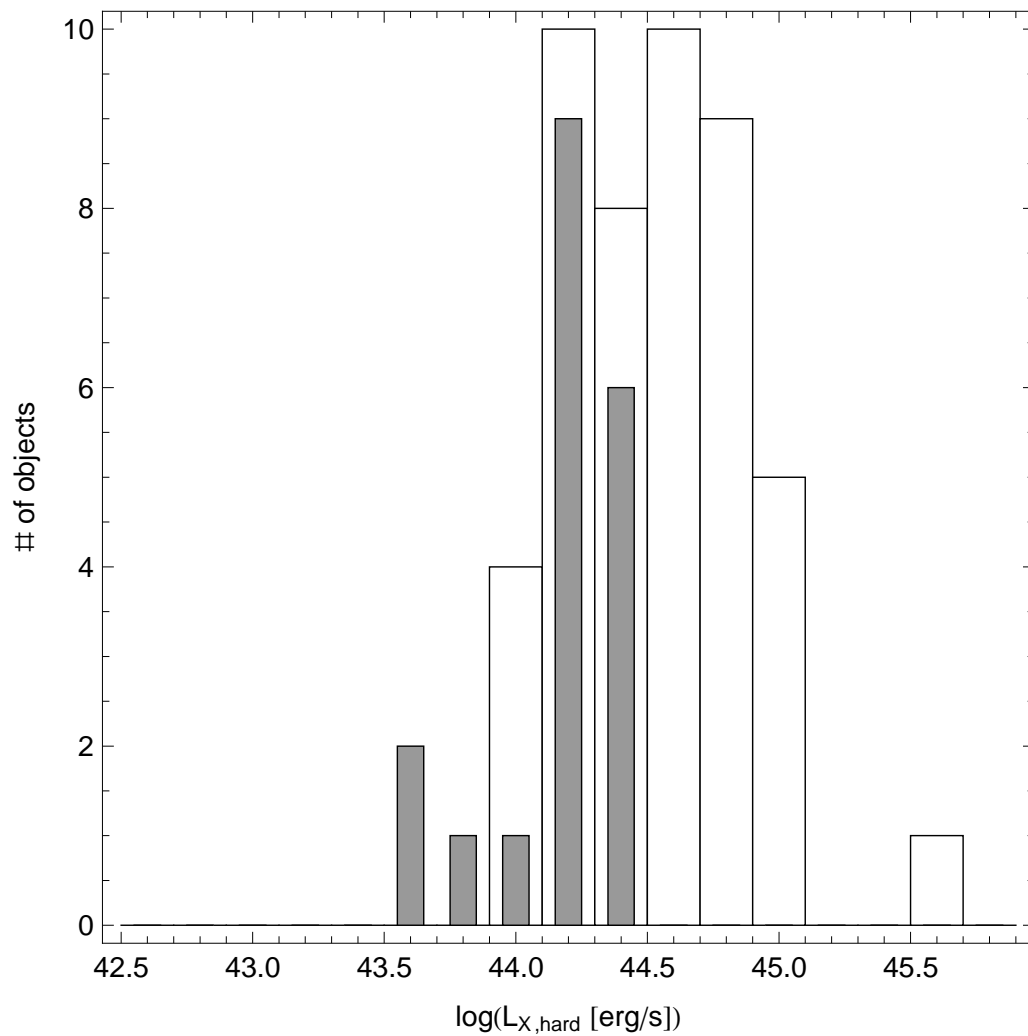
Figure 4.4 X-ray luminosity distribution of all XUVR AGN candidates. White bars come from values reported as flux measurements, and gray bars come from flux upper limits in the combined X-ray catalog. Luminosities were estimated assuming all sources were located at the mean redshift 3.2. The dispersion around this value due to the span in redshift mentioned in the previous plot is of 0.14 dex.
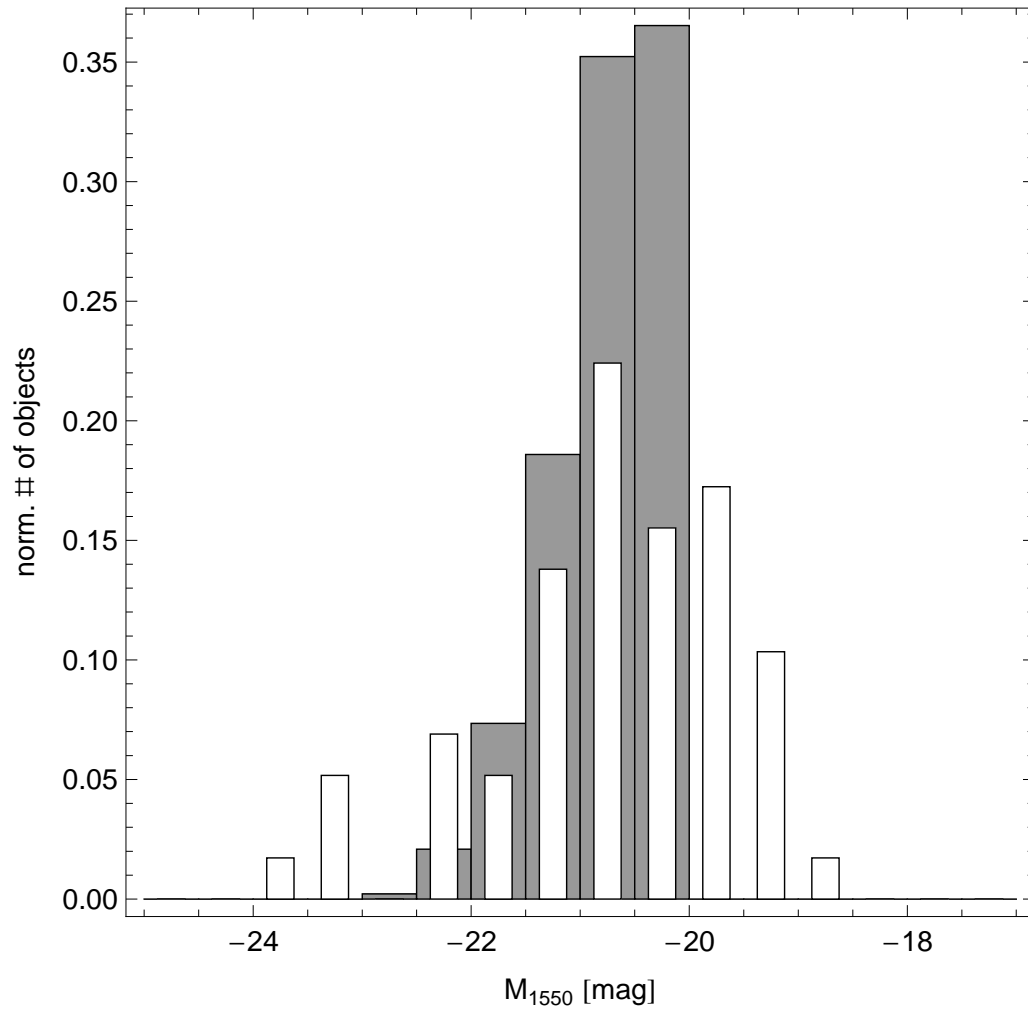
Figure 4.5 UV magnitude histogram of XUVR sources. Gray bars are LBG candidates, white bars are AGN candidates. Just like in the previous plot, all sources have been assumed to be at the median reshift 3.2.

4.4   Cross-correlation Analysis

We calculate the clustering of AGN via a two-step method, calculating the autocorrelation of LBG first, and then their cross-correlation with AGN. The main advantage of this approach is that it improves the statistics significantly, since the LBG population is much more numerous and its redshift distribution is similar to the AGN population sampled here, hence the AGN-LBG cross-correlation function is less noisy than the AGN autocorrelation function (Kauffmann & Haehnelt, 2002). Notice that the AGN and LBG that contribute to the cross-correlation function will be those that spatially overlap. Therefore, our results will reflect the clustering of AGN in the redshift range $2.7 < z < 3.8$.

For the UVR selected sources, we calculated the angular autocorrelation function over the ECDF-S field in the same fashion as in the previous chapter. To estimate the angular cross-correlation between UVR-XUVR sources, we used the cross-correlation estimator given by equation 1.11 (Croft et al., 1999):

$$\widehat{\omega}_{AGN-LBG}(\theta) = \frac{D_{AGN}D_{LBG}(\theta)}{D_{AGN}R_{LBG}(\theta)} - 1 \qquad (4.1)$$

where the data-data pairs are between AGN and LBG, and data-random between AGN and the optical random catalog. Notice that this estimator does not use a random catalog for the X-ray sources. The degradation of the PSF and sensitivity as a function of the off-axis angle in the Chandra data makes the X-ray survey geometry fairly complicated, but these resulting systematic uncertainties can be avoided in this cross-correlation scheme.

The estimates shown above were fitted to $\widehat{\omega}(\theta) = A\theta^{-\beta} - \mathrm{IC}(A, \beta)$, where the integral constraint factor IC is included as part of the model and we assumed the usual power law for the true angular correlation function. We kept $\beta = 0.8$ fixed, since our S/N does not allow us to put significant constraints on the slope. Fit limits were set to avoid the one-halo term regime on small angular scales. As a relatively conservative upper limit, we took 3 times the angular size of our expectation for their DM halo virial radii. Assuming 6×10^{11} and $10^{13} M_\odot$ for the LBGs and AGN halos

(with corresponding virial radii at z=3.2 of 60 and 160 physical kpc.), we estimated $30''$ and $60''$ for the lower fit limits, respectively. At large scales, we set the upper fit limit at half the field size, in order to avoid border effects of the estimator and where the sampling becomes poor anyway.

4.5 Results

The fit to the LBG autocorrelation function gave $A_{LBG} = 2.2 \pm 0.3$ arcsec$^\beta$ and the fit to the cross-correlation function $A_{AGN-LBG} = 2.9 \pm 1.1$ arcsec$^\beta$, with reduced χ^2 values of 1.24 and 1.21, respectively. The errors in these parameters were calculated using $\Delta\chi^2$ and correspond to 1-σ confidence level with 1 parameter. Figure 4.6 shows the measured angular correlation functions with their corresponding best fit models. Binning $\omega(\theta)/\sigma_\omega(\theta)$ over the entire fitting range gives estimates for the total signal-to-noise ratios of the auto- and cross-correlation functions measurements of 10.2 and 3.9, respectively. Additionally, to determine the probability of obtaining these values for the angular correlation functions parameters merely by chance, we generated correlation functions with zero mean and the same standard deviation as in the measured $\omega_{AGN-LBG}(\theta)$. We then fitted these realizations in the same way as we did for the measured correlation functions. From this distribution, we estimated that the probability of obtaining clustering this high from two unclustered populations of the same sizes as the LBG and AGN samples are <0.1% for the LBG autocorrelation and 1% for the AGN-LBG cross-correlation function.

Using smooth fits to the redshift distributions of the confirmed sources shown in figure 4.3, we deprojected the values obtained for the angular LBG autocorrelation and LBG-AGN cross-correlation amplitude using Limber's equation. We obtained spatial correlation lengths of $r_{0,LBG} = 6.7 \pm 0.5$ Mpc and $r_{0,AGN-LBG} = 8.7 \pm 1.9$ Mpc respectively, where the correlation length is defined by the usual fitting form $\xi(r) = (r/r_0)^{-\gamma}$ for the spatial correlation function. The random errors introduced in the deprojection are included as described in chapter 1.

To calculate the bias factor of the AGN sample in the approximation of linear,
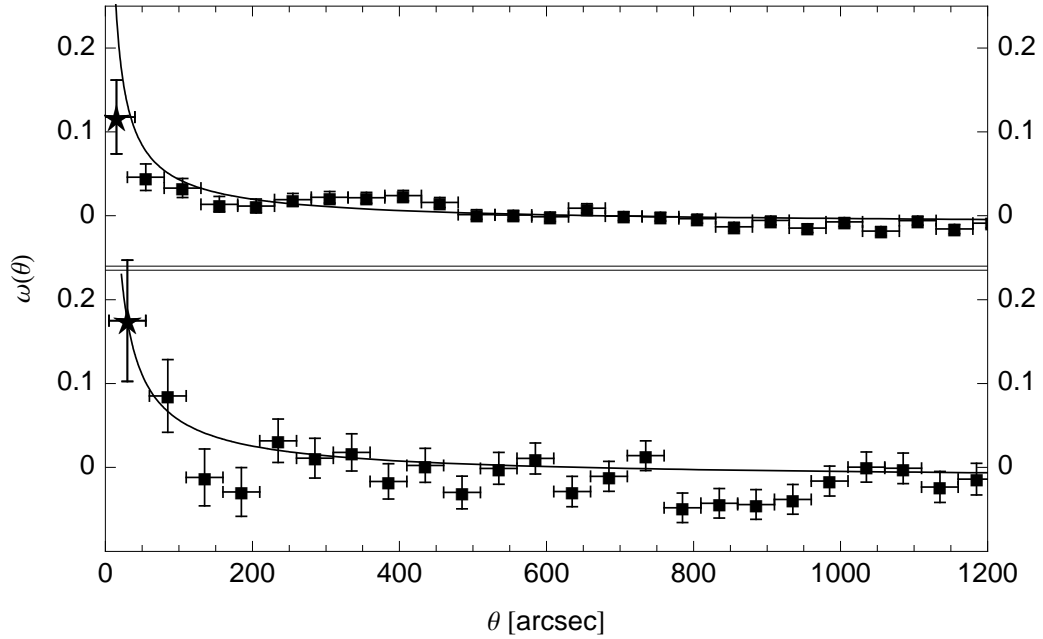
Figure 4.6 Top: Angular autocorrelation function for UVR selected sources, i.e. Lyman-break galaxies at $z \sim 3$. The solid line is the best fit to the data, $\omega_{lbg}(\theta) = 2.2 \pm 0.3 \ \theta^{-0.8}$. The point marked with a star is not included in the fit and is shown only for reference. Bottom: Analogous plot for the angular cross-correlation function with XUVR sources, i.e., AGN at $z \sim 3$. The best fit corresponds to $\omega_{agn-lbg}(\theta) = 2.9 \pm 1.1 \ \theta^{-0.8}$.

Table 4.2.   Summary of DM halo properties

	bias	$\log(M_{\min})$	n_{h} (Mpc^{-3})
LBG	3.5 ± 0.3	11.8 ± 0.15	$9 \pm 4 \times 10^{-4}$
AGN	5.5 ± 2	$12.6^{+0.5}_{-0.8}$	10^{-3} - 10^{-6}

Note. — Confidence intervals/ranges are 1-σ. Masses are given in logarithm of solar masses.

constant bias, we used the elementary relations from chapter 1: $b^2_{lbg} = \sigma^2_{8,lbg}/\sigma^2_{8,dm}$ and $b_{agn}b_{lbg} = \sigma^2_{8,agn-lbg}/\sigma^2_{8,dm}$, where $\sigma^2_{8,X}$ is the variance in spheres of radius $8h^{-1}_{100}$ Mpc in the random field described by the corresponding auto- or cross-correlation function. Although A/β and r_0/γ are highly degenerate parameter pairs, the bias factors are robust to variations in the slope β (or γ). Therefore, they are our preferred quantity for comparison with the literature. The bias, number densities and masses of the dark matter halos for our UVR and XUVR samples are shown in table 4.2.

4.6   Discussion

We have measured the autocorrelation strength of z=3 LBG and the cross-correlation between these galaxies and AGN selected using both X-ray and optical data. For the LBG sample, we obtained a bias factor of 3.5±0.3. This is somewhat higher than the value 2.8±0.3 found by Adelberger et al. (2005b) at $z = 3$. Hildebrandt et al. (2007) obtained a bias value of 3.2±0.2 for an equivalent LBG population, consistent with ours.

From the AGN-LBG cross-correlation and the bias calculated for the LBG, we have deduced a bias factor of 5.5±2 for our AGN sample. The active galaxies targeted in this study appear to cluster more than star-forming galaxies with similar

restframe-UV colors. This is consistent with cosmic downsizing of AGN, implying that typical SMBH tend to sit in more massive galaxies than the "normal" galaxy population. We need greater statistics to confirm this result, since the values are consistent within the uncertainties.

A05c performed this same calculation in an optically selected and spectroscopically confirmed sample of 79 AGN between $1.6 < z < 3.7$, dividing the AGN by UV luminosity into bright (25 sources with $-30 < M_{1350} < -25$) and dim (54 sources with $-25 < M_{1350} < -19$) samples. Although A05c does not report the autocorrelation length of the galaxy sample used to compute these values, we approximate it by averaging the result for LBG and BX galaxies presented in Adelberger et al. (2005b), obtaining a bias factor of 2.6±0.3. From this we infer bias factors of 3.9±3.0 and 4.7±1.7 for the A05c bright and dim AGN samples. Our AGN sample has an UV magnitude range between -26 and -20, almost identical to the A05c faint AGN set, and showing the same clustering strength. In figure 4.7 these results are compared to the present estimate and to that obtained by S07 for $\sim$2250 SDSS quasars at $2.9 < z < 3.5$. As can be seen in the plot, our current AGN sample does not include sources in the luminosity range of the slight disagreement between A05c and S07, at the bright end. Although this disagreement is only 1.6σ, resolving it in the future will give important constraints on AGN lifetimes (see e.g., Adelberger & Steidel, 2005a). Removing the bright magnitude cut we imposed on the Lyman-Break selection (eq. 3.4) and preferentially targeting bright UVR sources in the spectroscopic follow-up could include AGNs in our sample comparable to A05c bright sample, but to gather a sample big enough to have statistics similar to S07, much wider areas would need to be covered.

To estimate the dark matter halo mass of the typical descendant of the halos that host these $z \sim 3$ galaxy sets at the present time, we followed the same procedure as in the previous chapter using the descendants and ancestors conditional probability distribution at lower redshifts. Just as in the previous chapter, figure 4.8 shows our AGN results in comparison with literature values for other galaxy populations. Our AGN sample, with a bias of 5.5 at $z = 3.2$ would have a typical halo with a bias factor of $2.3^{+0.6}_{-0.3}$ at present, and median masses of $\log(M/M_\odot) = 14.1^{+0.3}_{-0.2}$. This is consistent with the clustering of the brightest galaxy sample from Zehavi et al.

(2005), having $4 - 10L^*$, which at the present time are most likely located in groups and galaxy clusters. The error bar in this measurement is significant, however, and a broad range of descendants are allowed for these AGN. It is not possible to say from this result alone, whether the SMBH is going to end up necessarily in the central galaxy of the halo. Some fraction of these SMBH might end up in satellite galaxies of the halo, of lower luminosity.

This confirms the result found by A05c for faint, optically selected AGN. Improving our understanding of the AGN-galaxy connection at $z \sim 3$ and constraining models such as Lidz et al. (2006) and Hopkins et al. (2007) for AGN clustering, requires resolving the discrepancy in the bias estimates for bright AGN seen in figure 4.7 improving the statistics. We have show that targeting high-redshift AGN for clustering analyses can be done very efficiently by means of deep optical and X-ray imaging, and for that reason, future surveys with deep X-ray coverage will be ideal for obtaining large samples of active galaxies in restricted redshift ranges, suitable for clustering studies. Our method can be applied to these surveys to obtain enough X-ray-selected AGN to reduce the current uncertainties.
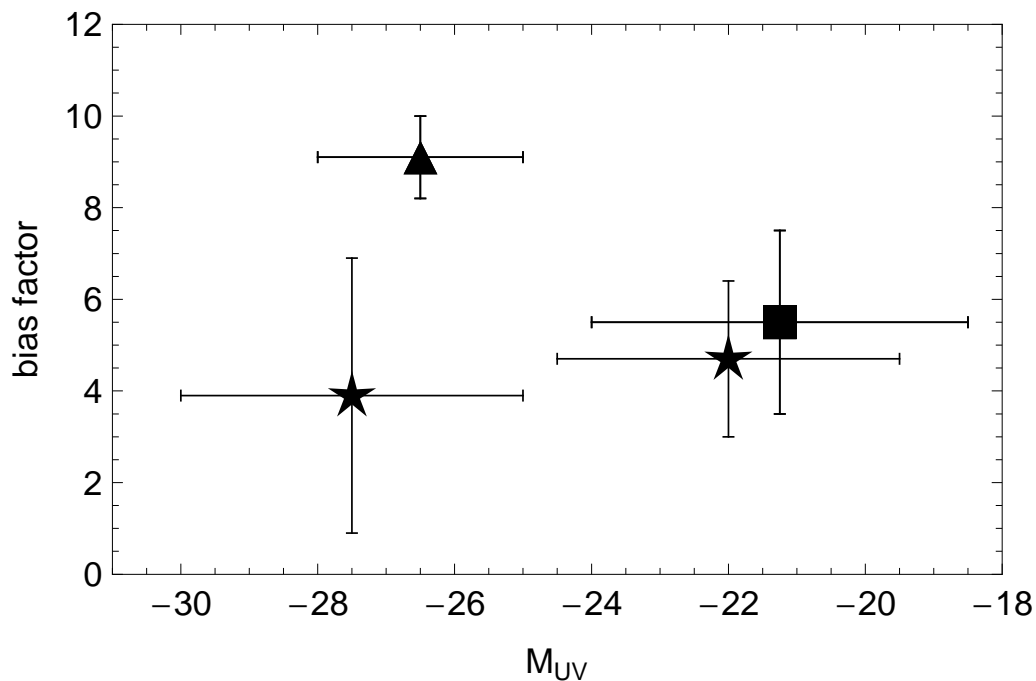
Figure 4.7 Comparison of AGN bias factors at $z \sim 3$ between this work (square), S07 (triangle) and A05c (stars). The UV luminosities are determined from different data sources, so the x-coordinate of these datapoints only approximate: Adelberger & Steidel's points have their continnum flux measured from spectra, near 1350 Å; Shen et al.'s point is estimated from the SDSS r magnitude and our result comes from the MUSYC R magnitude.
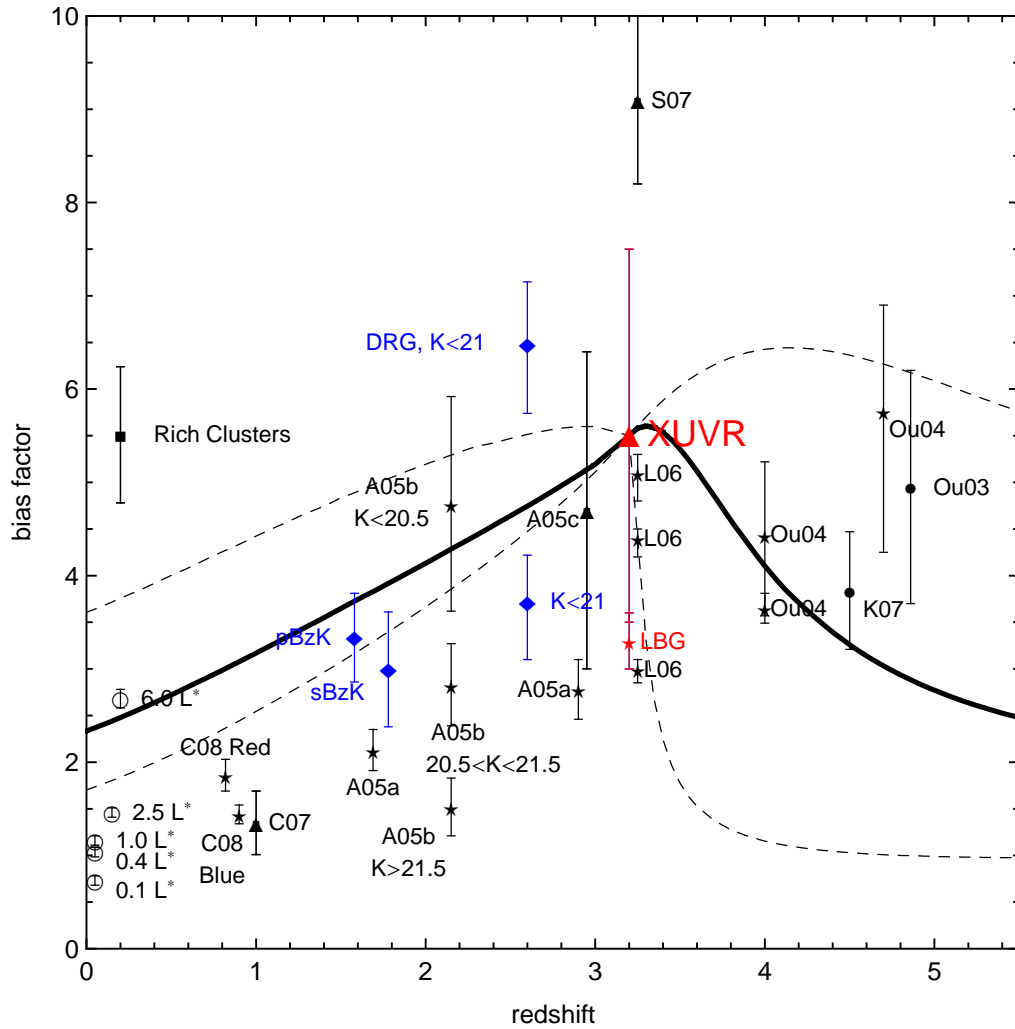
Figure 4.8 Bias evolution in redshift for our XUVR AGN sample (red triangle) and literature results for a variety of galaxy populations, with the same labels as in figure 3.15. Results for some AGN biases have been added with triangles: to the results at $z \sim 3$ from Shen et al. (2007) (S07) and Adelberger & Steidel (2005b) (A05c) cited in the text, we added the low redshift results from Coil et al. (2007) (C07).

Chapter 5

The Clustering of Lyman Alpha Emitters

5.1   Introduction

The idea of selecting primeval galaxies by searching for their Lyman-Alpha emission line is not new: Partridge & Peebles (1967) suggested that young galaxies might appear very bright at the redshifted wavelength of this transition, and that they could be easily detected at $1 + z < 6$. Although the first searches for high-redshift galaxies, started in the 70s and continued during the 80s and 90s, were focused on finding the redshifted Lyα emission line, this method only yielded positive results after the more rapid success of the Lyman-Break technique, which is based on continuum colors. This initial failure in narrow-band searches of the Lyα line was attributed to interstellar extinction. Since the Lyα is a resonant transition, interstellar hydrogen repeatedly scatters Lyα photons and increases their optical path, and thus, even a small amount of dust can significantly reduce this light. However, increasing the depth of this early surveys by a factor a few proved to be sufficient to make this technique feasible. Just as in the case of LBGs, successful searches started around quasars (Hu & McMahon, 1996; Steidel et al., 2000; Infante et al., 2003), but fairly

quickly blank field surveys proved fruitful too (Cowie & Hu, 1998; Hu et al., 1998; Kudritzki et al., 2000, among others). Although less efficient than the Lyman-Break technique, this method has become very important since it allows to probe the galaxy luminosity function much deeper.

Obtaining samples of these galaxies large enough to allow clustering studies is only now becoming a reality and currently, results for Lyman Alpha Emitters (LAEs) are relatively scarce and prone to systematics. Ouchi et al. calculated the spatial correlation function of 87 sources at $z = 4.86$ in an area of 543 arcmin^2 of the Subaru Deep Field and found a fairly large correlation length, even larger than LBG at this redshift. They also reported, however, significant contamination in their sample (40%), and the systematic uncertainty introduced by any attempt to correct for this effect is significant. Shimasaku et al. (2004) reported, in a follow-up work on the same field, that in a neighboring sample of 51 LAEs at $z = 4.79$ they found very weak clustering, and concluded that cosmic variance was a significant source of error in these surveys. Hayashino et al. (2004), imaging a 1088 arcmin^2 area around the SSA22a field previously studied by Steidel et al. (2000), found a significantly clustered population of 283 LAEs, something that is not a surprise, given the presence of a protocluster at $z = 3.09$ in the SSA22a subfield. They also detected 49 Lyα absorbers from their *decrement* in the narrow band flux with respect to the broadband flux. More recently, Kovač et al. (2007) were able to measure the spatial correlation function of a sample of 151 LAEs at $z = 4.5$ in the Large Area Lyman Alpha survey (LALA), covering an area of 1296 arcmin^2. This result also shows LAEs with significant clustering stregth, consistent with halos similar to LBG halos. They also suffered, however, from significant contamination (28%), and thus have the large systematic uncertainty coming from the correction indicated by eq. 1.33.

Measuring clustering properties requires the assembly of an homogeneous, spatially compact and numerous sample of objects; but also of a sample that shows a small level of contamination. The MUSYC coverage on ECDF-S, although of an area comparable to other surveys, provides very deep narrow-band and broad-band imaging, and an extensive spectroscopic database. This follow-up program has shown that at z=3.1, the typical interloper – [OII] emitter galaxies at $z = 0.34$ – have

Lyα equivalent widths smaller than 80 Å. Thus, setting and EW threshold let us eliminate almost all interlopers, something that at higher redshifts is less reliable, because interlopers show higher EWs. This enables us to select LAE down to fairly low equivalent widths, while at the same time keeping the contamination fraction to very low levels.

5.2 Narrow-Band object selection

The Lyman Alpha Emitter (LAE) sample used in this study comes Gronwall et al. (2007), constructed with the MUSYC narrow-band observations of the ECDF-S field. The selection consisted of a color cut in *NB5000 - BV*, where the *BV* magnitude was calculated from the mean flux between the *B* and *V* filters. Figure 5.1 ilustrates this idea: at $z \sim 3.1$, the Lyα emission of a galaxy falls inside the narrow band filter, and the galaxy becomes much more prominent in this band. Gronwall et al. (2007) showed that a color cut of *NB5000 - BV* < -1.03 corresponds to an observed frame equivalent width cut of 80 Å. Additionally, they applied an apparent magnitude cut of *NB5000* < 25.4, corresponding to a flux limit of 1.5×10^{-17} ergs cm^2 s^{-1}, where they showed the sample reaches a completeness of 90%. Figure 5.2 shows the color-magnitude plot for the ECDF-S, with the color and magnitude cuts described. After a visual inspection on the selected sources to eliminate spurious sources, a sample of 162 sources was obtained. To eliminate obvious AGN from this catalog, this list was cross-correlated with the MUSYC ECDF-S X-ray catalog, described in the previous chapter. This matched 2 sources and one other source showed X-ray emission when performing X-ray photometry at the position of the detection in the narrow-band image. These sources were excluded from the main LAE sample. The distribution of the remaining 159 LAEs used in this work is shown in figure 5.3. Notice that the fraction of AGN is very low, so even including those sources will not make any difference for our clustering analysis.
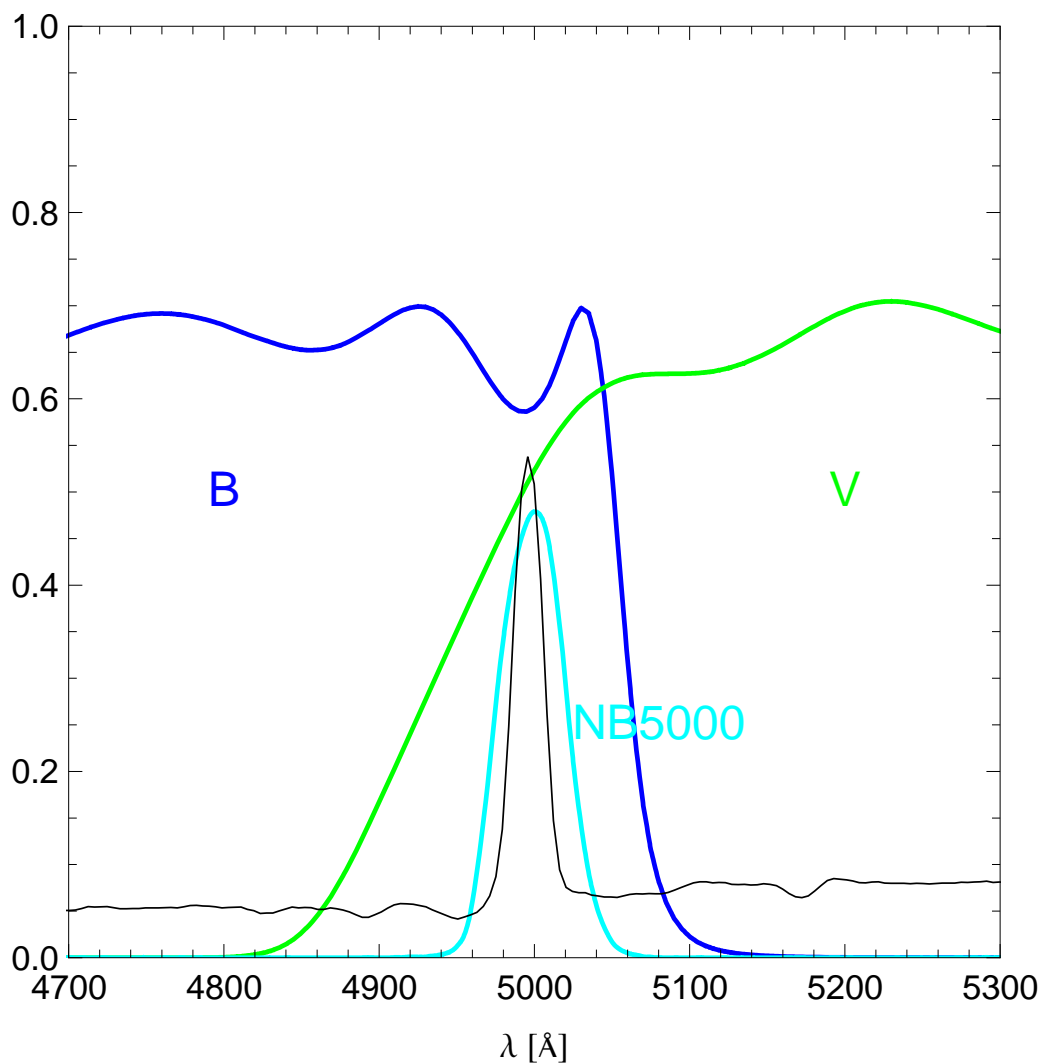
Figure 5.1 Transmission curves for the B,V and NB5000 filters used in the LAE selection, shown in blue, green and cyan, respectively. The black curve is a spectrum of a high-EW LBG, similar to an LAE spectrum. At $z \sim 3.1$, the Lyα emission lines falls in the NB5000 filter, creating an excess in the NB5000 - BV color.
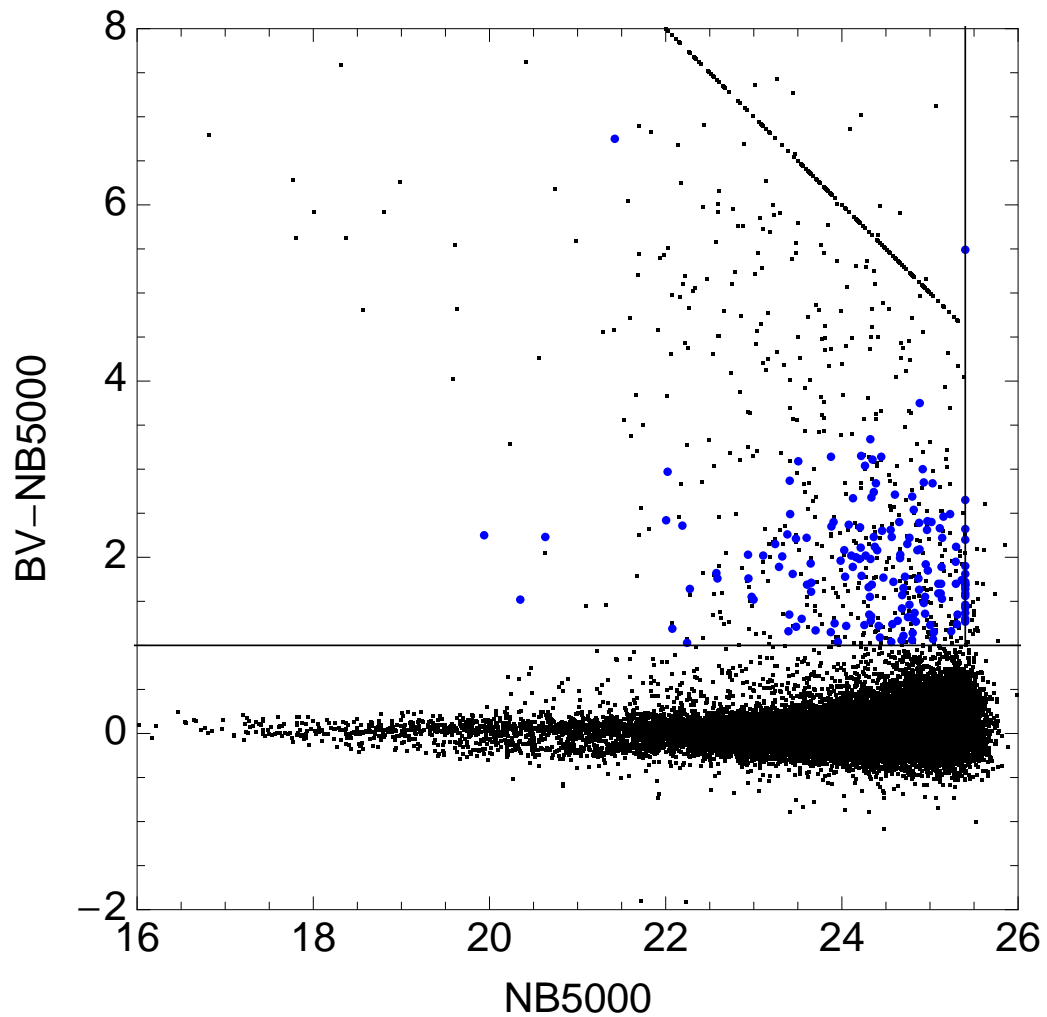
Figure 5.2 NB5000 versus B+V color-magnitude plot of all sources in the narrow-band detected catalog of ECDF-S (black points). LAE candidates are marked with blue circles. The color excess cut is represented by the horizontal line, while the flux limit cut is the vertical line. Black points inside this selection box are sources rejected upon visual inspection of the images. The points lying in the diagonal line at the upper right side are dropouts in the BV magnitude, and have been set to BV = 30.
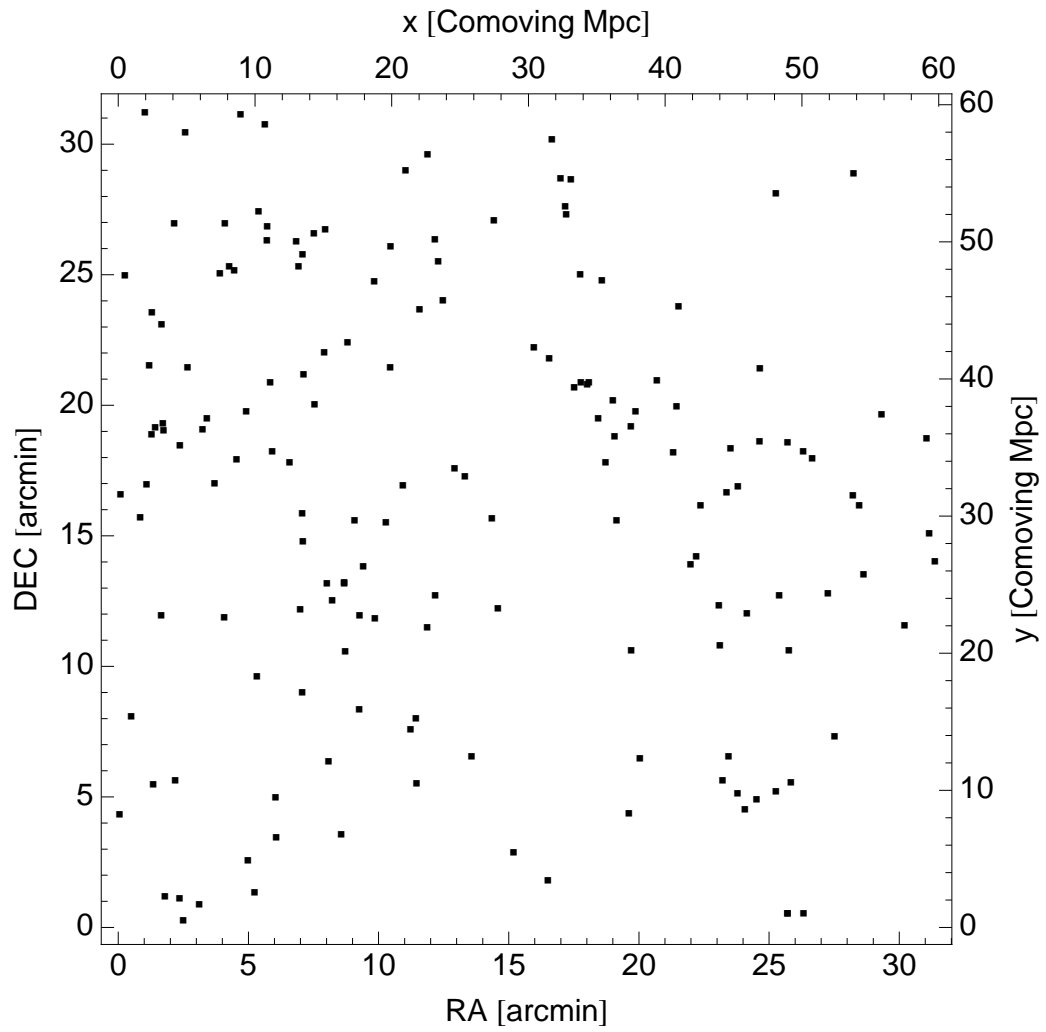
Figure 5.3 R.A. and Dec. positions of LAEs selected in ECDF-S.

5.3 Spectroscopic confirmations

From the spectroscopic follow-up program, we confirmed 84 LAE candidates in three fields, listed in appendix A. In the case of LAEs, the only feature that appears in the spectrum is the Lyα emission line. Despite this, distinguishing a real LAE from a low redshift emission line galaxy (ELG) (the typical interloper) was not complicated: no matter what line from an ELG happens to trigger a detection in the narrow band filter, it is always accompanied by at least a couple of neighboring nebular lines, as can be seen in the bright example in figure 2.6. With the broad wavelength coverage available in IMACS spectra, this discrimination was fairly easy. Another issue for these sources is that the EW determination is almost impossible from the spectra, since the continuum is not detected there.

Figure 5.4 shows the observed distribution in redshift of the confirmed LAEs in SDSS 1030, CW 1255 and ECDF-S, as well as the whole sample. Notice that the transmission curve of the narrow-band filter is not shaped as a top-hat function, but rather more similar to a gaussian. Because of this, galaxies with higher EW are more likely to get selected than galaxies with a low EW in the wings of the curve. Since galaxies with high EW are less numerous than galaxies with low EWs, the actual selection function is narrower than the filter transmission function. To model this effect, a Monte-Carlo simulation was perfomed, where a large number of LAE were distributed over the redshift trange $3.0 < z < 3.2$, and equivalent widths were assigned to them following the distribution determined by Gronwall et al. (2007). This distribution was determined from a deconvolution of the observed NB5000 flux distribution of the entire LAE sample, taking into account the photometric error function and the filter transmission curve. After applying the color cut for LAE selection we obtained the expected redshift distribution, which was then used to deproject the angular correlation function. An additional effect detected in this redshift distribution was that, at the telescope, the mean wavelength of this narrow-band filter is 9 Å bluer than what is reported by CTIO. The precise shift was constrained from observations of planetary nebulae (N. Bond, priv. comm.). Another thing to notice is the presence of a fairly clustered region in ECDF-S: figure 5.4 shows a spike in the redshift distribution obtained in this field between $3.085 < z < 3.09$,

and figure 5.5 shows the R.A.-Dec. positions of the spectroscopically confirmed LAEs in this field. Note that the redshift distribution function that enters the deprojection equation is meant to describe the *selection* in redshift, and should not include the spikes caused by large scale clustering.

5.4 Deprojection of narrow redshift distributions

Recently, Simon (2007) showed that Limber's approximation for the deprojection equation 1.28 starts to break down for narrow-band surveys such as these. To understand why, we need to take a closer look at the derivation of Limber's equation. The general relation, which only assumes spatial isotropy and homogeneity is eq. 1.27 from chapter 1:

$$\omega(\theta) = \int_0^\infty \int_0^\infty dr_1 dr_2 p_1(r_1) p_2(r_2) \xi \left(\sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta)} \right) \qquad (5.1)$$

(this equation also assumes that $\xi(r)$ is a constant within the survey volume, see Simon (2007) for the most general case). Changing integration variables to the mean and difference between the previous variables,

$$\bar{r} = \frac{r_1 + r_2}{2} \qquad \text{and} \qquad \Delta r = r_1 - r_2, \qquad (5.2)$$

make the previous double integral read:

$$\omega(\theta) = \int_0^\infty \int_{-2\bar{r}}^{2\bar{r}} d\bar{r} d\Delta r p_1 \left(\bar{r} - \frac{\Delta r}{2} \right) p_2 \left(\bar{r} + \frac{\Delta r}{2} \right) \xi(R) \qquad (5.3)$$

where

$$R(\bar{r}, \Delta r) = \sqrt{2\bar{r}^2 (1 - \cos\theta) + \frac{\Delta r^2}{2}(1 + \cos\theta)}. \qquad (5.4)$$

The approximation made by Limber consists in assuming that the width of the functions p_1 and p_2 is much greater than the correlation length of ξ. This makes the
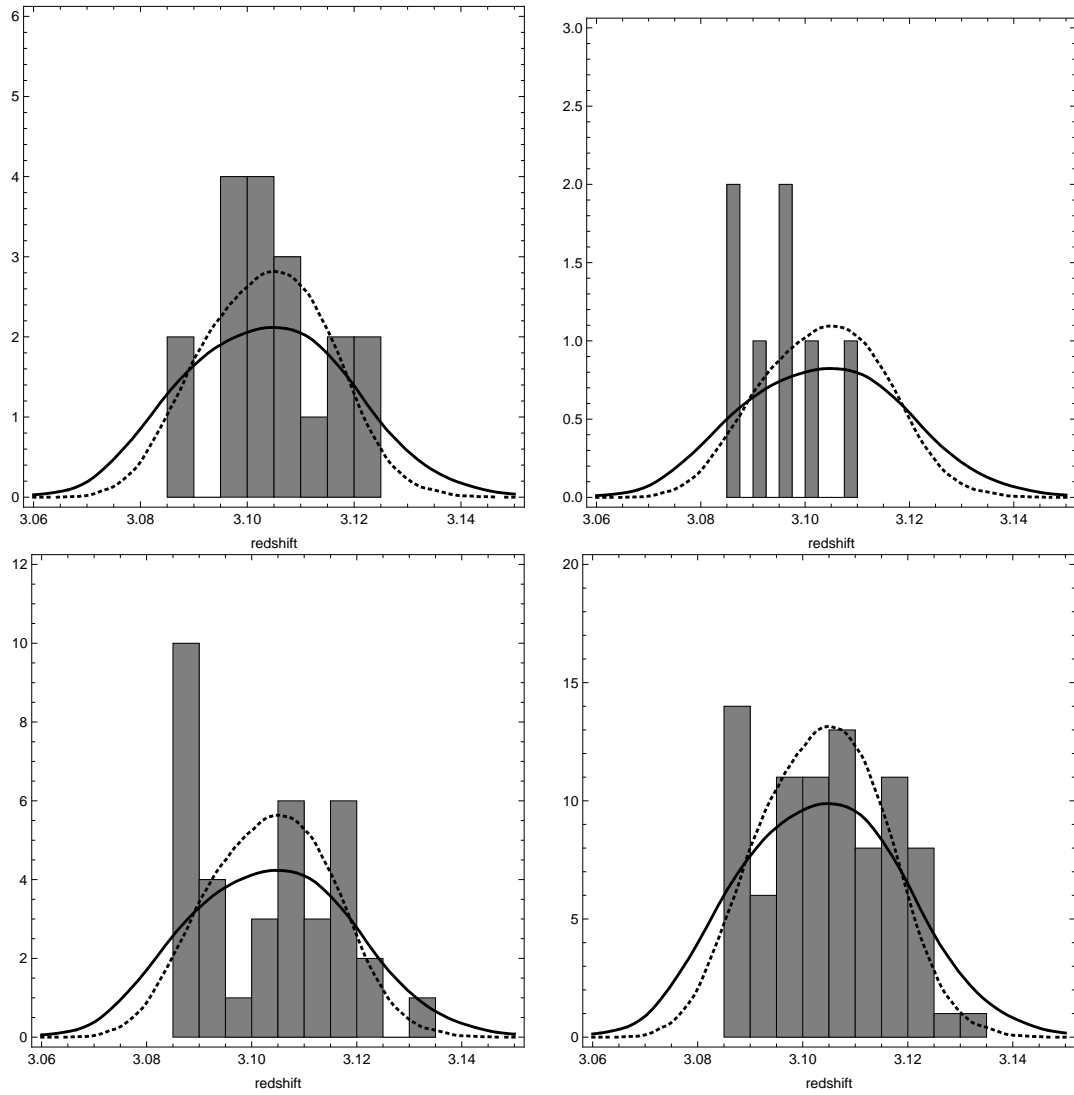
Figure 5.4 Redshift histograms of confirmed LAEs, together with the filter curve in redshift (solid curve) and the expected redshift distribution (dotted curve). Upper left panel corresponds to the SDSS 1030 field, upper right to CW 1255, lower left ECDF-S, and lower right is the full dataset over the three fields.
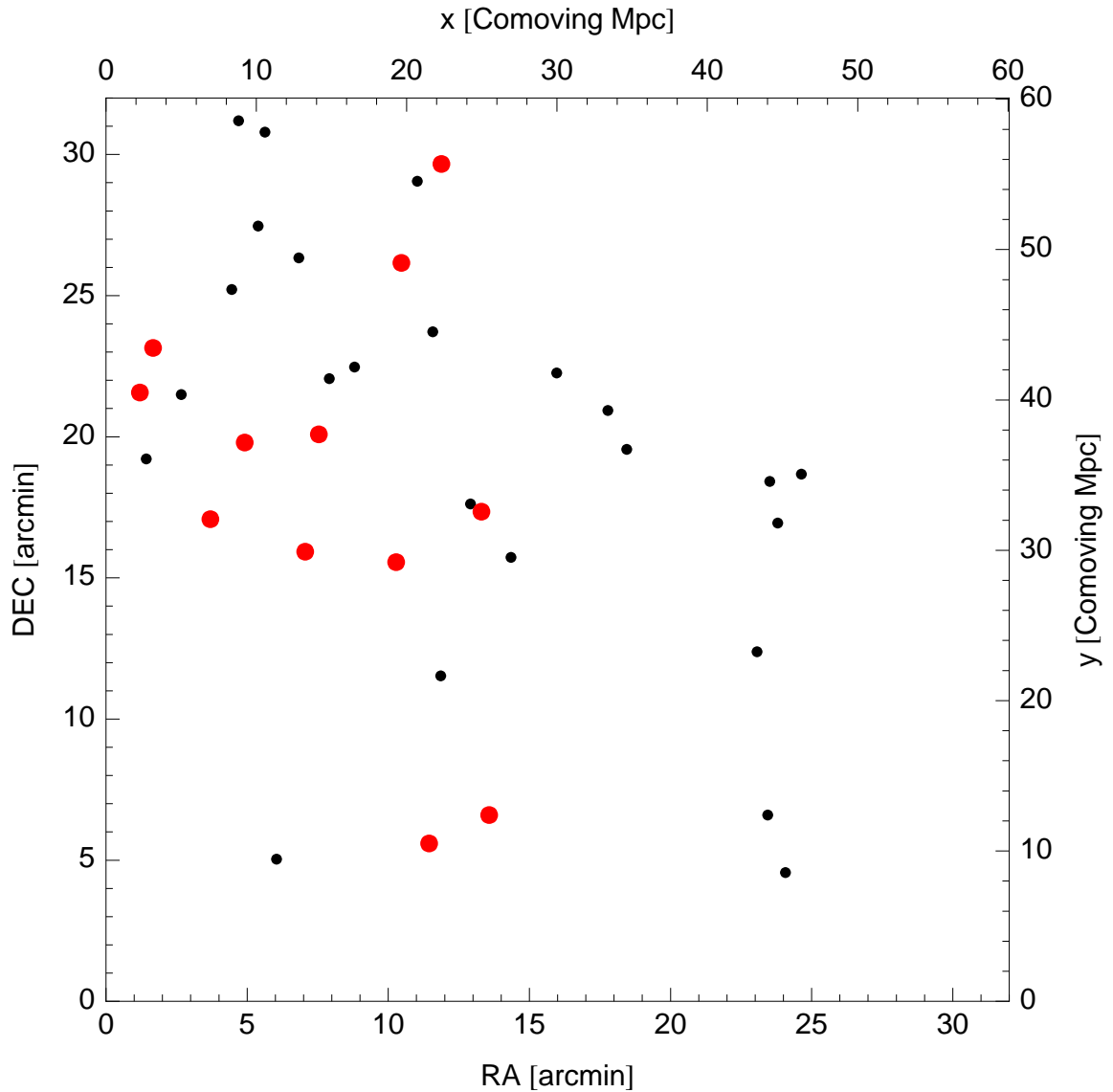
Figure 5.5 R.A.-Dec. positions of spectroscopically confirmed LAEs in ECDF-S. Red dots correspond to LAEs with redshift in the range $3.085 < z < 3.09$, corresponding to the spike in the redshift distribution of the previous figure, in the lower left plot. From these, the group in the center-left seem to be slightly more clustered than the rest, explaining the spike in the redshift distribution.

following approximations possible: first,

$$p_1\left(\bar{r} - \frac{\Delta r}{2}\right) p_2\left(\bar{r} + \frac{\Delta r}{2}\right) \xi(R) \approx p_1(\bar{r})p_2(\bar{r})\xi(R) \tag{5.5}$$

and then,

$$\int_{-2\bar{r}}^{2\bar{r}} d\Delta r \xi(R) \approx \int_{-\infty}^{\infty} d\Delta r \xi(R). \tag{5.6}$$

The first one comes from the fact that $p(r)$ changes significantly only in scales much bigger than the correlation length, and at those scales ξ decays to zero. Because of that same reason the integral in Δr can be extended to infinity. This decouples both integrals and simplifies the equation:

$$\omega(\theta) = \int_0^{\infty} d\bar{r} p_1(\bar{r})p_2(\bar{r}) \int_{-\infty}^{\infty} d\Delta r \xi(R(\bar{r}, \Delta r)), \tag{5.7}$$

which, if an analytical expression for ξ is given, can be simplified into a single integral over the product of the comoving distance distributions (or the comoving distance distribution squared, in the case of auto-correlation), eq. 1.28.

The difference between the two is basically that Limber's equation yields a power law ω for a power law ξ, while the expression from Simon, given in 5.1, yields a broken power law ω for a power law ξ. If the slope of the spatial correlation function ξ is γ, then Limber's equation says that the observed angular correlation function ω will have a slope of γ - 1, while Simon's equation says that this will be true only on small scales, but for angular separations bigger than θ_{break}, the slope of ω will become steeper, equal to γ. Notice that the Limber approximation has nothing to do with the "small angle" approximation used to simplify the trigonometric functions in the previous equations.

Since in Simon's formula we no longer obtain a power law for the angular correlation function, we need to directly fit ω with the parameters r_0 and γ of the spatial correlation function. To do so, we replace $\xi(R) = (R/r_0)^{-\gamma}$ into 5.1, obtaining:

$$\omega(\theta) = r_0^{\gamma} \int_0^{\infty} \int_0^{\infty} dr_1 dr_2 p_1(r_1)p_2(r_2) R^{-\gamma}. \tag{5.8}$$

Notice that the double integral term depends on the redshift distribution and the slope γ. To compute $\omega(\theta)$ we first compute and tabulate the quantity $\omega(\theta)/r_0^\gamma$ for our LAE redshift distribution for a range of γ values, and then we can evaluate the angular correlation function for any (r_0, γ) pair multiplying by the missing term.

To compare both deprojections, is also convenient to use the quantity $\omega(\theta)/r_0^\gamma$. We evaluated it from 5.8 and from Limber's equation, and in figure 5.6 we show the difference between these two curves, for two values of γ. Simon (2007) also gave a simple rule of thumb to estimate the angular scale where the knee of this broken power law appears: if one approximates the comoving distance distribution of the survey with a top hat function of width W and central value R_c, then θ_{break} $\approx KW/R_c$ arcmin, where K is a constant that depends on γ and varies between 750 for $\gamma = 1.6$, and 1020 for $\gamma = 1.9$. We can see immediately from this why this is not a problem for the LBG and AGN case: since $W_{\mathrm{LBG}} \approx 1100$ Mpc and $R_{\mathrm{c,LBG}} \approx$ 6600 Mpc, $\theta_{\mathrm{break}} \approx 170$ arcmin, which is well above the angular size of each of the MUSYC fields. In the LAE case, on the other hand, $W_{\mathrm{LAE}} \approx 40$ Mpc and $R_{\mathrm{c,LAE}} \approx$ 6500 Mpc, therefore $\theta_{\mathrm{break}} \approx 6.3$ arcmin. This is within our range of interest, so we have to avoid Limber's approximation.

5.5   Results

We have measured the angular correlation function for our LAE sample using again the Landy&Szalay estimator, just as described in previous chapters, and fitted this curve for $\theta > 30''$. This lower limit corresponds to ~ 1 comoving Mpc at $z=3.1$ and, as in the LBG and AGN case, prevents us from reaching scales affected by the 1-halo term of the autocorrelation function. Since the halos hosting LAEs turned out to be smaller than those hosting LBG and AGN, we were able to decrease this lower limit slightly. To perform the fit, we actually fixed the slope of the correlation function to $\gamma=1.8$, just as in the AGN case, due to the fairly low signal to noise of this measurement, which doesn't allow us to constrain both r_0 and γ simultaneously. Figure 5.7 shows the measured angular correlation function, from which we derive a spatial correlation length of $r_0 = 3.6^{+0.8}_{-1.0}$ Mpc. This clustering strength corresponds
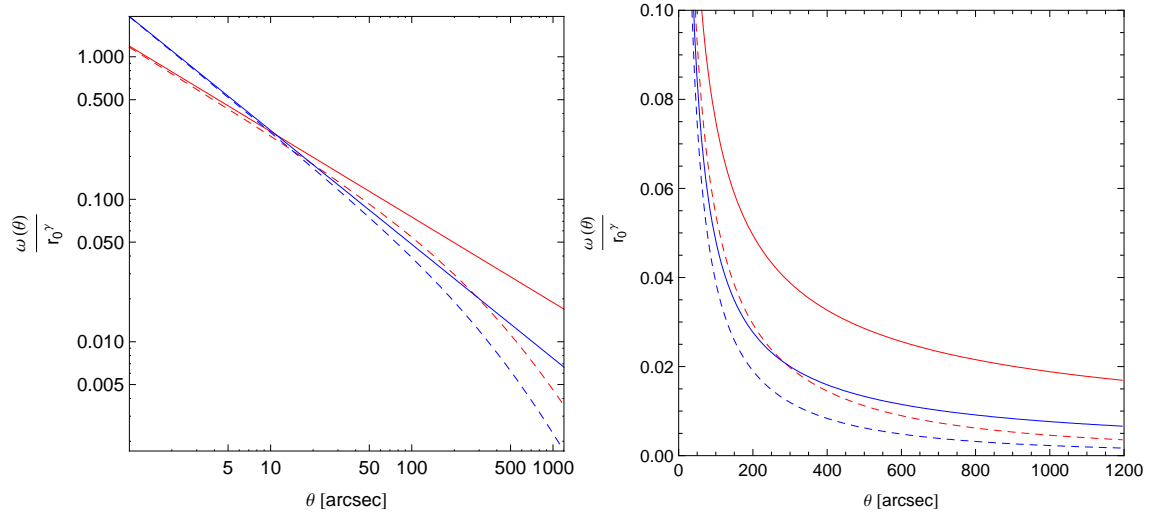
Figure 5.6 Limber's versus Simon's deprojection integrals, plotted solid and dashed, respectively. Red curves correspond to a slope γ=1.6, while blue curves correspond to γ=1.8. The curves in the left plot are the same as in the right plot, only the axes are switched from logarithmic to linear. Notice that these curves are the ideal $\omega(\theta)$ and therefore do not include the integral constraint.

to halos with an effective bias factor of $b = 1.9^{+0.4}_{-0.5}$. Within our simple halo occupation model, this bias is consistent with a population having a minimum mass of $\log(M_{\mathrm{min}}/M_{\odot}) = 10.5^{+0.5}_{-0.9}$, which corresponds to a population with median mass of $\log(M_{\mathrm{med}}/M_{\odot}) = 10.8^{+0.5}_{-0.9}$. The number density of this halo population is $n_{\mathrm{h}} = 3^{+27}_{-2} \times 10^{-2}$ Mpc^{-3}. Gawiser et al. (2007) calculated the number density of this LAE sample to be $1.5 \pm 0.3 \times 10^{-3}$ Mpc^{-3}, and thus, we can deduce an occupation fraction of 4^{+11}_{-3} %.

To put these clustering results in context, just as it was done in previous chapters, we review plot 3.15, with the LAE result included. Figure 5.8 shows the bias of this LAE sample among the same literature results presented in chapter 3, with the corresponding bias evolution of the median descendants and ancestors. The most striking result from this analysis is that the median descendant at $z = 0$ is within halos that can host L* galaxies, i.e. galaxies like the Milky Way. Another interesting result is that the K-faint sample of BX galaxies targeted by Adelberger et al. (2005a) at $z = 2.1$ also seem to be typical descendants of our LAEs, as well as blue galaxies at $z = 1$ in the DEEP2 survey (Coil et al., 2008). BM galaxies at $z = 1.7$

and red galaxy sample from DEEP2 might also be part of the high mass end of the descendants' distribution. On the other hand, there are no clustering results in the literature that fall below the upper 68% percentile envelope of the ancestors' bias track. This is not surprising, since such low mass halos would most likely contain very low mass, low luminosity galaxies, that would be hard to detect.

5.6 Discussion

We have measured the clustering strength of the MUSYC LAE sample at $z = 3.1$ in the ECDF-S field, and found that the correlation length of these galaxies is consistent with being hosted by halos with a typical mass of $\sim 6 \times 10^{10}$ $M_{\odot}$. This is the lowest dark matter mass measured for any galaxy population at this redshift and is the only population known so far at high redshift that could have evolved into halos that could host galaxies like the Milky Way at the present time. Notice, however, that only a fraction of today's L^* galaxies could have LAEs at $3.08 < z < 3.14$ as progenitors: the number density of LAEs is 1.5×10^{-3} Mpc^{-3}, a factor of $\sim$4 smaller than ϕ^* in the local universe (6.5×10^{-3} Mpc^{-3}, Lin et al., 1996) Additionally, the upper limit in the bias track of figure 5.8 shows that the possible descendants span a wide range of halo biases. The left plot on figure 5.9 shows this at present time, in terms of dark matter halo mass. Another way to look at this point is to compute the ancestor's conditional mass distribution at $z = 3.1$ for halos that at present time have a mass 10^{12} $M_{\odot}$, which is shown on the left plot of this last figure. The probability for such a halo to have an ancestor within the mass range of our LAEs hosting halos is nearly 30%, and a wide range of many smaller halos are accreted or merge to form current L^* halos. The right plot on figure 5.9 shows the distribution of these ancestors at z=3.1.
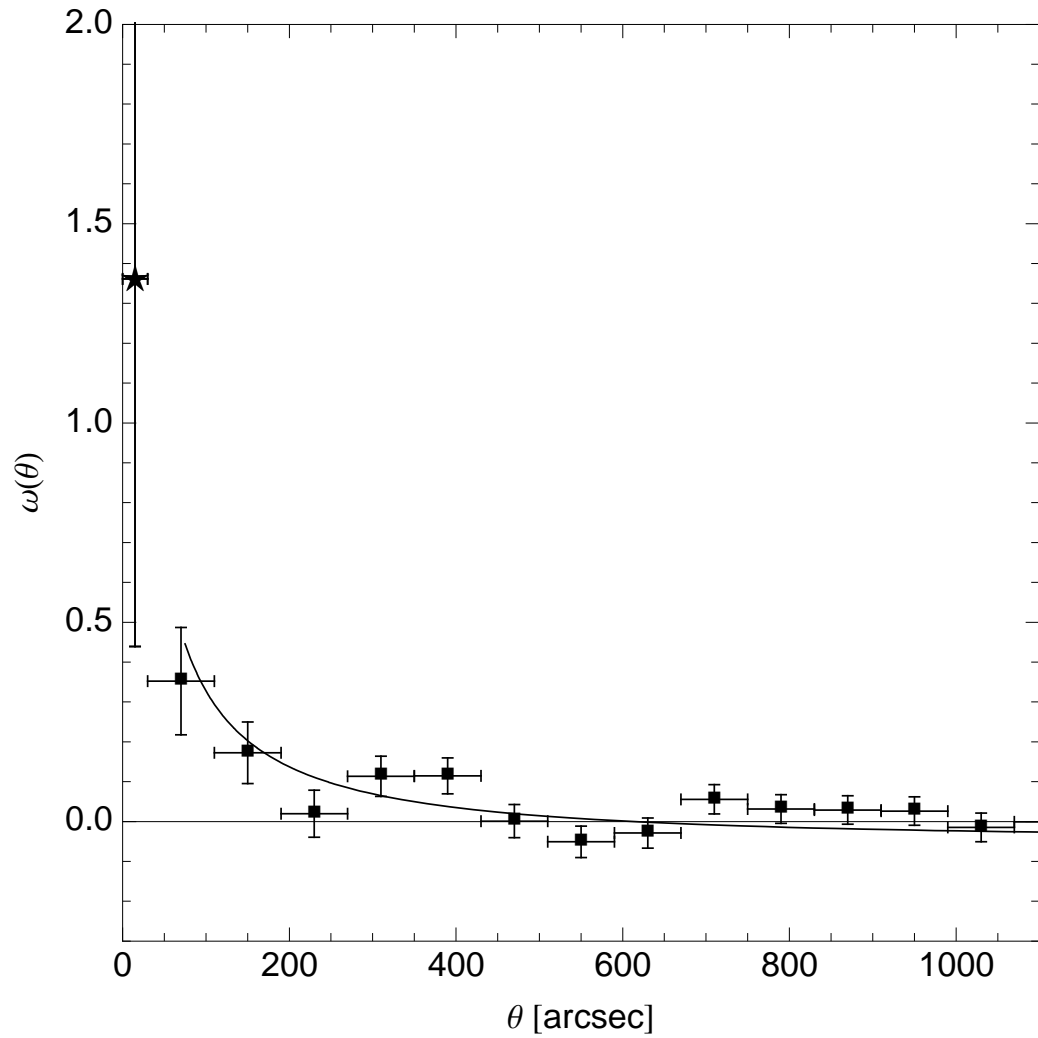
Figure 5.7 Angular autocorrelation function of LAEs in ECDF-S. First bin (star) is not included in the fit, but only shown for reference. Solid line is the best fit model.
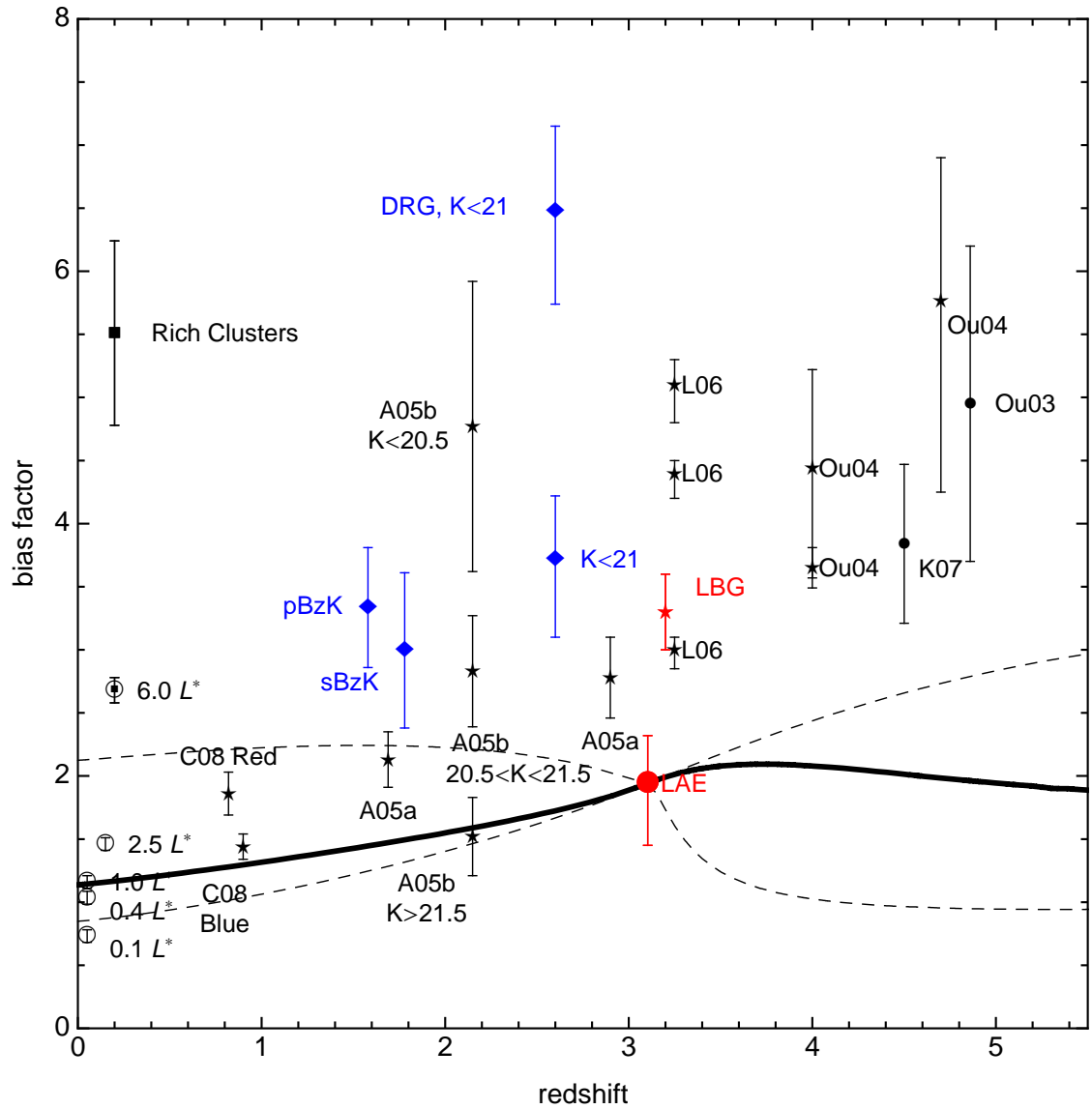
Figure 5.8 Bias evolution in redshift for our LAEs (red filled circle) and literature results for a variety of galaxy populations, including LBGs from this work (red star) and other results derived from MUSYC data (blue symbols). Stars are galaxies selected from their UV colors/magnitudes, diamonds are K selected, filled circles are narrow band selected and open circles are low-redshift samples. Literature points have the same labels as in figure 3.15.
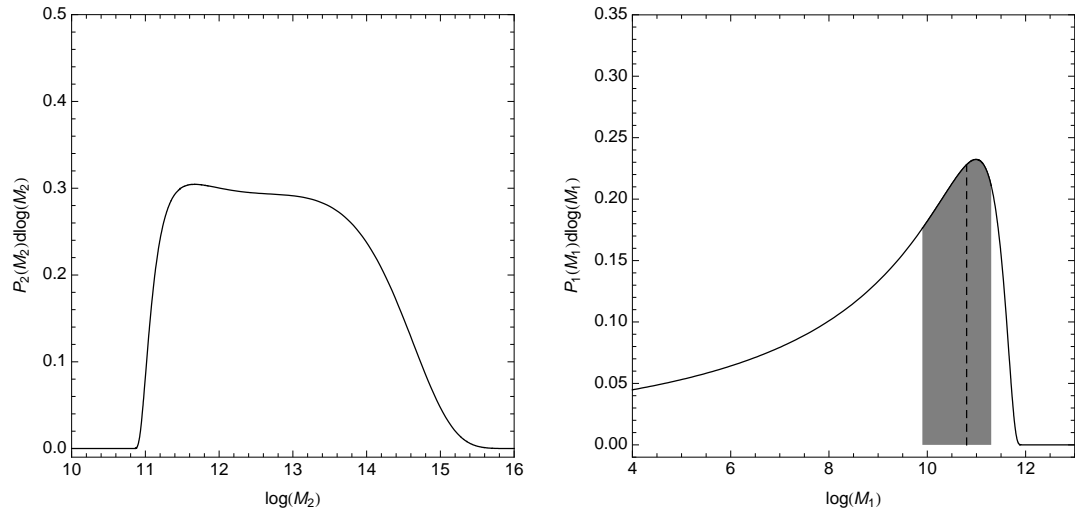
Figure 5.9 *Left:* Conditional probability distribution function of present time descendants of halos with a mass of $10^{10.8}$ M$_\odot$ (the central value for our LAE halo mass determination) at $z = 3.1$. *Right:* Conditional probability distribution function of the ancestors of a 10^{12} M$_\odot$ halo at $z = 3.1$. Shaded area corresponds to the range of masses included within the 1-σ limits around the median value estimated for LAE halos ($10^{9.9-11.3}$ M$_\odot$). Dashed vertical line marks the median value: $10^{10.8}$ M$_\odot$. Integrating within this shaded area gives a probability of 29.7%, below $10^{9.9}$ gives 64.1%, and above $10^{11.3}$ gives 6.2%.

Summary and Conclusions

In this work we have measured the clustering properties of Lyman-Break Galaxies, Active Galatic Nuclei and Lyman Alpha Emitter galaxies at redshift three in the MUSYC survey, and we have determined the basic properties of the typical dark matter halos hosting these galaxies. Additionally, based on the expected clustering evolution of dark matter halos within a ΛCDM cosmology, we have constrained the possible descendants and ancestors of these halos, with the aim of drawing possible evolutionary sequences between them.

In chapter 3 we calculated the autocorrelation function of LBG at $z \sim 3$, obtaining a clustering measurement that agrees well with previous results from Adelberger et al. (2005b), Hildebrandt et al. (2007) and Lee et al. (2006): these galaxies sit in fairly massive dark matter halos, of $\log(M_{\mathrm{med}}/M_\odot) = 11.9 \pm 0.2$, less massive than DRGs and bright K-selected galaxies' halos (Quadri et al., 2007), but more massive than the halos of the LAE sample also analysed in this work. We computed the number density of halos having this mass and obtained $n_{\mathrm{h}} = 1.2^{+1.1}_{-0.6} \times 10^{-3}$ Mpc^{-3}, which compared to the number density of our observed LBGs ($2.88 \pm 0.05 \times 10^{-4}$ Mpc^{-3}), tells us that these galaxies populate 23^{+23}_{-11} % of their hosting halos. The dark matter halo mass also lets us put a constraint on the median baryonic mass available to these galaxies for star formation: since the halo should have the same baryonic to dark matter ratio as the rest of the universe, $M_{\mathrm{b}} \approx (\Omega_{\mathrm{b}}/\Omega_{\mathrm{m}})M_{\mathrm{h}} \approx 1^{+0.5}_{-0.4}10^{11}$. Comparing this to the typical LBG stellar mass $2.5^{+5.5}_{-1.9} \times 10^{10}$ M$_\odot$ (Shapley et al., 2001) tells us that typically, about $25^{+55}_{-20}\%$ of the baryonic mass inside the halo is in stars in galaxies like the LBGs. The occupation fraction measured for these galaxies also means that our selection technique is not severely undersampling the halos, but we cannot

say anything about the galaxies that might exist in the halos of the same mass not occupied by LBGs. A natural extension of this work would be to constrain the LBG halo occupation distribution (HOD), something that would require a higher signal to noise measurement of the correlation function at small scales. Another spin-off, one that could already be attempted with the current dataset, would be to measure the clustering strength dependence on UV luminosity.

From our analysis of the LBG's conditional mass function we inferred the following possible evolutionary sequence:

$$\text{high-redshift LAEs} \Rightarrow \text{LBGs} \Rightarrow \text{BXs} \Rightarrow \text{BMs} \Rightarrow \text{Red galaxies} \Rightarrow \text{Ellipticals}$$
$$(z = 4.5\text{-}5.0) \quad (z = 3.2) \quad (z = 2.2) \quad (z = 1.7) \quad (z = 1.0) \quad (z = 0)$$

The halos of high-redshift LAEs, such as the ones studied by Ouchi et al. (2003) and Kovač et al. (2007) at redshift $\sim$4.5-5 could be the ones containing LBGs at redshift $\sim$3, which at redshift $\sim$2 could evolve into BX galaxies of intermediate K brightness, and later on into BM galaxies. At lower redshift we would expect these halos to contain early type galaxies, as in the DEEP2 samples from Coil et al. (2008), and finally contain elliptical galaxies at present time. Another way to visualize the evolution of these halos is to track the number density of each galaxy population as the halo mass grows. Figure 5.10 shows the number density of the halos hosting these galaxies as a function of their mass, along with the trajectories in this diagram of the halos presented in the bias tracks for each of these galaxies. As can be seen in this plot, the mass gain and number density evolution differs strongly with the initial halo mass, although all three halo tracks start at $z \sim 3$. Notice that although the mass and clustering strength of the halos in this sequence are consistent, baryonic content does not necessarily follow. To complete this argument, the information about the galaxies' stellar populations, star formation history, metallicity, and merger history of these halos should be included, something that could be achieved by using semi-analytical simulations of galaxy formation. The baryonic physics involved are, however, complex and not completely understood today. This is outside the scope of this thesis, but is something to keep in mind when interpreting these results, as with all the evolutionary sequences described in this work.

In chapter 4, we selected a sample of AGN based on optical and X-ray imaging, and performed clustering analysis on them. To circumvent the problem of the small size of this sample, we took advantage of the spatial overlap between the AGN and the LBGs, and calculated the cross-correlation function between these two set of galaxies. Using our previous result for LBGs we deduced the clustering strength for these AGN in a two-step scheme. This work showed that this method of X-ray selection, previously applied by Nandra et al. (2005) and Aird et al. (2008) to constrain the luminosity function of AGN, can be applied to clustering studies of these sources at high-redshift too. From the cross-correlation analysis, we deduced that the halos hosting these intermediate X-ray luminosity AGN are about one order of magnitude more massive than the halos of the LBG, reaching $4^{+8}_{-3} \times 10^{12} M_\odot$. This is similar to what was found by Adelberger et al. (2005a), but smaller than the halos of QSOs targeted by the Sloan Digital Sky Survey (Shen et al., 2007), which have masses of $2.2 \pm 0.7 \times 10^{13} M_\odot$. This is consistent with more massive galaxies tending to have more massive and more luminous AGN. The occupation fraction we estimated for these sources has a central value of order unity, but it is poorly constrained with present data. Increasing the precision in this measurement will put a constraint on the duty cycle of these AGN. A survey with wider sky coverage will be needed, however, to target enough bright sources and resolve the disagreement between the bright UV luminosity AGN sample from Adelberger et al. and Shen et al. Following the median halo descendant of our AGN sample to present time, we found that they would be at the high end of the LBG descendants distribution (see fig. 5.11), possibly hosting groups of galaxies, or poor galaxy clusters. Although these are significantly massive halos at z=0, we cannot rule out the possibility that these SMBHs end up in satellite galaxies inside the halo, and thus, not necesarily hosted by very massive galaxies.

In the case of our LAE sample examined in chapter 5, we obtained the lowest clustering strength measured for any population of $z \sim 3$ galaxies, implying dark matter halos of $\sim 8 \times 10^{10}$ $M_\odot$. This clustering result is at odds with the results obtained for blank fields at higher redshift by Ouchi et al. (2003) and Kovač et al. (2007), which show significantly more clustered population, similar to LBGs at that same redshift. Additionally, the result from the spectral energy distribution fits in Gawiser et al. (2007) showed that the bulk of this galaxy sample has a fairly low

stellar mass and star formation rate ($\sim 10^9$ M$_\odot$ and ~ 2 M$_\odot$/yr) and are also young and nearly dust free. This would be consistent with a low mass galaxy population, having their first burst of star formation. The comparison between our set of $z \sim 3$ LAE and those at higher redshift is not direct, since those samples were selected using slightly different treatment of the photometry, have different broadband and narrowband depths and thus, might be selecting different galaxy populations. Additionally, the extra gigayear of cosmic evolution between $z \sim 4.5$ - 5 to $z \sim 3$ could be enough to change the stellar populations, dust content and metallicity of the environment where these galaxies grow, changing the type of object effectively targeted even having equivalent observational selection.

From our analysis of the conditional mass function our LAE sample, we also inferred a possible evolutionary sequence:

$$\begin{array}{cccccc} \text{LAEs} & \Rightarrow & \text{K-dim BXs} & \Rightarrow & \text{Blue galaxies} & \Rightarrow & \text{L}^* \text{ galaxies} \\ (z = 3.1) & & (z = 2.2) & & (z = 1.0) & & (z = 0) \end{array}$$

The halos hosting our LAEs could be the ones hosting BX galaxies with $K > 21.5$ at $z \sim 2$, such as those selected by Adelberger et al. (2005a), while at $z \sim 1$ could host blue galaxies, such as the ones targeted by Coil et al. (2008). During the gigayear passed between $z = 3$ and 2, LAE would have to form stars at an average rate of 14 M$_\odot$/yr to reach stellar masses of 1.5×10^{10} M$_\odot$ typical of the BX galaxies with the K magnitude range mentioned above (Adelberger et al., 2005a). This is not unreasonable, given that these BX galaxies have been shown to have star formation rates of 25 M$_\odot$/yr (Reddy et al., 2005), so this first evolutionary step could be feasible, given that these LAEs increase their star formation rates correspondingly.

Given the increasing importance the astronomical comunity has given to deep, wide, multiwavelength extragalactic surveys in the last decade and in the planning of future astronomical facilities, the time for performing analysis of large scale structure has never been better. The wealth of data allows not only better statistics of these measurements, but also, given the amount of complementary information in modern surveys, to separate galaxies based on more physical parameters. This has enabled astronomers not only to learn about large scale structure and cosmology, but also

to study galaxy evolution, as we have done in the present work. In comparison with the limited amount of information that was available in photographic plate surveys in the past, this is naturally a huge improvement, and one a that will continue in the future. Many future galaxy surveys are aimed at the measurement of baryonic accoustic oscillations (BAOs), the imprint on large scale structure left by the hot plasma of the early universe. Since this imprint should be observed at very large scales ($\sim$150 Mpc), very wide surveys are required. Projects like the Dark Energy Survey (DES), VST-ATLAS or VISTA-VIKING, will cover very wide regions of the sky, but to shallower depths than MUSYC. These will allow, besides the measurements of the BAO wiggles, a high signal-to-noise measurement of the correlation function for galaxies at the bright end of the galaxy luminosity function at high redshifts, something that current surveys like the Sloan Digital Sky Survey can only accomplish at low redshift. Naturally, to expand the present work the depth would also be needed, in order to reach the bulk of the LBG population and other faint galaxies as the LAEs. The Large Synoptic Survey Telescope (LSST), on the other hand, will completely revolutionize this field, by imaging the entire southern sky with single exposures of $r \sim 24.5$ in depth, and ultimately reaching $r \sim 27.5$. By itself, these datasets will allow a huge improvement in the most classical broad-band galaxy selections, based on the Lyman and Balmer breaks. Searching for Lyman alpha or other emission lines will require, however, complementary surveys with wide coverage.

When it comes to actually answering the most fundamental questions about galaxy formation and evolution, it is clear that follow-up datasets will be critical. Radio and X-ray data will be very important in disentangling and pondering the role of AGN in the galaxy populations being studied; near and far infrared, and submillimeter data will be also crucial to deduce physical properties of the galaxies by means of Spectral Energy Distribution fitting. This will allow us to answer questions such as: Are the star formation histories of these galaxies consistent with the evolutionary sequences we have inferred for them? Is the growth of SMBH consistent with these sequences too? Is the timescale of AGN activity consistent with the models? How much do AGNs affect the star formation of the galaxies around them? Can current semianalytic models of galaxy formation reproduce the quantities we have measured here for all type of galaxies and at all redshifts?
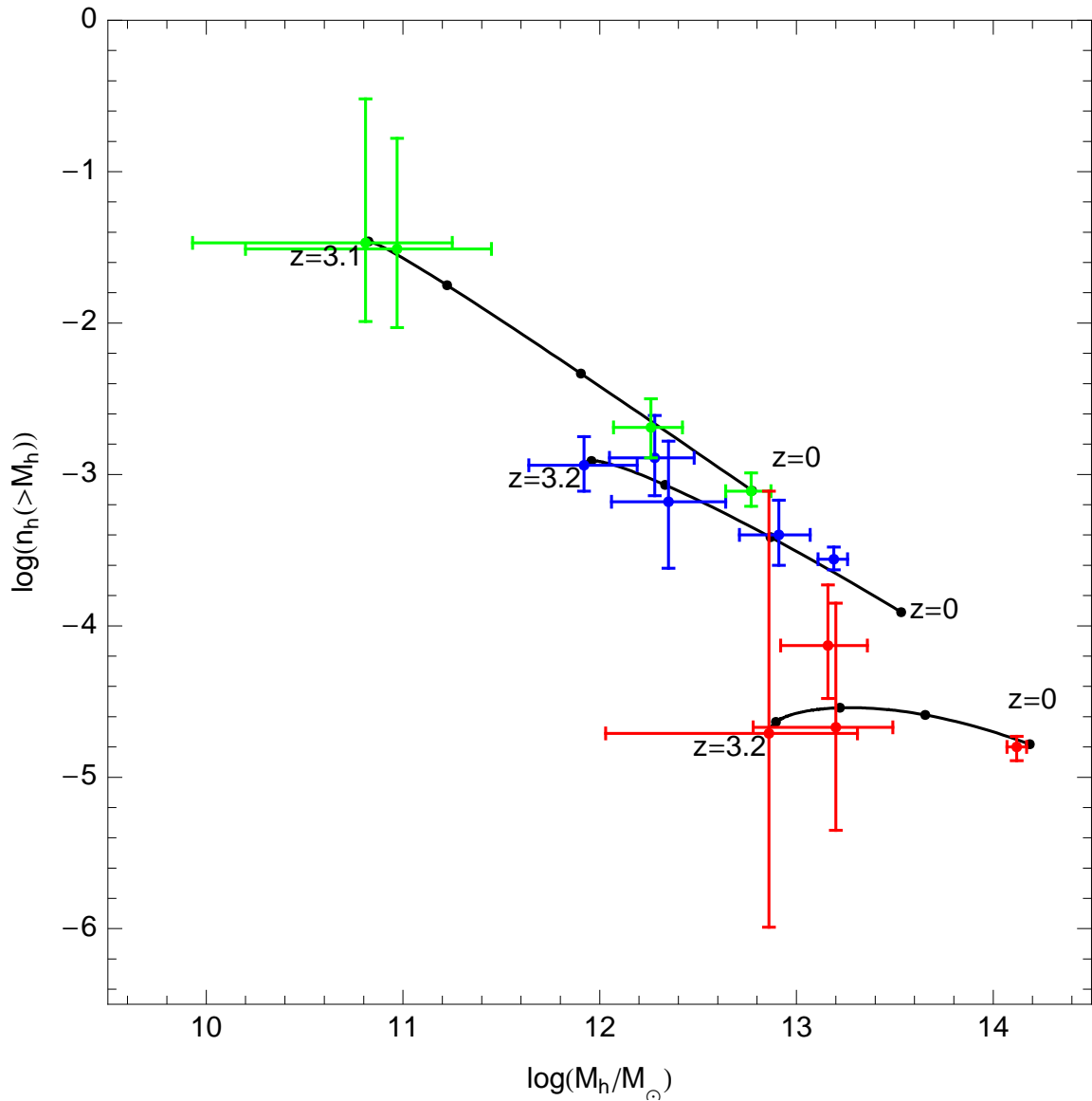
Figure 5.10 Evolution of number density as a function of halo mass, for the LAE evolutionary track (line with green points), the LBG track (line with blue points) and the AGN track (line with red points). Black points mark the position at which $z = 3, 2, 1$ and 0. Datapoints in the LAE track from left to right are: this work, dim K-selected BX galaxies at $z = 2.2$, blue galaxies at $z \sim 1$ and L^* galaxies at $z = 0$. Datapoints in the LBG track from left to right: this work, bulk of BX galaxies at $z = 2.2$, red galaxies at $z \sim 1$, BM galaxies at $z = 1.7$ and $2.5L^*$ galaxies at $z = 0$. Datapoints in the AGN track from left to right: this work, bright K-selected BX galaxies at $z = 2.2$ and $6L^*$ galaxies at $z = 0$. References to these numbers are in the text.
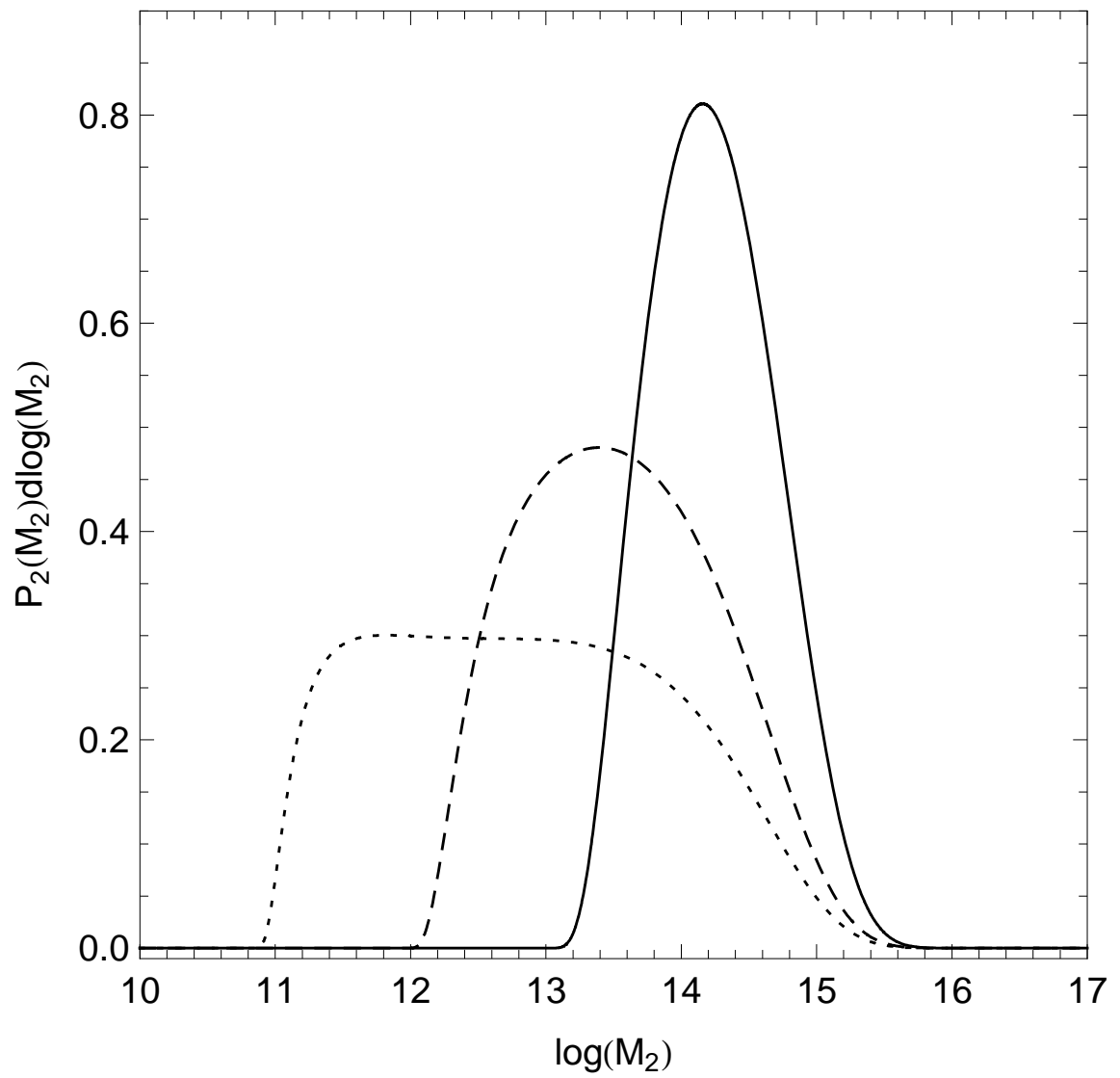
Figure 5.11 Halo mass distribution for LAE (dotted), LBG (dashed) and AGN (solid) at $z = 0$, obtained from the conditional mass function for the median halos of each of those populations.

Appendix A

Lists of spectroscopically confirmed sources

Table A.1: Spectroscopically confirmed LBGs

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	EW[2]	fwhm[3]	redshift
523	3 32 25.29	-28 3 30.85	27.0 ± 0.60	25.6 ± 0.07	24.9 ± 0.07	25.0 ± 0.07	25.7 ± 0.80	25.5 ± 1.00	9.4, 16.0, 26.0	1.8	3.254
3517	3 31 58.71	-28 2 21.91	> 27.7	25.8 ± 0.08	25.2 ± 0.08	25.1 ± 0.06	25.6 ± 0.60	25.8 ± 2.00	3.4, 6.8, 34.0	2.3	3.393
4937	3 31 49.49	-28 1 52.57	> 26.8	24.9 ± 0.10	24.3 ± 0.08	23.9 ± 0.06	23.9 ± 0.20	23.7 ± 0.40	6.8, 11.0, 18.0	2.2	3.398
8123	3 32 11.34	-28 0 47.48	> 23.0	22.3 ± 0.40	22.2 ± 0.60	22.0 ± 0.50	21.3 ± 0.80	21.2 ± 2.00	3.0, 4.9, 6.2	1.5	2.683
12485	3 31 57.08	-27 59 3.77	> 26.8	23.8 ± 0.04	23.2 ± 0.03	22.8 ± 0.02	22.7 ± 0.07	22.6 ± 0.20	11.0, 14.0, 15.0	3.9	3.38
13043	3 31 32.86	-27 58 52.28	25.5 ± 0.50	23.5 ± 0.04	22.9 ± 0.04	22.7 ± 0.03	22.8 ± 0.10	22.5 ± 0.20	3.8, 5.1, 7.6	2.4	2.95
16800	3 32 4.58	-27 57 19.51	> 27.8	25.8 ± 0.08	25.5 ± 0.09	25.3 ± 0.08	25.9 ± 0.60	24.6 ± 0.50	2.4, 8.0, 29.0	2.5	3.134
19563	3 31 34.73	-27 56 21.88	> 26.2	24.7 ± 0.20	24.3 ± 0.20	24.4 ± 0.20	24.9 ± 1.00	23.6 ± 0.70	13.0, 21.0, 39.0	2.2	3.131
19686	3 33 4.85	-27 56 18.49	> 26.6	24.8 ± 0.10	24.1 ± 0.09	23.8 ± 0.07	24.1 ± 0.30	23.4 ± 0.40	4.5, 9.0, 13.0	3.0	3.465
22517	3 31 29.30	-27 55 15.06	> 27.7	24.9 ± 0.04	24.3 ± 0.04	24.2 ± 0.03	24.6 ± 0.30	23.9 ± 0.30	12.0, 20.0, 29.0	1.9	3.175
24989	3 32 5.88	-27 54 22.18	> 27.7	24.9 ± 0.04	24.4 ± 0.03	24.2 ± 0.03	24.3 ± 0.10	23.9 ± 0.30	–	–	2.97
28238	3 31 31.19	-27 52 57.40	> 27.6	25.7 ± 0.08	24.9 ± 0.06	24.8 ± 0.05	24.3 ± 0.20	24.5 ± 0.50	4.2, 6.2, 8.6	1.6	3.06
32292	3 31 48.51	-27 51 28.40	> 25.1	25.2 ± 0.70	24.3 ± 0.50	24.1 ± 0.50	24.6 ± 2.00	24.1 ± 3.00	13.0, 20.0, 36.0	2.5	3.475
34428	3 32 14.78	-27 50 46.43	> 26.9	24.3 ± 0.05	23.8 ± 0.05	23.7 ± 0.05	23.9 ± 0.20	23.3 ± 0.30	8.3, 12.0, 15.0	2.3	3.216
37346	3 32 40.80	-27 49 35.80	26.2 ± 0.80	24.7 ± 0.10	24.0 ± 0.08	23.8 ± 0.07	23.8 ± 0.30	23.5 ± 0.40	2.2, 4.2, 7.1	1.7	3.066
37771	3 32 29.94	-27 49 28.74	> 26.1	24.7 ± 0.10	24.0 ± 0.10	23.8 ± 0.10	24.3 ± 0.70	23.9 ± 1.00	1.7, 3.5, 7.1	1.7	3.238
38048	3 31 48.51	-27 49 18.59	> 27.7	26.3 ± 0.10	25.4 ± 0.09	25.3 ± 0.09	24.9 ± 0.30	26.3 ± 2.00	4.3, 8.7, 15.0	0.9	3.598
42620	3 33 1.07	-27 47 34.04	> 27.3	25.5 ± 0.09	24.7 ± 0.07	24.8 ± 0.07	25.0 ± 0.40	24.8 ± 0.70	6.9, 11.0, 17.0	2.1	3.35
43624	3 33 16.74	-27 47 12.05	> 27.2	26.0 ± 0.20	24.9 ± 0.10	24.7 ± 0.08	25.0 ± 0.50	24.1 ± 0.40	15.0, 24.0, 43.0	4.8	3.653
50153	3 32 36.70	-27 44 57.44	> 26.7	24.5 ± 0.07	24.0 ± 0.07	23.8 ± 0.06	24.2 ± 0.30	24.0 ± 0.60	7.3, 14.0, 24.0	2.2	3.09
50813	3 32 0.88	-27 44 43.98	> 25.7	24.4 ± 0.20	23.7 ± 0.20	23.5 ± 0.10	24.5 ± 1.00	23.5 ± 1.00	16.0, 25.0, 40.0	2.2	3.487
51327	3 31 18.27	-27 44 31.09	> 27.0	24.5 ± 0.04	23.7 ± 0.04	23.4 ± 0.03	23.5 ± 0.10	23.2 ± 0.20	1.2, 2.5, 3.5	1.5	3.021
52053	3 32 28.94	-27 44 11.69	> 27.1	25.3 ± 0.10	24.6 ± 0.10	24.4 ± 0.10	24.5 ± 0.40	25.2 ± 1.00	9.2, 18.0, 32.0	1.6	3.368
53532	3 32 29.22	-27 43 40.48	> 26.8	26.6 ± 0.50	25.3 ± 0.20	25.2 ± 0.20	25.3 ± 0.80	24.5 ± 0.90	29.0, 45.0,∞	1.5	3.463
57440	3 32 35.60	-27 42 15.48	> 27.8	25.6 ± 0.06	25.1 ± 0.07	24.8 ± 0.05	25.1 ± 0.30	24.2 ± 0.40	2.8, 4.2, 5.5	1.8	3.325
58795	3 31 40.06	-27 41 44.48	26.5 ± 0.40	25.1 ± 0.07	24.9 ± 0.09	24.9 ± 0.10	24.1 ± 0.20	23.6 ± 0.50	1.6, 4.2, 13.0	2.4	3.286
63389	3 32 31.61	-27 40 8.47	> 27.5	24.5 ± 0.03	24.0 ± 0.03	23.8 ± 0.03	24.2 ± 0.10	23.8 ± 0.30	2.5, 4.9, 8.6	–	3.08
63566	3 32 22.10	-27 40 3.65	> 26.5	23.9 ± 0.05	23.4 ± 0.06	23.3 ± 0.06	23.7 ± 0.30	23.4 ± 0.50	12.0, 16.0, 22.0	2.0	3.06
67501	3 31 36.89	-27 38 32.21	> 27.8	24.5 ± 0.02	23.9 ± 0.02	23.8 ± 0.02	23.9 ± 0.20	23.8 ± 0.20	2.0, 3.0, 5.6	–	3.29
69935	3 33 20.07	-27 37 38.17	> 26.3	25.1 ± 0.20	24.5 ± 0.20	24.3 ± 0.10	24.4 ± 0.50	24.3 ± 1.00	4.6, 10.0, 23.0	1.4	3.363

Continued on Next Page...

[1]Magnitudes upper limits are 1-σ in the U and B bands, and 0.1-σ in the rest of the bands

[2]Lyα restframe equivalent width, in Å. First and third numbers are the rough lower and upper 95% confidence intervals for the central value, dominated by the uncertainty in the placement of the continuum in the spectra.

[3]Restframe full width at half maximum of the Lyα line, in Å, calculated from a gaussian fit.

[4]Object shows a double Lyα line profile. The EW value given in the table is for the full profile, but the blue and red components have EW = 3.8 and 3.3 Åin the restframe, respectively. The gaussian FWHM is given in the table for these two components too.

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	EW[2]	fwhm[3]	redshift
70051	3 32 51.34	-27 37 36.59	> 26.4	24.8 ± 0.10	24.5 ± 0.20	24.4 ± 0.10	25.0 ± 0.80	24.6 ± 1.00	5.7, 8.9, 13.0	2.9	3.06
70318	3 31 23.59	-27 37 29.82	> 27.4	26.0 ± 0.10	25.2 ± 0.10	25.3 ± 0.10	25.5 ± 0.70	27.1 ± 5.00	16.0, 24.0, 49.0	5.0	3.29
73451	3 32 27.90	-27 36 22.21	> 26.0	25.0 ± 0.20	24.2 ± 0.20	24.3 ± 0.20	25.1 ± 2.00	> 26.4	27.0, 40.0, 69.0	2.2	3.225
75562	3 32 6.78	-27 35 32.53	> 27.8	25.2 ± 0.05	25.1 ± 0.08	25.1 ± 0.07	26.1 ± 0.80	24.7 ± 0.60	5.0, 10.0, 25.0	1.3	2.993
81659	3 31 39.50	-27 33 17.60	25.7 ± 0.60	24.2 ± 0.08	23.3 ± 0.06	23.0 ± 0.05	23.1 ± 0.20	22.8 ± 0.30	1.8, 2.5, 4.1	1.6	3.441
7536	22 30 57.12	-60 55 46.20	> 27.7	25.7 ± 0.20	25.0 ± 0.10	25.0 ± 0.10	25.3 ± 0.40	25.5 ± 1.00	4.7, 7.0, 10.0	1.9	3.298
10803	22 31 11.14	-60 54 23.08	> 27.6	25.2 ± 0.10	24.3 ± 0.06	24.0 ± 0.05	23.9 ± 0.10	23.8 ± 0.30	3.6, 6.1, 7.7	4.4	3.41
15594	22 31 33.10	-60 31 57.18	> 27.6	26.1 ± 0.20	24.8 ± 0.08	25.2 ± 0.10	24.7 ± 0.20	25.0 ± 0.80	42.0, 58.0, 79.0	1.8	3.463
15617	22 31 32.38	-60 45 47.16	> 27.1	25.0 ± 0.10	24.0 ± 0.08	24.2 ± 0.10	23.9 ± 0.20	23.9 ± 0.50	6.2, 11.0, 17.0	2.0	3.194
16261	22 31 35.81	-60 34 41.99	> 27.1	25.1 ± 0.10	24.5 ± 0.10	24.5 ± 0.10	24.3 ± 0.30	23.7 ± 0.50	5.6, 10.0, 15.0	1.8	3.269
17799	22 31 42.36	-60 47 7.55	> 27.6	25.8 ± 0.20	24.8 ± 0.08	24.5 ± 0.06	24.1 ± 0.10	23.6 ± 0.20	11.0, 16.0, 33.0	2.2	3.35
18557	22 31 45.96	-60 36 33.44	> 25.5	25.1 ± 0.80	24.2 ± 0.80	24.6 ± 1.00	24.5 ± 3.00	24.2 ± 5.00	25.0, 29.0, 46.0	2.9	3.162
20205	22 31 52.29	-60 47 34.12	25.6 ± 0.50	24.4 ± 0.20	23.8 ± 0.20	23.7 ± 0.10	23.7 ± 0.40	23.3 ± 0.60	3.7, 7.4, 9.8	1.8	3.063
20334	22 31 52.58	-60 59 33.83	> 26.4	25.5 ± 0.50	24.4 ± 0.30	24.5 ± 0.40	24.7 ± 1.00	23.7 ± 1.00	6.9, 8.4, 14.0	1.9	3.173
22511	22 32 1.58	-61 0 30.28	> 26.8	24.4 ± 0.10	23.5 ± 0.09	23.5 ± 0.09	23.2 ± 0.20	24.0 ± 0.80	11.0, 15.0, 20.0	5.3	3.135
23163	22 32 5.28	-60 43 54.19	> 25.9	25.6 ± 1.00	24.4 ± 0.60	24.2 ± 0.50	24.3 ± 2.00	23.8 ± 2.00	20.0, 28.0, 70.0	1.6	3.772
25144	22 32 13.97	-60 51 12.35	> 27.7	25.6 ± 0.20	24.9 ± 0.08	25.2 ± 0.10	25.4 ± 0.40	26.1 ± 2.00	12.0, 21.0, 35.0	2.3	3.341
25345	22 32 13.39	-60 53 8.48	> 27.6	24.9 ± 0.09	24.4 ± 0.05	24.3 ± 0.06	24.0 ± 0.10	24.2 ± 0.40	9.2, 13.0, 17.0	1.9	3.24
27724	22 32 25.15	-60 50 49.70	26.8 ± 0.50	25.5 ± 0.10	25.0 ± 0.10	25.3 ± 0.10	25.2 ± 0.30	25.5 ± 1.00	13.0, 17.0, 24.0	2.0	3.05
27784	22 32 25.15	-60 49 43.90	> 27.2	25.2 ± 0.20	24.5 ± 0.10	24.8 ± 0.20	25.1 ± 0.60	24.4 ± 0.80	7.4, 12.0, 19.0	1.5	3.053
30581	22 32 36.17	-60 51 35.89	> 26.5	24.4 ± 0.20	23.9 ± 0.20	24.0 ± 0.20	23.8 ± 0.40	23.6 ± 0.80	0.9, 1.6, 2.1	4.4	3.276
30879	22 32 37.61	-61 3 27.14	25.0 ± 0.60	23.8 ± 0.30	23.8 ± 0.50	23.8 ± 0.50	23.7 ± 1.00	24.8 ± 9.00	0.8, 1.0, 1.3	3.5	2.853
32091	22 32 43.68	-60 44 17.59	> 26.9	25.2 ± 0.20	25.2 ± 0.30	25.4 ± 0.40	> 28.0	24.2 ± 0.80	18.0, 23.0, 36.0	1.8	3.134
33232	22 32 48.17	-60 54 6.16	> 27.5	25.8 ± 0.20	24.8 ± 0.08	25.0 ± 0.10	24.5 ± 0.20	> 27.8	11.0, 14.0, 19.0	2.2	3.301
36779	22 33 2.28	-60 35 50.10	> 27.2	25.1 ± 0.10	24.4 ± 0.10	24.1 ± 0.09	24.0 ± 0.20	23.9 ± 0.50	−	−	2.899
37404	22 33 5.43	-60 55 13.44	> 27.4	27.1 ± 0.70	25.3 ± 0.10	25.1 ± 0.10	24.6 ± 0.20	24.4 ± 0.40	−	−	3.507
37542	22 33 5.90	-60 44 33.43	> 27.1	26.0 ± 0.40	25.0 ± 0.20	25.4 ± 0.30	25.3 ± 0.70	> 27.2	15.0, 22.0, 31.0	1.8	3.217
39944	22 33 16.34	-60 49 53.36	> 23.2	> 22.3	22.1 ± 2.00	22.0 ± 3.00	21.5 ± 7.00	> 20.8	3.2, 5.0, 7.5	1.2	3.02
40695	22 33 19.73	-60 50 53.38	> 27.6	26.1 ± 0.30	24.7 ± 0.07	24.6 ± 0.07	24.8 ± 0.20	24.5 ± 0.50	13.0, 17.0, 26.0	1.7	3.471
41359	22 33 22.54	-60 38 49.63	> 27.4	25.2 ± 0.10	24.5 ± 0.10	24.5 ± 0.09	24.8 ± 0.30	25.2 ± 1.00	4.0, 5.7, 8.3	2.0	3.205
44929	22 33 38.49	-60 52 47.10	> 27.3	25.4 ± 0.20	24.4 ± 0.07	24.8 ± 0.10	24.8 ± 0.30	25.1 ± 1.00	12.0, 21.0, 35.0	1.7	3.343
47436	22 33 48.27	-60 30 22.64	> 27.2	25.8 ± 0.20	24.8 ± 0.10	24.7 ± 0.10	24.5 ± 0.30	23.8 ± 0.40	2.7, 4.2, 14.0	2.1	3.486
51106	22 34 5.13	-60 54 14.94	> 26.9	26.2 ± 0.50	25.1 ± 0.30	25.1 ± 0.30	26.0 ± 1.00	24.6 ± 1.00	9.5, 15.0, 26.0	2.1	3.232
51321	22 34 5.38	-60 41 1.14	> 27.6	25.9 ± 0.20	25.1 ± 0.10	25.0 ± 0.10	25.5 ± 0.40	25.1 ± 0.90	10.0, 14.0, 27.0	2.6	3.287
55951	22 34 24.67	-60 35 19.64	26.4 ± 0.30	25.4 ± 0.10	25.0 ± 0.10	25.3 ± 0.10	25.1 ± 0.30	24.8 ± 0.70	11.0, 17.0, 26.0	2.7	2.851
56360	22 34 27.02	-60 56 26.09	> 26.4	24.4 ± 0.20	23.6 ± 0.20	23.6 ± 0.20	23.6 ± 0.40	23.6 ± 1.00	0.4, 0.5, 0.8	1.6	3.119
56887	22 34 29.97	-60 57 24.44	> 25.9	24.8 ± 0.40	24.3 ± 0.60	24.6 ± 0.70	24.3 ± 1.00	> 25.5	6.5, 13.0, 26.0	4.6	2.839
213	10 29 28.32	+5 9 25.97	> 26.9	24.7 ± 0.09	24.3 ± 0.05	24.0 ± 0.04	24.0 ± 0.09	23.8 ± 0.30	8.9, 15.0, 25.0	2.5	2.928
608	10 29 25.80	+5 9 40.89	> 27.0	25.7 ± 0.20	25.4 ± 0.10	25.4 ± 0.20	25.1 ± 0.30	25.6 ± 2.00	6.9, 12.0, 23.0	4.1	3.066
2529	10 31 15.70	+5 10 40.79	> 27.0	26.0 ± 0.20	25.0 ± 0.08	25.3 ± 0.10	25.9 ± 0.40	25.8 ± 2.00	7.7, 13.0, 16.0	2.9	3.28
2876	10 29 33.48	+5 10 51.82	> 27.1	25.6 ± 0.20	25.1 ± 0.10	25.4 ± 0.10	25.4 ± 0.30	24.5 ± 0.60	7.2, 12.0, 24.0	2.6	3.138

Continued on Next Page. . .

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	EW[2]	fwhm[3]	redshift
4093	10 29 19.46	+5 11 21.49	> 27.2	25.6 ± 0.20	24.7 ± 0.07	24.8 ± 0.09	25.3 ± 0.30	25.0 ± 0.70	12.0, 20.0, 29.0	2.8	3.068
8458	10 30 14.57	+5 13 16.57	26.7 ± 0.50	25.0 ± 0.08	24.8 ± 0.06	24.8 ± 0.07	24.9 ± 0.10	24.2 ± 0.30	19.0, 36.0, 64.0	2.8	2.891
13890	10 29 24.91	+5 15 29.33	> 27.3	25.2 ± 0.09	24.5 ± 0.04	24.1 ± 0.04	23.9 ± 0.05	23.5 ± 0.10	12.0, 37.0, 46.0	2.6	3.33
15874	10 29 24.31	+5 16 16.70	26.1 ± 0.40	24.3 ± 0.04	24.0 ± 0.03	23.8 ± 0.03	23.8 ± 0.05	23.6 ± 0.20	3.5, 5.8, 11.0	2.9	2.95
15920[4]	10 29 42.31	+5 16 12.86	27.5 ± 1.00	25.1 ± 0.08	24.3 ± 0.04	24.1 ± 0.03	24.2 ± 0.06	23.8 ± 0.20	3.8, 7.6, 13.0	2.2/2.7	2.945
16074	10 30 0.89	+5 16 27.53	> 27.6	26.9 ± 0.40	25.5 ± 0.10	25.4 ± 0.10	25.4 ± 0.20	25.0 ± 0.60	7.1, 13.0, 22.0	2.5	3.62
16711	10 30 21.57	+5 16 41.44	> 27.7	25.9 ± 0.20	24.9 ± 0.06	24.6 ± 0.05	24.6 ± 0.10	24.3 ± 0.30	8.1, 11.0, 16.0	2.4	3.56
17983	10 30 14.81	+5 17 12.92	> 27.6	26.4 ± 0.30	25.5 ± 0.10	25.9 ± 0.20	25.4 ± 0.20	25.0 ± 0.60	7.3, 17.0, 24.0	3.5	3.128
18656	10 29 31.85	+5 17 27.97	> 27.3	25.5 ± 0.10	25.0 ± 0.07	25.0 ± 0.08	25.4 ± 0.20	25.5 ± 1.00	2.4, 5.4, 11.0	2.6	3.084
19162	10 31 20.18	+5 17 38.64	> 27.1	25.4 ± 0.10	24.5 ± 0.04	24.3 ± 0.04	23.9 ± 0.05	23.6 ± 0.20	4.8, 8.4, 11.0	2.3	3.19
19737	10 31 10.73	+5 17 53.21	> 27.4	26.5 ± 0.30	25.1 ± 0.07	25.1 ± 0.09	23.6 ± 0.04	22.2 ± 0.05	5.0, 11.0, 19.0	1.4	3.24
22340	10 29 43.99	+5 18 52.61	> 27.6	25.2 ± 0.10	24.3 ± 0.04	24.2 ± 0.04	24.2 ± 0.06	24.0 ± 0.20	13.0, 21.0, 33.0	3.2	3.298
23455	10 30 12.91	+5 19 18.57	> 27.6	25.8 ± 0.20	24.9 ± 0.07	24.5 ± 0.06	24.4 ± 0.08	24.1 ± 0.30	2.8, 4.2, 8.5	1.8	3.714
27576	10 29 48.05	+5 21 12.68	26.0 ± 0.20	24.9 ± 0.09	24.2 ± 0.05	24.1 ± 0.05	24.1 ± 0.07	23.4 ± 0.20	1.5, 3.5, 6.6	2.7	2.944
29373	10 29 50.11	+5 22 2.30	> 27.6	25.8 ± 0.20	25.1 ± 0.10	25.1 ± 0.10	25.3 ± 0.20	> 27.7	7.2, 12.0, 16.0	3.1	3.003
30449	10 30 50.28	+5 22 29.72	> 27.5	25.7 ± 0.20	25.0 ± 0.07	24.7 ± 0.06	24.7 ± 0.10	24.7 ± 0.40	1.9, 3.1, 4.8	1.8	3.139
36104	10 31 32.83	+5 25 1.75	> 27.2	25.4 ± 0.10	25.0 ± 0.07	24.8 ± 0.06	24.9 ± 0.10	24.5 ± 0.40	9.0, 13.0, 18.0	2.2	2.87
36112	10 30 26.86	+5 25 4.13	> 27.7	25.5 ± 0.10	24.9 ± 0.08	24.7 ± 0.07	24.6 ± 0.10	24.4 ± 0.40	3.4, 6.9, 12.0	2.2	2.786
37419	10 31 3.67	+5 25 34.32	> 27.5	25.3 ± 0.10	24.4 ± 0.04	24.0 ± 0.03	23.8 ± 0.05	23.6 ± 0.20	16.0, 23.0, 35.0	2.2	3.267
41694	10 31 31.54	+5 27 31.02	> 27.3	27.2 ± 0.60	25.5 ± 0.10	25.6 ± 0.10	25.2 ± 0.20	25.1 ± 0.70	8.4, 16.0, 34.0	1.9	3.418
41894	10 30 35.11	+5 27 34.91	> 27.5	25.5 ± 0.10	24.9 ± 0.06	24.8 ± 0.06	24.8 ± 0.10	24.9 ± 0.50	4.7, 14.0, 24.0	5.2	3.22
45347	10 29 45.02	+5 29 15.05	> 27.6	25.5 ± 0.20	24.9 ± 0.08	25.0 ± 0.10	25.1 ± 0.20	24.4 ± 0.50	11.0, 20.0, 27.0	3.0	2.988
50140	10 29 18.86	+5 31 35.17	> 27.4	25.9 ± 0.20	25.4 ± 0.10	25.2 ± 0.10	25.6 ± 0.30	24.3 ± 0.40	2.1, 3.7, 6.1	1.9	3.27
50834	10 31 11.35	+5 31 53.38	> 27.4	25.4 ± 0.10	24.9 ± 0.06	24.8 ± 0.06	24.6 ± 0.09	24.7 ± 0.40	5.0, 13.0, 18.0	2.3	2.993
51592	10 30 42.02	+5 32 18.38	> 27.5	25.9 ± 0.20	25.2 ± 0.08	25.4 ± 0.10	25.4 ± 0.20	> 28.2	6.5, 11.0, 17.0	2.6	3.146
55652	10 30 10.41	+5 34 17.14	> 27.6	25.8 ± 0.20	25.1 ± 0.10	25.2 ± 0.10	26.7 ± 0.90	24.9 ± 0.70	8.4, 12.0, 17.0	1.9	3.173
61912	10 31 7.68	+5 37 14.48	> 27.5	26.5 ± 0.30	25.5 ± 0.10	25.3 ± 0.10	25.3 ± 0.20	24.5 ± 0.40	2.1, 3.4, 4.3	1.4	3.39
64416	10 31 6.70	+5 38 38.44	> 27.5	26.3 ± 0.30	25.3 ± 0.10	25.3 ± 0.10	25.5 ± 0.20	24.9 ± 0.70	10.0, 16.0, 31.0	1.2	3.486
711	12 54 35.59	+1 13 44.80	> 28.1	26.5 ± 0.20	26.5 ± 0.20	26.7 ± 0.30	26.5 ± 0.80	25.3 ± 0.50	13.0, 22.0, 67.0	1.7	3.467
1049	12 54 36.33	+0 57 53.07	26.5 ± 0.20	25.2 ± 0.06	24.9 ± 0.06	24.8 ± 0.06	25.1 ± 0.20	24.8 ± 0.30	5.3, 8.3, 11.0	1.9	2.962
5056	12 54 46.18	+1 11 41.61	25.8 ± 0.10	24.5 ± 0.04	24.4 ± 0.04	24.3 ± 0.04	24.7 ± 0.20	24.2 ± 0.20	4.9, 6.1, 8.2	2.9	2.91
5586	12 54 47.69	+1 10 54.65	> 28.1	26.6 ± 0.20	25.7 ± 0.10	25.4 ± 0.10	25.4 ± 0.30	25.6 ± 0.80	4.3, 6.5, 11.0	1.6	3.637
5620	12 54 47.78	+0 58 39.00	27.3 ± 0.50	25.7 ± 0.09	25.4 ± 0.09	25.6 ± 0.10	25.7 ± 0.40	25.9 ± 0.90	1.2, 2.2, 4.9	1.5	3.089
6681	12 54 49.85	+1 8 7.40	> 28.0	25.5 ± 0.09	25.0 ± 0.07	24.7 ± 0.06	25.0 ± 0.20	25.1 ± 0.50	–	–	3.185
9562	12 54 56.93	+1 2 51.55	> 28.1	26.1 ± 0.10	25.8 ± 0.10	26.0 ± 0.20	27.1 ± 1.00	> 28.5	49.0, 66.0, 120.0	3.2	3.058
9686	12 54 57.29	+1 5 40.22	> 28.1	26.3 ± 0.20	26.1 ± 0.20	26.1 ± 0.20	26.2 ± 0.70	25.6 ± 0.80	6.1, 11.0, 21.0	2.7	2.962
11400	12 55 1.30	+1 6 44.03	> 28.1	26.7 ± 0.30	25.9 ± 0.20	25.6 ± 0.10	24.7 ± 0.20	24.6 ± 0.30	11.0, 15.0, 22.0	3.6	3.534
11967	12 55 2.42	+1 11 0.48	> 28.1	26.7 ± 0.30	25.7 ± 0.10	25.3 ± 0.10	25.7 ± 0.40	24.9 ± 0.40	13.0, 21.0, 53.0	1.5	3.695
12876	12 55 4.58	+1 3 17.28	27.5 ± 0.70	26.0 ± 0.10	26.3 ± 0.20	26.6 ± 0.40	27.2 ± 2.00	27.3 ± 4.00	7.5, 15.0, 25.0	3.4	3.006
13041	12 55 4.92	+1 8 16.45	> 28.0	26.9 ± 0.30	25.8 ± 0.10	25.5 ± 0.10	25.6 ± 0.40	25.2 ± 0.50	8.8, 15.0, 27.0	2.1	3.52
13317	12 55 5.42	+1 22 20.90	27.3 ± 0.50	26.0 ± 0.10	25.6 ± 0.10	25.6 ± 0.10	26.0 ± 0.60	26.5 ± 2.00	–	2.0	3.101

Continued on Next Page. . .

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	EW[2]	fwhm[3]	redshift
13373	12 55 5.78	+0 58 48.01	> 28.2	26.3 ± 0.20	25.6 ± 0.10	25.6 ± 0.10	25.3 ± 0.30	26.8 ± 2.00	2.3, 4.6, 6.9	3.9	3.342
15833	12 55 10.99	+1 20 10.20	27.9 ± 0.80	25.4 ± 0.07	24.6 ± 0.04	24.1 ± 0.03	23.8 ± 0.06	23.6 ± 0.10	2.3, 4.6, 6.9	2.6	3.37
17315	12 55 14.66	+0 55 46.08	> 28.2	26.7 ± 0.20	25.1 ± 0.06	25.3 ± 0.08	25.5 ± 0.30	24.8 ± 0.30	43.0, 87.0, 870.0	3.2	3.608
18749	12 55 17.93	+1 9 47.30	> 28.2	26.6 ± 0.20	25.4 ± 0.09	25.5 ± 0.10	25.9 ± 0.50	> 28.6	13.0, 29.0, 89.0	1.6	3.488
22595	12 55 27.02	+1 11 53.95	> 28.2	26.4 ± 0.20	25.8 ± 0.10	26.0 ± 0.20	26.7 ± 1.00	26.8 ± 2.00	9.0, 27.0, 67.0	2.8	3.445
23047	12 55 28.10	+1 7 13.56	27.5 ± 0.60	26.0 ± 0.10	26.1 ± 0.20	26.1 ± 0.20	25.8 ± 0.40	> 28.5	13.0, 19.0, 30.0	2.8	2.943
23279	12 55 28.70	+0 59 38.77	> 28.1	26.2 ± 0.20	25.5 ± 0.10	25.6 ± 0.10	25.5 ± 0.30	25.9 ± 1.00	15.0, 21.0, 32.0	2.2	3.058
24479	12 55 31.30	+1 2 45.22	> 28.1	26.0 ± 0.10	25.5 ± 0.10	25.0 ± 0.07	24.9 ± 0.20	24.6 ± 0.30	2.4, 3.9, 6.6	1.6	3.094
25460	12 55 33.77	+0 54 18.97	> 28.1	26.0 ± 0.10	25.4 ± 0.10	25.2 ± 0.08	25.2 ± 0.20	26.2 ± 1.00	7.0, 9.3, 19.0	1.4	3.307
26884	12 55 36.81	+1 1 47.18	> 28.2	26.2 ± 0.20	25.4 ± 0.09	25.5 ± 0.10	26.4 ± 0.70	25.2 ± 0.50	23.0, 33.0, 43.0	2.4	3.153
27249	12 55 37.51	+1 0 22.01	25.9 ± 0.10	24.7 ± 0.04	24.4 ± 0.04	24.5 ± 0.05	24.5 ± 0.10	24.2 ± 0.20	2.5, 4.3, 8.6	3.0	2.94
29493	12 55 42.91	+1 8 17.28	> 28.1	26.7 ± 0.30	26.4 ± 0.20	26.3 ± 0.30	26.5 ± 0.80	25.7 ± 0.80	8.8, 15.0, 22.0	1.3	3.563
36227	12 55 57.86	+1 18 0.81	> 28.0	26.2 ± 0.20	24.8 ± 0.06	24.2 ± 0.04	24.3 ± 0.10	24.1 ± 0.20	13.0, 27.0, 40.0	2.1	3.75
41042	12 56 9.07	+1 4 43.95	27.9 ± 0.90	26.2 ± 0.20	25.5 ± 0.10	25.3 ± 0.10	25.4 ± 0.30	25.1 ± 0.40	4.3, 7.5, 11.0	3.4	3.641
41755	12 56 10.49	+1 12 36.30	26.7 ± 0.30	25.6 ± 0.09	24.8 ± 0.06	24.8 ± 0.06	24.9 ± 0.20	24.9 ± 0.40	1.3, 2.8, 5.8	1.7	2.992
42767	12 56 12.81	+1 5 42.99	27.0 ± 0.40	25.3 ± 0.07	25.0 ± 0.06	25.1 ± 0.07	25.3 ± 0.30	25.3 ± 0.50	17.0, 22.0, 32.0	2.0	3.074
44182	12 56 15.91	+1 18 44.59	> 28.2	27.3 ± 0.40	26.5 ± 0.20	26.4 ± 0.20	26.0 ± 0.50	> 28.6	0.7, 1.4, 2.4	2.5	3.226
46309	12 56 20.21	+1 2 11.78	> 28.2	25.6 ± 0.09	24.8 ± 0.05	24.3 ± 0.04	24.3 ± 0.10	24.5 ± 0.30	4.3, 6.4, 13.0	2.2	3.194
48307	12 56 24.65	+1 15 31.15	27.3 ± 0.50	25.9 ± 0.10	25.9 ± 0.10	26.1 ± 0.20	26.7 ± 1.00	> 28.5	7.5, 13.0, 23.0	1.8	2.98
51406	12 56 31.03	+1 17 41.67	> 28.2	26.1 ± 0.10	25.3 ± 0.08	25.1 ± 0.08	25.4 ± 0.30	25.6 ± 0.70	1.2, 1.7, 2.4	2.8	3.145
52562	12 56 33.60	+1 13 25.56	27.9 ± 0.90	25.7 ± 0.10	25.6 ± 0.10	26.0 ± 0.20	25.7 ± 0.40	> 28.5	6.6, 10.0, 15.0	2.7	2.955
55120	12 56 39.53	+0 55 35.23	> 28.1	26.4 ± 0.20	25.7 ± 0.10	25.8 ± 0.10	26.8 ± 1.00	> 28.5	11.0, 14.0, 29.0	2.1	3.202
56502	12 56 42.69	+0 57 41.07	> 28.1	26.8 ± 0.30	25.9 ± 0.10	25.8 ± 0.10	27.6 ± 2.00	25.7 ± 0.80	6.9, 12.0, 20.0	1.5	3.334
57501	12 56 44.81	+1 7 49.53	> 28.0	26.1 ± 0.20	25.7 ± 0.10	25.5 ± 0.10	25.3 ± 0.30	> 28.4	7.2, 17.0, 34.0	6.2	3.16
59049	12 56 47.21	+1 2 24.55	> 28.0	28.0 ± 0.90	26.2 ± 0.20	26.1 ± 0.20	26.3 ± 0.70	26.8 ± 2.00	13.0, 32.0,∞	1.9	3.702

Table A.2: Spectroscopically confirmed AGN

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	class[2]	redshift
54242	3 31 30.65	-27 43 25.21	26.9 ± 0.60	25.8 ± 0.08	25.6 ± 0.10	25.8 ± 0.10	25.0 ± 0.40	25.9 ± 2.00	OAGN	2.253
71136	3 32 26.13	-27 37 12.54	> 27.5	26.9 ± 0.30	24.9 ± 0.08	24.9 ± 0.07	25.3 ± 0.40	24.7 ± 0.60	UAGN	3.61
70783	3 31 37.86	-27 37 24.06	> 25.9	23.8 ± 0.08	24.0 ± 0.20	23.6 ± 0.10	23.3 ± 0.40	23.5 ± 0.90	UAGN	3.00
30949	3 32 41.85	-27 52 02.39	> 27.7	23.6 ± 0.01	23.0 ± 0.01	22.5 ± 0.01	22.4 ± 0.03	22.0 ± 0.04	UAGN	3.588
54097	3 32 13.25	-27 43 29.75	> 21.4	> 21.9	> 23.7	> 23.5	> 22.4	> 21.7	OAGN	3.117
11620	3 33 03.39	-27 59 22.52	25.4 ± 0.30	23.9 ± 0.04	23.7 ± 0.05	23.6 ± 0.04	23.4 ± 0.10	23.2 ± 0.30	UAGN	2.745
13169	3 32 51.54	-27 58 42.42	> 27.7	26.1 ± 0.10	24.9 ± 0.06	24.7 ± 0.05	24.9 ± 0.30	24.6 ± 0.60	UAGN	3.195
24190	3 33 10.94	-27 54 40.25	25.7 ± 0.50	24.3 ± 0.07	23.9 ± 0.08	23.8 ± 0.08	24.0 ± 0.30	23.6 ± 0.50	UAGN	2.63
9233	3 32 30.19	-28 00 19.76	23.2 ± 0.02	22.2 ± 0.01	22.0 ± 0.01	22.0 ± 0.01	21.6 ± 0.01	21.7 ± 0.04	UAGN	2.305
51272	3 32 04.93	-27 44 31.70	> 27.9	24.8 ± 0.03	23.8 ± 0.02	23.7 ± 0.02	23.8 ± 0.10	23.5 ± 0.20	OAGN	3.47
36752	3 31 45.16	-27 49 49.58	26.4 ± 0.40	24.3 ± 0.02	23.7 ± 0.02	23.4 ± 0.02	23.4 ± 0.08	23.3 ± 0.20	UAGN	2.715
28523	3 32 44.32	-27 52 51.24	27.7 ± 1.00	24.6 ± 0.02	23.9 ± 0.02	23.7 ± 0.02	23.9 ± 0.10	24.2 ± 0.30	UAGN	3.47
66907	3 31 24.91	-27 38 51.50	> 27.1	24.0 ± 0.03	23.0 ± 0.02	22.5 ± 0.01	22.4 ± 0.09	22.3 ± 0.10	UAGN	3.45
6736	3 33 16.85	-28 01 05.20	26.9 ± 0.70	24.8 ± 0.05	24.5 ± 0.06	24.8 ± 0.07	25.0 ± 0.40	24.7 ± 0.60	OAGN	3.092
70686	3 32 04.05	-27 37 25.46	24.4 ± 0.05	24.1 ± 0.02	23.0 ± 0.01	23.3 ± 0.01	23.2 ± 0.06	22.2 ± 0.06	UAGN	0.97
11931	22 31 16.15	-60 54 06.55	27.5 ± 1.00	25.3 ± 0.10	24.4 ± 0.06	24.2 ± 0.05	23.7 ± 0.09	23.8 ± 0.30	UAGN	3.162
26289	22 32 18.53	-60 57 47.05	27.0 ± 0.70	24.7 ± 0.08	24.2 ± 0.06	24.2 ± 0.07	24.1 ± 0.20	24.3 ± 1.00	UAGN	3.42
40737	22 33 19.49	-60 43 03.86	23.2 ± 0.02	22.4 ± 0.01	21.9 ± 0.01	21.9 ± 0.01	21.9 ± 0.02	21.8 ± 0.04	UAGN	3.25
6791	10 31 02.71	5 12 32.17	25.9 ± 0.30	22.8 ± 0.01	22.0 ± 0.01	22.1 ± 0.01	22.0 ± 0.01	22.0 ± 0.04	UAGN	3.24
4989	10 31 31.73	5 11 46.30	> 27.1	24.9 ± 0.09	23.5 ± 0.02	23.1 ± 0.01	22.7 ± 0.02	22.3 ± 0.06	UAGN	3.58
13442	10 29 50.16	5 15 19.29	> 27.6	25.2 ± 0.09	23.7 ± 0.02	23.3 ± 0.02	23.1 ± 0.02	23.2 ± 0.10	UAGN	3.7
50398	10 29 30.09	5 31 42.86	> 27.3	26.1 ± 0.30	24.6 ± 0.06	24.7 ± 0.08	24.6 ± 0.10	25.4 ± 1.00	UAGN	3.494
32970	10 31 39.29	5 23 36.54	26.2 ± 0.40	24.0 ± 0.03	23.0 ± 0.01	22.7 ± 0.01	22.3 ± 0.01	22.1 ± 0.05	UAGN	3.4
66496	10 30 45.55	5 39 45.40	26.7 ± 0.50	24.8 ± 0.08	24.3 ± 0.05	24.2 ± 0.05	24.0 ± 0.06	23.9 ± 0.30	UAGN	2.23
53921	10 29 29.47	5 33 18.22	23.0 ± 0.02	22.1 ± 0.01	21.4 ± 0.01	21.3 ± 0.01	21.1 ± 0.01	20.7 ± 0.02	UAGN	3.51
6927	10 30 03.36	5 12 37.36	26.3 ± 0.40	24.2 ± 0.04	24.0 ± 0.03	23.7 ± 0.03	23.8 ± 0.05	23.4 ± 0.10	OAGN	2.975
46792	10 29 38.71	5 29 50.36	> 27.4	24.6 ± 0.07	23.3 ± 0.02	23.2 ± 0.02	23.0 ± 0.03	22.8 ± 0.10	UAGN	3.6
10796	10 30 13.63	5 14 15.89	27.1 ± 0.70	25.8 ± 0.20	25.1 ± 0.07	25.0 ± 0.08	24.8 ± 0.10	24.9 ± 0.60	OAGN	3.14
49934	12 56 27.55	1 05 26.81	26.0 ± 0.20	23.9 ± 0.02	23.2 ± 0.01	23.0 ± 0.01	23.1 ± 0.04	23.2 ± 0.08	UAGN	3.03
40795	12 56 08.04	0 59 40.66	23.6 ± 0.02	22.7 ± 0.01	21.7 ± 0.01	21.9 ± 0.01	21.7 ± 0.01	21.7 ± 0.02	UAGN	1.48
14729	12 55 08.50	1 07 37.29	27.9 ± 1.00	24.8 ± 0.05	24.0 ± 0.03	24.0 ± 0.03	24.0 ± 0.10	24.3 ± 0.20	UAGN	3.158
24189	12 55 30.43	1 10 08.43	26.2 ± 0.20	24.8 ± 0.04	24.5 ± 0.04	24.7 ± 0.06	24.8 ± 0.20	24.2 ± 0.20	UAGN	2.439
32977	12 55 50.66	0 59 18.78	24.1 ± 0.03	23.0 ± 0.01	22.2 ± 0.01	22.3 ± 0.01	22.3 ± 0.02	22.3 ± 0.04	UAGN	3.15
27146	12 55 37.27	1 19 17.37	> 28.1	26.0 ± 0.10	25.1 ± 0.07	24.7 ± 0.05	25.2 ± 0.20	24.7 ± 0.30	UAGN	3.14

[1]Magnitudes upper limits are 1-σ in the U and B bands, and 0.1-σ in the rest of the bands

[2]AGN classification in MUSYC spectroscopic database, UAGN=Unobscured AGN, OAGN=Obscured AGN. Separation based on fwhm of Lyα and C IV lines: sources with fwhm < 1000 Km/s restframe are classified as obscured and sources with broader lines as unobscured.

Table A.3: Spectroscopically confirmed LAEs

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	EW[2]	fwhm[3]	redshift
11982	3 31 50.36	-27 59 9.96	> 27.7	26.1 ± 0.10	26.0 ± 0.20	26.2 ± 0.20	25.9 ± 0.70	24.9 ± 0.70	6.1, 9.2, 17.0	2.1	3.123
13100	3 33 12.02	-27 58 39.86	> 27.8	26.8 ± 0.20	26.6 ± 0.30	26.2 ± 0.20	26.2 ± 0.80	25.0 ± 0.80	11.0, 15.0, 21.0	3.2	3.095
14746	3 32 47.52	-27 58 7.57	> 26.6	24.8 ± 0.10	24.4 ± 0.10	24.4 ± 0.10	24.7 ± 0.50	24.5 ± 1.00	22.0, 42.0, 81.0	1.6	3.088
17275	3 31 53.21	-27 57 7.92	> 27.5	26.2 ± 0.20	26.0 ± 0.20	26.1 ± 0.20	25.7 ± 0.60	> 27.9	4.9, 11.0, 19.0	1.7	3.115
17369	3 32 37.88	-27 57 7.02	26.9 ± 0.90	25.5 ± 0.10	25.2 ± 0.10	25.4 ± 0.10	25.3 ± 0.50	24.8 ± 0.90	5.9, 11.0, 23.0	1.4	3.09
30091	3 32 45.61	-27 52 10.96	> 27.8	27.1 ± 0.30	26.6 ± 0.30	26.5 ± 0.20	> 28.9	26.2 ± 2.00	9.7, 15.0, 23.0	2.0	3.107
32493	3 31 54.88	-27 51 20.99	> 27.7	25.6 ± 0.07	25.3 ± 0.08	25.3 ± 0.08	25.2 ± 0.40	27.1 ± 5.00	4.9, 15.0, 34.0	2.2	3.115
41056	3 32 52.69	-27 48 9.25	27.5 ± 0.90	25.2 ± 0.05	25.2 ± 0.07	25.5 ± 0.10	27.2 ± 2.00	> 27.9	18.0, 31.0, 46.0	2.0	3.085
41352	3 32 34.32	-27 47 59.46	27.3 ± 0.70	26.8 ± 0.20	26.0 ± 0.10	26.2 ± 0.20	27.3 ± 3.00	27.0 ± 4.00	6.3, 8.5, 25.0	2.2	3.1
42125	3 33 7.26	-27 47 46.50	25.3 ± 0.60	24.6 ± 0.20	24.6 ± 0.30	24.6 ± 0.30	24.5 ± 0.90	23.9 ± 1.00	19.0, 39.0, 84.0	2.0	3.085
44661	3 31 51.56	-27 46 47.14	> 27.8	26.7 ± 0.20	26.3 ± 0.20	26.9 ± 0.40	27.1 ± 3.00	> 27.9	13.0, 29.0, 69.0	1.9	3.108
45284	3 33 22.43	-27 46 36.55	> 26.9	24.9 ± 0.08	24.5 ± 0.09	24.7 ± 0.10	24.9 ± 0.60	24.7 ± 0.90	5.9, 9.3, 13.0	1.2	3.088
45857	3 32 39.02	-27 46 22.15	> 26.8	26.6 ± 0.40	25.9 ± 0.40	26.2 ± 0.40	26.8 ± 4.00	25.8 ± 3.00	9.8, 13.0, 16.0	1.7	3.085
46801	3 32 40.79	-27 46 5.99	> 27.4	25.3 ± 0.07	24.8 ± 0.06	25.0 ± 0.07	25.7 ± 0.70	25.7 ± 1.00	6.8, 8.8, 10.0	1.7	3.092
48935	3 31 52.83	-27 45 18.79	> 26.9	26.8 ± 0.50	26.0 ± 0.40	26.2 ± 0.50	26.7 ± 3.00	> 27.2	4.9, 7.1, 9.0	2.9	3.106
49643	3 31 47.80	-27 45 3.49	> 27.8	26.3 ± 0.10	26.1 ± 0.20	26.5 ± 0.30	26.2 ± 1.00	> 27.9	7.3, 16.0, 35.0	0.9	3.107
51126	3 33 32.68	-27 44 28.25	> 27.7	26.3 ± 0.10	25.8 ± 0.10	26.3 ± 0.20	> 28.9	25.2 ± 0.90	6.1, 9.5, 17.0	1.2	3.116
52127	3 32 15.76	-27 44 10.21	27.4 ± 1.00	26.4 ± 0.20	25.7 ± 0.20	25.9 ± 0.20	27.4 ± 3.00	25.5 ± 1.00	3.2, 6.8, 14.0	1.9	3.119
52768	3 33 16.90	-27 43 53.94	> 27.2	26.5 ± 0.30	26.6 ± 0.40	26.4 ± 0.40	> 28.5	> 27.7	11.0, 15.0, 84.0	2.0	3.086
53583	3 33 5.03	-27 43 37.27	> 27.7	26.4 ± 0.20	26.2 ± 0.20	26.3 ± 0.20	25.9 ± 0.70	> 27.9	15.0, 23.0, 41.0	1.6	3.087
56399	3 32 18.77	-27 42 47.56	> 24.0	24.4 ± 1.00	24.1 ± 1.00	23.9 ± 1.00	24.3 ± 6.00	23.6 ± 6.00	10.0, 21.0, 35.0	1.9	3.116
57373	3 33 27.02	-27 42 11.48	> 27.7	26.7 ± 0.20	26.3 ± 0.20	27.2 ± 0.50	26.3 ± 1.00	> 28.0	9.5, 12.0, 16.0	2.4	3.101
57848	3 33 33.66	-27 42 6.88	> 25.7	24.9 ± 0.30	24.9 ± 0.50	24.6 ± 0.40	24.5 ± 1.00	24.0 ± 1.00	4.9, 9.5, 13.0	1.5	3.088
58863	3 33 3.33	-27 41 38.72	> 27.7	27.1 ± 0.30	26.8 ± 0.30	27.1 ± 0.40	26.4 ± 1.00	> 28.0	6.8, 16.0, 39.0	0.7	3.103
59765	3 32 26.94	-27 41 27.56	25.0 ± 0.80	24.4 ± 0.30	24.2 ± 0.40	24.1 ± 0.40	23.8 ± 1.00	23.4 ± 1.00	3.2, 7.3, 15.0	1.0	3.113
60052	3 32 59.29	-27 41 14.57	> 26.0	26.1 ± 0.60	25.6 ± 0.60	25.9 ± 0.90	25.6 ± 2.00	> 26.5	7.0, 17.0, 38.0	1.0	3.12
62109	3 33 31.51	-27 40 32.38	> 26.7	24.9 ± 0.10	24.6 ± 0.10	24.6 ± 0.10	24.6 ± 0.40	24.8 ± 1.00	21.0, 29.0, 46.0	1.0	3.087
63537	3 32 46.77	-27 39 59.90	26.0 ± 1.00	26.0 ± 0.50	25.6 ± 0.60	25.7 ± 0.70	27.1 ± 9.00	25.8 ± 6.00	9.5, 27.0, 99.0	2.7	3.112
67519	3 33 18.90	-27 38 28.54	> 23.5	> 24.1	24.0 ± 2.00	24.2 ± 3.00	> 24.7	23.7 ± 8.00	13.0, 25.0, 53.0	3.2	3.106
70071	3 32 51.77	-27 37 33.46	> 27.7	26.1 ± 0.10	25.9 ± 0.10	26.2 ± 0.20	28.0 ± 5.00	25.2 ± 0.90	25.0, 43.0, 64.0	1.7	3.086
70528	3 33 8.15	-27 37 21.90	> 27.7	27.0 ± 0.30	26.4 ± 0.20	26.6 ± 0.30	> 28.9	> 28.0	5.8, 8.0, 11.0	0.7	3.113
73622	3 33 14.64	-27 36 13.90	> 27.7	27.0 ± 0.20	26.2 ± 0.20	26.2 ± 0.20	> 28.8	25.3 ± 0.90	14.0, 25.0, 47.0	2.4	3.107
77873	3 32 49.16	-27 34 39.90	26.8 ± 0.90	25.6 ± 0.10	25.2 ± 0.20	25.2 ± 0.20	25.6 ± 1.00	> 27.4	8.8, 12.0, 14.0	1.2	3.093

Continued on Next Page...

[1]Magnitudes upper limits are 1-σ in the U and B bands, and 0.1-σ in the rest of the bands

[2]Lyα restframe equivalent width, in Å. First and third numbers are the rough lower and upper 95% confidence intervals for the central value, dominated by the uncertainty in the placement of the continuum in the spectra.

[3]Restframe full width at half maximum of the Lyα line, in Å, calculated from a gaussian fit.

MUSYC ID	RA (J2000)	Dec. (J2000)	U mag[1]	B mag	V mag	R mag	I mag	z mag	EW[2]	fwhm[3]	redshift
79401	3 32 45.34	-27 34 3.29	> 23.1	> 23.6	24.3 ± 4.00	24.1 ± 4.00	24.0 ± 10.00	> 23.4	7.3, 15.0, 22.0	2.1	3.09
82344	3 33 13.52	-27 32 54.49	> 27.1	25.8 ± 0.20	25.4 ± 0.20	25.3 ± 0.20	25.1 ± 0.50	> 27.7	1.5, 1.9, 6.6	1.5	3.118
83999	3 33 17.71	-27 32 30.34	> 26.9	25.9 ± 0.20	25.5 ± 0.20	25.5 ± 0.20	> 28.2	> 27.6	5.3, 8.0, 20.0	3.4	3.13
19259	10 29 39.96	+5 17 42.89	> 27.5	26.6 ± 0.30	25.5 ± 0.10	25.6 ± 0.10	26.2 ± 0.40	25.7 ± 1.00	4.2, 19.0, 35.0	2.4	3.087
20883	10 30 24.87	+5 18 21.08	> 27.7	26.7 ± 0.40	25.6 ± 0.10	25.6 ± 0.10	25.4 ± 0.20	25.2 ± 0.70	5.3, 7.8, 14.0	2.7	3.122
21400	10 31 27.62	+5 18 35.38	> 27.2	27.0 ± 0.50	26.1 ± 0.20	26.7 ± 0.40	26.4 ± 0.50	27.8 ± 7.00	9.8, 12.0, 19.0	2.1	3.102
25597	10 29 19.99	+5 20 22.80	> 27.3	26.8 ± 0.50	25.9 ± 0.20	26.4 ± 0.40	26.6 ± 0.70	25.2 ± 1.00	8.0, 12.0, 33.0	1.9	3.124
28510	10 30 43.73	+5 21 41.25	> 27.5	26.8 ± 0.40	25.7 ± 0.10	25.8 ± 0.20	26.4 ± 0.50	> 28.2	29.0, 56.0,∞	3.1	3.097
29949	10 31 22.08	+5 22 16.68	> 27.2	26.2 ± 0.20	25.2 ± 0.08	25.0 ± 0.08	24.8 ± 0.10	25.3 ± 0.80	10.0, 26.0, 64.0	1.2	3.102
34124	10 31 30.31	+5 24 11.74	> 27.2	27.1 ± 0.50	26.1 ± 0.20	26.3 ± 0.20	26.3 ± 0.50	26.0 ± 1.00	5.3, 15.0, 19.0	1.2	3.115
34964	10 31 2.31	+5 24 32.88	> 27.5	25.7 ± 0.10	24.9 ± 0.06	25.0 ± 0.08	25.3 ± 0.20	24.8 ± 0.50	38.0, 49.0, 69.0	2.4	3.097
37685	10 30 45.39	+5 25 46.25	> 27.5	26.8 ± 0.40	25.9 ± 0.20	26.0 ± 0.20	26.2 ± 0.40	26.3 ± 2.00	7.6, 12.0, 20.0	0.5	3.099
37852	10 30 52.89	+5 25 49.49	> 27.5	25.9 ± 0.20	25.0 ± 0.07	24.6 ± 0.05	24.3 ± 0.07	24.0 ± 0.20	3.9, 6.1, 9.8	1.7	3.099
41045	10 30 18.69	+5 27 13.67	> 27.7	27.5 ± 1.00	26.3 ± 0.30	25.7 ± 0.20	25.5 ± 0.30	26.6 ± 4.00	4.9, 8.1, 10.0	3.1	3.086
42775	10 30 37.87	+5 27 59.41	> 27.5	26.5 ± 0.30	25.3 ± 0.09	25.1 ± 0.08	25.2 ± 0.20	25.0 ± 0.60	8.5, 14.0, 19.0	2.4	3.103
45693	10 30 38.26	+5 29 26.43	> 27.5	26.5 ± 0.30	26.0 ± 0.20	25.8 ± 0.20	25.9 ± 0.30	24.9 ± 0.50	7.3, 12.0, 17.0	1.2	3.102
45706	10 31 6.14	+5 29 24.96	> 27.5	26.1 ± 0.20	25.3 ± 0.09	25.3 ± 0.10	25.2 ± 0.20	24.4 ± 0.30	19.0, 39.0, 97.0	–	3.119
49048	10 30 21.34	+5 31 3.68	> 27.7	25.6 ± 0.20	25.1 ± 0.10	25.1 ± 0.10	25.5 ± 0.20	24.5 ± 0.50	12.0, 24.0, 33.0	3.2	3.106
52333	10 31 39.43	+5 32 43.06	> 27.3	26.8 ± 0.40	26.0 ± 0.20	26.1 ± 0.20	28.1 ± 3.00	> 28.0	10.0, 18.0, 39.0	4.6	3.111
55827	10 31 11.57	+5 34 21.95	> 27.4	25.6 ± 0.10	24.9 ± 0.06	24.8 ± 0.06	24.8 ± 0.10	25.1 ± 0.60	22.0, 35.0, 61.0	1.9	3.106
60592	10 31 9.41	+5 36 41.22	> 27.5	27.2 ± 0.60	26.3 ± 0.20	26.6 ± 0.30	> 29.8	> 28.2	17.0, 30.0, 76.0	1.9	3.108
14186	12 55 7.46	+1 10 43.29	> 28.0	26.3 ± 0.20	25.9 ± 0.20	25.8 ± 0.20	25.6 ± 0.40	25.0 ± 0.50	12.0, 17.0, 24.0	2.0	3.086
19674	12 55 19.97	+1 8 37.82	27.1 ± 0.40	25.9 ± 0.10	25.6 ± 0.10	25.6 ± 0.10	25.9 ± 0.50	27.7 ± 5.00	32.0, 54.0, 76.0	1.7	3.085
23313	12 55 28.61	+1 17 0.21	27.2 ± 0.50	26.0 ± 0.10	25.6 ± 0.10	25.4 ± 0.10	25.6 ± 0.40	26.7 ± 2.00	9.0, 14.0, 19.0	1.5	3.109
26159	12 55 35.37	+1 0 38.85	28.0 ± 0.90	26.9 ± 0.30	26.0 ± 0.20	26.2 ± 0.20	25.7 ± 0.40	24.6 ± 0.30	4.9, 7.8, 13.0	1.5	3.101
29383	12 55 42.82	+0 56 36.89	> 27.7	26.9 ± 0.40	26.0 ± 0.20	25.8 ± 0.20	25.5 ± 0.50	25.2 ± 0.70	9.3, 12.0, 16.0	1.7	3.096
31617	12 55 48.07	+0 59 21.20	> 28.1	26.3 ± 0.20	25.9 ± 0.10	26.0 ± 0.20	25.6 ± 0.40	25.6 ± 0.70	7.1, 10.0, 18.0	2.0	3.092
34250	12 55 54.12	+0 56 46.39	> 28.1	26.5 ± 0.20	26.1 ± 0.20	26.0 ± 0.20	26.4 ± 0.80	24.6 ± 0.30	9.3, 17.0, 37.0	1.0	3.097

References

Adelberger, K. L., Erb, D. K., Steidel, C. C., Reddy, N. A., Pettini, M., & Shapley, A. E. 2005a, ApJ, 620, L75

Adelberger, K. L. & Steidel, C. C. 2005a, ApJ, 630, 50

—. 2005b, ApJ, 627, L1

Adelberger, K. L., Steidel, C. C., Giavalisco, M., Dickinson, M., Pettini, M., & Kellogg, M. 1998, ApJ, 505, 18

Adelberger, K. L., Steidel, C. C., Pettini, M., Shapley, A. E., Reddy, N. A., & Erb, D. K. 2005b, ApJ, 619, 697

Adelberger, K. L., Steidel, C. C., Shapley, A. E., Hunt, M. P., Erb, D. K., Reddy, N. A., & Pettini, M. 2004, ApJ, 607, 226

Aird, J., Nandra, K., Georgakakis, A., Laird, E. S., Steidel, C. C., & Sharon, C. 2008, MNRAS, 387, 883

Alexander, D. M., Bauer, F. E., Brandt, W. N., Schneider, D. P., Hornschemeier, A. E., Vignali, C., Barger, A. J., Broos, P. S., Cowie, L. L., Garmire, G. P., Townsley, L. K., Bautz, M. W., Chartas, G., & Sargent, W. L. W. 2003, AJ, 126, 539

Bahcall, N. A., Dong, F., Hao, L., Bode, P., Annis, J., Gunn, J. E., & Schneider, D. P. 2003, ApJ, 599, 814

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Basilakos, S., Georgakakis, A., Plionis, M., & Georgantopoulos, I. 2004, ApJ, 607, L79

Basilakos, S., Plionis, M., Georgakakis, A., & Georgantopoulos, I. 2005, MNRAS, 356, 183

Blanc, G. A., Lira, P., Barrientos, L. F., Aguirre, P., Francke, H., Taylor, E. N., Quadri, R., Marchesini, D., Infante, L., Gawiser, E., Hall, P. B., Willis, J. P., Herrera, D., & Maza, J. 2008, ApJ, 681, 1099

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Brusa, M., Comastri, A., Daddi, E., Pozzetti, L., Zamorani, G., Vignali, C., Cimatti, A., Fiore, F., Mignoli, M., Ciliegi, P., & Röttgering, H. J. A. 2005, A&A, 432, 69

Burstein, D. & Heiles, C. 1978, ApJ, 225, 40

Cardamone, C. N., Urry, C. M., Damen, M., van Dokkum, P., Treister, E., Labbé, I., Virani, S. N., Lira, P., & Gawiser, E. 2008, ApJ, 680, 130

Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499

Chandrasekhar, S. 1943, Reviews of Modern Physics, 15, 1

Coil, A. L., Hennawi, J. F., Newman, J. A., Cooper, M. C., & Davis, M. 2007, ApJ, 654, 115

Coil, A. L., Newman, J. A., Croton, D., Cooper, M. C., Davis, M., Faber, S. M., Gerke, B. F., Koo, D. C., Padmanabhan, N., Wechsler, R. H., & Weiner, B. J. 2008, ApJ, 672, 153

Collins, C. A., Heydon-Dumbleton, N. H., & MacGillivray, H. T. 1989, MNRAS, 236, 7P

Collins, C. A., Nichol, R. C., & Lumsden, S. L. 1992, MNRAS, 254, 295

Cowie, L. L. & Hu, E. M. 1998, AJ, 115, 1319

Croft, R. A. C., Dalton, G. B., & Efstathiou, G. 1999, MNRAS, 305, 547

Croom, S. M., Boyle, B. J., Shanks, T., Smith, R. J., Miller, L., Outram, P. J., Loaring, N. S., Hoyle, F., & da Angela, J. 2005, MNRAS, 356, 415

Davis, M. & Peebles, P. J. E. 1983, ApJ, 267, 465

Dodelson, S. 2003, Modern cosmology (Modern cosmology / Scott Dodelson. Amsterdam (Netherlands): Academic Press. ISBN 0-12-219141-2, 2003, XIII + 440 p.)

Ferrarese, L. 2002, ApJ, 578, 90

Franx, M., Illingworth, G. D., Kelson, D. D., van Dokkum, P. G., & Tran, K.-V. 1997, ApJ, 486, L75+

Gawiser, E., Francke, H., Lai, K., Schawinski, K., Gronwall, C., Ciardullo, R., Quadri, R., Orsi, A., Barrientos, L. F., Blanc, G. A., Fazio, G., Feldmeier, J. J., Huang, J.-s., Infante, L., Lira, P., Padilla, N., Taylor, E. N., Treister, E., Urry, C. M., van Dokkum, P. G., & Virani, S. N. 2007, ApJ, 671, 278

Gawiser, E. et al. 2006a, ApJS, 162, 1

—. 2006b, ApJ, 642, L13

Gebhardt, K. et al. 2000, ApJ, 539, L13

Giacconi, R. et al. 2002, ApJS, 139, 369

Giavalisco, M. & Dickinson, M. 2001, ApJ, 550, 177

Giavalisco, M., Steidel, C. C., Adelberger, K. L., Dickinson, M. E., Pettini, M., & Kellogg, M. 1998, ApJ, 503, 543

Giavalisco, M., Steidel, C. C., & Szalay, A. S. 1994, ApJ, 425, L5

Gilli, R., Daddi, E., Zamorani, G., Tozzi, P., Borgani, S., Bergeron, J., Giacconi, R., Hasinger, G., Mainieri, V., Norman, C., Rosati, P., Szokoly, G., & Zheng, W. 2005, A&A, 430, 811

Gronwall, C., Ciardullo, R., Hickey, T., Gawiser, E., Feldmeier, J. J., van Dokkum, P. G., Urry, C. M., Herrera, D., Lehmer, B. D., Infante, L., Orsi, A., Marchesini, D., Blanc, G. A., Francke, H., Lira, P., & Treister, E. 2007, ApJ, 667, 79

Groth, E. J. & Peebles, P. J. E. 1977, ApJ, 217, 385

—. 1986a, ApJ, 310, 507

—. 1986b, ApJ, 310, 499

Hamilton, A. J. S. 1993, ApJ, 417, 19

Hayashino, T. et al. 2004, AJ, 128, 2073

Hildebrandt, H., Pielorz, J., Erben, T., Schneider, P., Eifler, T., Simon, P., & Dietrich, J. P. 2007, A&A, 462, 865

Hopkins, P. F., Lidz, A., Hernquist, L., Coil, A. L., Myers, A. D., & Cox, Thomas J.and Spergel, D. N. 2007, ApJ, 662, 110

Hu, E. M., Cowie, L. L., & McMahon, R. G. 1998, ApJ, 502, L99+

Hu, E. M. & McMahon, R. G. 1996, Nature, 382, 231

Hubble, E. 1934, ApJ, 79, 8

Huchra, J. P., Geller, M. J., de Lapparent, V., & Corwin, Jr., H. G. 1990, ApJS, 72, 433

Infante, L. 1994, A&A, 282, 353

Infante, L., Varela, J., Moles, M., Hertling, G., García, A., & Menanteau, F. 2003, ApJ, 588, 90

Kaiser, N. 1984, ApJ, 284, L9

Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999, MNRAS, 303, 188

Kauffmann, G. & Haehnelt, M. 2002, MNRAS, 332, 529

Kovač, K., Somerville, R. S., Rhoads, J. E., Malhotra, S., & Wang, J. 2007, ApJ, 668, 15

Kudritzki, R.-P., Méndez, R. H., Feldmeier, J. J., Ciardullo, R., Jacoby, G. H., Freeman, K. C., Arnaboldi, M., Capaccioli, M., Gerhard, O., & Ford, H. C. 2000, ApJ, 536, 19

Lacey, C. & Cole, S. 1993, MNRAS, 262, 627

Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64

Lee, K.-S., Giavalisco, M., Gnedin, O. Y., Somerville, R. S., Ferguson, H. C., Dickinson, M., & Ouchi, M. 2006, ApJ, 642, 63

Lehmer, B. D. et al. 2005, ApJS, 161, 21

Lidz, A., Hopkins, P. F., Cox, T. J., Hernquist, L., & Robertson, B. 2006, ApJ, 641, 41

Limber, D. N. 1953, ApJ, 117, 134

—. 1954, ApJ, 119, 655

Lin, H., Kirshner, R. P., Shectman, S. A., Landy, S. D., Oemler, A., Tucker, D. L., & Schechter, P. L. 1996, ApJ, 471, 617

Lupton, R. 1993, Statistics in theory and practice (Princeton, N.J.: Princeton University Press, —c1993)

Madau, P. 1995, ApJ, 441, 18

Maddox, S. J., Efstathiou, G., & Sutherland, W. J. 1996, MNRAS, 283, 1227

Magorrian, J. et al. 1998, AJ, 115, 2285

McCarthy, P. J. 1993, ARA&A, 31, 639

Miyaji, T., Zamorani, G., Cappelluti, N., Gilli, R., Griffiths, R. E., Comastri, A., Hasinger, G., Brusa, M., Fiore, F., Puccetti, S., Guzzo, L., & Finoguenov, A. 2007, ApJS, 172, 396

Mo, H. J. & White, S. D. M. 1996, MNRAS, 282, 347

Mullis, C. R., Henry, J. P., Gioia, I. M., Böhringer, H., Briel, U. G., Voges, W., & Huchra, J. P. 2004, ApJ, 617, 192

Myers, A. D., Brunner, R. J., Richards, G. T., Nichol, R. C., Schneider, D. P., Vanden Berk, D. E., Scranton, R., Gray, A. G., & Brinkmann, J. 2006, ApJ, 638, 622

Nandra, K., Laird, E. S., & Steidel, C. C. 2005, MNRAS, 360, L39

Ouchi, M., Hamana, T., Shimasaku, K., Yamada, T., Akiyama, M., Kashikawa, N., Yoshida, M., Aoki, K., Iye, M., Saito, T., Sasaki, T., Simpson, C., & Yoshida, M. 2005, ApJ, 635, L117

Ouchi, M., Shimasaku, K., Okamura, S., Doi, M., Furusawa, H., Hamabe, M., Kimura, M., Komiyama, Y., Miyazaki, M., Miyazaki, S., Nakata, F., Sekiguchi, M., Yagi, M., & Yasuda, N. 2001, ApJ, 558, L83

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., & Yasuda, N. 2004, ApJ, 611, 685

Ouchi, M. et al. 2003, ApJ, 582, 60

Partridge, R. B. & Peebles, P. J. E. 1967, ApJ, 147, 868

Peebles, P. J. E. 1967, ApJ, 147, 859

—. 1980, The large-scale structure of the universe (Research supported by the National Science Foundation. Princeton, N.J., Princeton University Press, 1980. 435 p.)

Peebles, P. J. E. & Hauser, M. G. 1974, ApJS, 28, 19

Press, W. H. & Schechter, P. 1974, ApJ, 187, 425

Quadri, R., van Dokkum, P., Gawiser, E., Franx, M., Marchesini, D., Lira, P., Rudnick, G., Herrera, D., Maza, J., Kriek, M., Labbé, I., & Francke, H. 2007, ApJ, 654, 138

Reddy, N. A., Erb, D. K., Steidel, C. C., Shapley, A. E., Adelberger, K. L., & Pettini, M. 2005, ApJ, 633, 748

Shapley, A. E., Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., & Pettini, M. 2001, ApJ, 562, 95

Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65

Shectman, S. A., Landy, S. D., Oemler, A., Tucker, D. L., Lin, H., Kirshner, R. P., & Schechter, P. L. 1996, ApJ, 470, 172

Shen, Y. et al. 2007, AJ, 133, 2222

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

Sheth, R. K. & Tormen, G. 1999, MNRAS, 308, 119

—. 2002, MNRAS, 329, 61

Shimasaku, K. et al. 2004, ApJ, 605, L93

Simon, P. 2007, A&A, 473, 711

Spergel, D. N. e. a. 2007, ApJS, 170, 377

Steidel, C. C., Adelberger, K. L., Dickinson, M., Giavalisco, M., Pettini, M., & Kellogg, M. 1998, ApJ, 492, 428

Steidel, C. C., Adelberger, K. L., Shapley, A. E., Pettini, M., Dickinson, M., & Giavalisco, M. 2000, ApJ, 532, 170

—. 2003, ApJ, 592, 728

Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996a, AJ, 112, 352

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996b, ApJ, 462, L17

Steidel, C. C., Hunt, M. P., Shapley, A. E., Adelberger, K. L., Pettini, M., Dickinson, M., & Giavalisco, M. 2002, ApJ, 576, 653

Szokoly, G. P. et al. 2004, ApJS, 155, 271

Treister, E. & Urry, C. M. 2005, ApJ, 630, 115

Treister, E. et al. 2004, ApJ, 616, 123

Virani, S. N., Treister, E., Urry, C. M., & Gawiser, E. J. 2006, AJ, 131, 2373

Yang, Y., Mushotzky, R. F., Barger, A. J., & Cowie, L. L. 2006, ApJ, 645, 68

Zehavi, I., Zheng, Z., Weinberg, D. H., Frieman, J. A., Berlind, A. A., Blanton, M. R., Scoccimarro, R., Sheth, R. K., Strauss, M. A., Kayo, I., Suto, Y., Fukugita, M., Nakamura, O., Bahcall, N. A., Brinkmann, J., Gunn, J. E., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Loveday, J., Meiksin, A., Schlegel, D. J., Schneider, D. P., Szapudi, I., Tegmark, M., Vogeley, M. S., & York, D. G. 2005, ApJ, 630, 1

Zehavi, I. et al. 2002, ApJ, 571, 172