PROBING THE LARGE SCALE STRUCTURE OF THE UNIVERSE THROUGH LUMINOUS RED GALAXIES

by

Abhishek Prakash

B.Sc., IGNOU, 2007

M.Sc., University of Hyderabad, 2009

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

DEPARTMENT OF PHYSICS AND ASTRONOMY

This dissertation was presented

by

Abhishek Prakash

It was defended on

July 18th 2017

and approved by

Jeffrey A. Newman, Prof., Physics & Astronomy, University of Pittsburgh

Sandhya Rao, Research Prof., Physics & Astronomy, University of Pittsburgh

Michael Wood-Vasey, Assoc. Prof., Physics & Astronomy, University of Pittsburgh

Shirley Ho, Assoc. Prof., Physics & Astronomy, Carnegie Mellon University

Andrew R. Zentner, Assoc. Prof., Physics & Astronomy, University of Pittsburgh

James Mueller, Assoc. Prof., Physics & Astronomy, University of Pittsburgh

Dissertation Director: Jeffrey A. Newman, Prof., Physics & Astronomy, University of Pittsburgh

PROBING THE LARGE SCALE STRUCTURE OF THE UNIVERSE THROUGH LUMINOUS RED GALAXIES

Abhishek Prakash, PhD

University of Pittsburgh, 2017

The study of Baryon Acoustic Oscillations (BAO) has become a key tool for exploring the nature of Dark Energy. BAO refers to periodic fluctuations in the density of baryonic matter in the universe that were caused by acoustic oscillations created by radiation pressure differences. Luminous Red Galaxies (LRGs) have played a vital role in the detection of BAO. These are the most massive galaxies in the z$\sim$1 universe, showing a characteristic 4000Å break in their spectral energy distributions (SEDs). This thesis establishes a new algorithm for selecting high redshift LRGs for spectroscopic surveys like SDSS-IV/eBOSS and DESI, which aim to precisely measure the BAO signal at high redshifts. We further adapt these methods for assembling the next generation eBOSS LRG sample and explain these methods in detail. These methods have been used by other international teams to assemble high redshift LRG sample for various investigations.

Large-scale clustering measurements are prone to systematic uncertainties associated with imaging surveys. This thesis establishes modern statistical techniques to overcome these challenges. A multivariate regression analysis, primarily applied on the eBOSS LRG and quasar samples, is presented for a better understanding of the systematic uncertainties and their effects on the clustering measurements. Another key aspect of this work is the development of a machine learning algorithm for estimating photometric redshifts. Applying this machinery yields accurate estimates of the distances of galaxies using their photometric properties alone, i.e., their colors and magnitudes. This work culminates with new clustering measurements that use these photometric redshifts in an attempt to detect BAO in the

eBOSS LRG sample. We also presented initial results from an ongoing effort to measure the clustering of LRGs around quasars using the first two years of spectroscopic data from eBOSS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Nearly a century ago, in 1929, Edwin Hubble established that our universe is expanding (Hubble, 1929). Following that discovery, the field of cosmology has grown considerably, adding to our knowledge of the universe in which we live. For many subsequent decades, physicists believed that the expansion of the universe should be slowing with time. This was motivated by the understanding that if the universe contains just matter and radiation, then the gravitational attraction between matter should slow down this expansion. In 1998, measurements of distances to Type Ia supernovae (SN Ia) surprised the whole scientific community by establishing that the expansion of the universe is, in fact, accelerating (Riess et al., 1998; Perlmutter et al., 1999). This was also demonstrated by the constraints on the flatness of the universe by Cosmic Microwave Background (CMB) data. This discovery has inspired massive efforts to probe the expansion history of the universe through observational techniques. The origin of the accelerating universe continues to be the most important unknown in physics.

1.1 AN UNKNOWN ENERGY OR MANIFESTATION OF GRAVITY?

Many theories have been proposed to explain the accelerated expansion of the universe. However, the two widely accepted plausible theories are 1) the presence of an unknown component of the energy density of the universe, commonly referred to as "dark energy" or 2) the breakdown of Einstein's General Relativity (GR) at large scales. It is possible to decouple scenarios of acceleration that require dark energy from those that require modifications to GR by independently probing both the cosmic expansion history and the growth rate of

the structure. For this thesis, I focus on scenarios that require the presence of dark energy. Recent measurements of the CMB by the *Planck* satellite (Planck Collaboration et al., 2014) support a picture where the acceleration is driven by "dark energy" in a spatially flat universe. Although (if we discard the possibility of deviations from GR) evidence for the presence of dark energy is conclusive, its nature is not understood.

Many theoretical explanations for dark energy have been put forward. The simplest explanation for dark energy is that it is an intrinsic, fundamental energy of space. This is also referred to as "vacuum energy" because it is the energy density of an empty vacuum (Weinberg, 1987). This contribution to the energy density of the universe could be explained by Einstein's cosmological constant (Λ), introduced into his General Theory of Relativity (GTR) to counteract the attractive force of gravity. Another popular theory proposes that the accelerated expansion is driven by the potential energy of a smooth dynamical scalar field, commonly referred to as "quintessence" (Peebles & Ratra, 1988). This field must be very sparse (not dense) in order to not clump and form structure like matter. Unlike the cosmological constant, a scalar field can vary in space-time.

1.2 EFFORTS TOWARDS UNDERSTANDING DARK ENERGY

Dark energy may be characterized by the equation of state $w = P/\rho$, where P is the pressure and ρ is mass density of the dark energy when treated as a perfect fluid. The parameter w, usually called the "equation of state parameter", is a dimensionless number. Evolution of w with time may be described by a second parameter, $w_a = dw/da(t)$, where $a(t)$ is the scale factor describing the relative expansion of the universe (Johri & Rath, 2007). One of the central goals of many past, current, and future cosmological surveys is determining constraints on these dark energy parameters (e.g., the Sloan Digital Sky Survey (SDSS) II, III/Baryon Oscillation Spectroscopic survey (BOSS), IV/ extended Baryon Oscillation Spectroscopic survey (eBOSS), the Dark Energy Survey (DES), and the Dark Energy Spectroscopic Instrument (DESI); York et al. (2000a); Eisenstein et al. (2001); Cannon et al. (2006); Dawson et al. (2013a)). Astronomers can distinguish different models for dark energy based on the

values of these parameters. For example, an equation of state where $w = -1$ with $w_a = 0$ would be equivalent to the cosmological constant model. On the other hand, significant deviations from these values would rule out a cosmological constant.

A measurement of the expansion of the universe as a function of redshift probes the energy content of the universe and can be used to study the properties of the dark energy. Currently, four primary observational techniques are accepted as the most powerful toward obtaining this goal (e.g., Albrecht et al. 2006): SN Ia, weak lensing, galaxy clusters, and baryon acoustic oscillations (BAO). This thesis is focused on the BAO probe. Spectroscopy of galaxies and quasars provides a 3D map of the large-scale structure of the universe in which the BAO feature is embedded.

1.2.1 What are Baryon Acoustic Oscillations (BAO)?

The term BAO refers to periodic fluctuations in the density of the visible baryonic matter of the universe caused by acoustic oscillations due to differences in radiation pressure in the early universe (analogous to sound waves created in air by pressure differences). These pressure differences arise due to density perturbations in the early universe and the interaction between radiation and matter. After the big bang, the early universe was hot and opaque which caused photons and baryons to move outwards together. This expansion continued for $\sim 10^5$ years. As the universe expanded and cooled, the electrons and baryons combined to form neutral atoms, making the universe essentially transparent to photons. This phase of the universe occurred approximately 400,000 years after the big bang and is referred to as "recombination." The photons, no longer interacting strongly with the baryonic matter, diffused away to be observed as the CMB. This ceased the outward pressure on the system; for a perfectly isolated, compact density fluctuation, the consequence would be a shell of baryonic matter at a fixed radius. This radius is often referred to as the "sound horizon." It is expected that these slight under-densities and over-densities of baryons were frozen into place at the moment of recombination and should be imprinted on the galaxy distribution today.

The BAO feature may be used to compare the apparent size of the length scale today

(and at different redshifts or different times) to the sound horizon at the time of recombination. These sizes may be determined using galaxy clustering (closer to the present epoch) or the CMB (for the sound horizon at recombination). Thus, the BAO provides a standard ruler that can be used to understand the accelerated expansion of the universe completely independent of other methods.

Using BAO to measure the expansion of the universe is a well-developed field. BAO signals have been detected and measured to ever-greater precision using data from a number of large galaxy surveys including the SDSS-I and -II (York et al., 2000a; Abazajian et al., 2009), SDSS-III/BOSS (Dawson et al., 2013a), Two-degree-Field Galaxy Redshift Survey (2dFGRS, Croom et al., 2005), and WiggleZ (Blake et al., 2011).

1.2.1.1 BAO using Luminous Red Galaxies (LRGs)

A precise measurement of the BAO scale requires large ensembles of galaxies with their redshifts measured accurately using wide-field spectroscopy spanning a large volume of the universe. This presents a great challenge and requires careful planning in assembling uniformly-distributed galaxy or quasar samples. This is due to the fact that the BAO signal is very weak, so clustering of samples due to systematics on BAO-like scales can obscure the signal. Traditionally, massive galaxies are used for BAO studies, since the detection of BAO requires measuring the clustering of matter (or galaxies). Dark matter halos are the matter over-densities in which galaxy formation takes place. Their presence is inferred from their gravitational effects on galaxy rotation curves. Massive galaxies, and in particular red, elliptical galaxies, tend to reside in massive dark matter halos and cluster strongly, which in turn enhances the BAO signal (e.g., Postman & Geller, 1984; Kauffmann et al., 2004). The most luminous and reddest (in rest-frame color) galaxies are typically referred to as 'Luminous Red Galaxies' or LRGs. LRGs populate a narrow range of rest-frame colors (Postman & Lauer, 1995). Given both their bright intrinsic luminosities (allowing them to be studied to higher redshifts than typical L^* galaxies) and their strong clustering, LRGs are excellent tracers of the large-scale structure of the universe.

LRGs have been previously used to study large-scale structure by a variety of investi-

gations, most notably the SDSS (York et al., 2000a) and the *SDSS-III*/BOSS, as well as the 2dF-SDSS LRG and QSO survey (e.g. Eisenstein et al., 2001; Cannon et al., 2006). Due to this, LRGs are currently the best-understood sample for large scale structure studies. It is worth noting that the BAO signature was first detected in the LRG sample of SDSS-I and II (Eisenstein et al., 2005). Increasing our current sample of LRGs to higher redshifts will allow measurements of the BAO feature, and hence of the expansion rate of the universe (Seo & Eisenstein, 2003; Lin & Mohr, 2003; Ross et al., 2008), during the era when accelerated expansion began.

1.3 THE SLOAN DIGITAL SKY SURVEY (SDSS)

For almost 20 years, the SDSS (York et al., 2000a) has provided the largest spectroscopic samples for cosmological analysis using a 2.5-meter telescope at Apache Point Observatory (APO). The first two generations of SDSS, SDSS-I and -II, measured the redshifts of nearly one million galaxies (Abazajian et al., 2009). This led to the first detection of the BAO scale (Eisenstein et al., 2005) which is one of the signature accomplishments of SDSS. BOSS (Dawson et al., 2013a) performed spectroscopic observations of more than 1.5 million galaxies and more than 150,000 quasars. Measurements of BAO with BOSS have led to 1%-2% precision measurements of the cosmological distance scale for redshifts $z < 0.6$ and $z \cong 2.5$. Building on the success of BOSS, SDSS-IV/eBOSS is performing a spectroscopic survey of galaxies and quasars at redshifts uncharted by BOSS. The work presented in this thesis has played a central role in building the eBOSS survey, which is explained in detail in the following chapters.

1.3.1 SDSS-IV/extended Baryon Oscillation Spectroscopic Survey (eBOSS)

eBOSS concentrates its efforts on the observation of galaxies and quasars, at redshift $0.6 < z < 2.2$, a time interval currently left relatively unexplored by other three-dimensional maps of the large-scale structure of the universe. In filling this gap, we will create the

largest volume survey of the universe to date. eBOSS is using four different tracers of the underlying matter density field to map large-scale structure. 1) Using more than 350,000 new, spectroscopically confirmed LRGs at a median redshift $z \sim 0.71$, we project that eBOSS will yield measurements of the angular diameter distance, $d_A(z)$, to an accuracy of 1.2% and measurements of Hubble parameter, $H(z)$, to 2.1% when combined with the $z < 0.6$ sample of BOSS galaxies. 2) With ∼195,000 new emission line galaxy (ELG) redshifts, we expect BAO measurements of $d_A(z)$ to an accuracy of 3.1% and H(z) to 4.7% at an effective redshift of $z \sim 0.87$. 3) A sample of more than 500,000 spectroscopically confirmed quasars will provide the first BAO distance measurements over the redshift range $0.9 < z < 2.2$, with expected precision of 2.8% and 4.2% on $d_A(z)$ and $H(z)$, respectively. This is the first instance of quasars being targeted as direct tracers of the underlying dark matter distribution to measure the BAO scale. 4) Finally, with 60,000 new quasars and the re-observation of 60,000 BOSS quasars, we will obtain new Ly-α forest measurements at redshifts $z > 2.1$; these new data will enhance the precision of $d_A(z)$ and $H(z)$ at $z > 2.1$ by a factor of 1.44 relative to SDSS-III/BOSS.

1.3.2 Dark Energy Spectroscopic Instrument (DESI)

DESI is a ground-based dark energy survey that will study BAO and the growth of structure with a wide-area galaxy and quasar redshift survey. Although DESI is a much larger survey and will map the universe at much higher density, it is very similar to eBOSS in survey design. eBOSS target selection methods are based on tests for DESI target selection. DESI will be used to conduct a five-year survey designed to cover 14,000 deg^2 of the sky. DESI aims to measure $d_A(z)$ and $H(z)$ with sub-percent level precision. As in eBOSS, spectroscopic targets will be selected in four classes from imaging data: 1) LRGs up to $z = 1.0$, extending the BOSS LRG survey in both redshift and survey area; 2) ELGs up to $z = 1.7$; 3) quasars as direct tracers of the underlying dark matter distribution, and, 4) at higher redshifts ($2.1 < z < 3.5$), the Ly-α forest absorption features in QSO spectra, which will be used to trace the distribution of neutral hydrogen. In addition, when moonlight prevents efficient observations, usually called "bright time," DESI will conduct a magnitude-limited Bright

Galaxy Survey (BGS) comprising approximately 10 million galaxies with median $z \sim 0.2$. Overall, ∼30 million galaxy and quasar redshifts will be obtained to measure the BAO feature, probe whether GR is accurate using redshift space distortion (Jennings et al., 2012), and constrain theories of galaxy evolution, among other science goals.

1.4 DISSERTATION OVERVIEW

The chapters of this thesis are designed as a series of steps toward the goal of using the BAO technique to study the expansion history of the universe and are relatively self-contained. This allows readers with various backgrounds to focus in on certain chapters or sections according to their level of interest or expertise.

In Chapter 2, I establish a new method of combining optical and infrared photometry to select luminous red galaxies (LRGs) at redshifts $z > 0.6$. I also demonstrate the need for this new method by showing the shortcomings of previously-used methods. I use a combination of optical photometry from the Canada France Hawaii Telescope Legacy Survey (CFHTLS) and the *Hubble Space Telescope* (HST), infrared photometry from the *Wide-Field Infrared Survey Explorer* (WISE) satellite, and spectroscopic or photometric redshifts from the DEEP2 Galaxy Redshift Survey and COSMOS. I present a variety of methods for testing the success of this selection and present methods for optimization given a set of rest-frame color and redshift requirements. I present tests of this selection in two different regions of the sky, the COSMOS and Extended Groth Strip (EGS) fields, to reduce the effect of cosmic/sample variance. I have used these methods to assemble large samples of LRGs for two different ancillary programs as a part of the SDSS-III/BOSS. I have shown that these methods can select high-redshift LRGs efficiently with minimal stellar contamination; this is extremely difficult to achieve with selections that rely on optical photometry alone.

In Chapter 3, I present the application of the methods developed in Chapter 2 for selecting ∼600,000 LRG targets for SDSS-IV/eBOSS, which began observations in Fall 2014. In a six-year program, eBOSS will provide us with secure and accurate redshift information for a sample of ∼375,000 LRGs over ∼7,500 deg^2 of the sky. This will allow us to measure

the BAO feature with $\sim 1\%$ precision in the redshift range $0.6 < z < 1.0$. I present results from an SDSS-III/BOSS ancillary program, SEQUELS, which was a survey validation effort for SDSS-IV/eBOSS. We perform a number of tests using spectroscopic data from SDSS-III/BOSS ancillary programs to determine the redshift reliability of our target selection and its ability to meet the science requirements of eBOSS. The SDSS spectra are of high enough signal-to-noise ratio that at least $\sim 89\%$ of the target sample yields secure redshift measurements. We also present tests of the uniformity and homogeneity of the sample, demonstrating that it should be clean enough for studies of the large-scale structure of the universe at higher redshifts than the SDSS-III/BOSS LRG sample.

In Chapter 4, I present a similar analysis for the SDSS-IV/eBOSS quasar sample, which will probe the large scale structure of the universe at a much higher redshift range, $0.8 < z < 2.2$. This will be the first study in which quasars are being targeted as direct tracers of the underlying matter density on BAO scales. The primary limitation of this approach is the sparse number density of quasars. eBOSS overcomes this difficulty by mapping an enormous volume over a broad redshift range. I discuss some characteristic features of this sample. Finally, I present the details of a uniformity and homogeneity analysis of the quasar sample, similar to the analysis of LRGs in Chapter 3, and establish the mask for large-scale structure measurements with the eBOSS survey.

In Chapter 5, I present an analysis for detecting the BAO signal within the eBOSS LRG sample using photometric redshifts. Traditionally, spectroscopic redshifts are used to estimate distances to galaxies and, in turn, to measure galaxy clustering. However, acquiring spectroscopic redshifts is a time consuming and expensive process. Large multi-band imaging surveys have made it possible to estimate more accurate photometric redshifts than was possible before. I present a machine-learning algorithm that can make improved estimations of photometric redshifts once augmented with a subset of spectroscopic redshift training samples. Although photometric redshifts are less accurate, they are significantly easier to obtain, and for a given amount of time, one can image wider areas to fainter limits than can be reached with spectroscopy, allowing one to probe both larger scales and larger volumes. This is possible as photometric redshift estimates depend on the strong features in the spectra of astronomical objects that can be detected by broad-band filters of large

imaging surveys.

In Chapter 6, I present an analysis of the cross-correlations of eBOSS quasars with eBOSS LRGs using spectroscopic redshifts from the first two years of eBOSS observations. Quasar clustering measurements can be used to infer their lifetimes and large-scale environments, which help address issues such as the physical mechanism fueling quasars and their impact on galaxy formation and evolution. For this study, we used samples of $\sim 30,600$ quasars and $\sim 120,000$ LRGs in the redshift range, $0.6 < z < 1.0$. Unfortunately, quasars are found to have relatively low number density, which causes large uncertainties in clustering amplitudes estimated from the autocorrelation function especially at low redshifts where they are rare. It is possible to make more precise measurements of quasar clustering by studying the clustering of galaxies around quasars (i.e., by measuring the cross-correlation function of quasars and galaxies), as the number density of galaxies is much higher.

Lastly, in Chapter 7, I summarize the many key techniques that this dissertation has yielded for current and future cosmological probes. I describe many key aspects of a spectroscopic survey which are essential to be understood for robust science results. I conclude by discussing ways that this work can be expanded upon in the future, especially through the further study of machine learning and advanced statistical methods.

2.0 LUMINOUS RED GALAXIES: SELECTION AND CLASSIFICATION BY COMBINING OPTICAL AND INFRARED PHOTOMETRY

The contents of this chapter have been published in Prakash, Licquia, Newman, and Rao, 2015 April 20, *The Astrophysical Journal*, Volume 803, Number 2.

2.1 INTRODUCTION

LRGs exhibit a strong 4000 Å break in their spectral energy distributions (SEDs; Eisenstein et al., 2005). At lower redshifts ($z \leq 0.6$), LRGs can be efficiently selected and their redshifts estimated using optical photometry alone, by taking advantage of this feature. This method was used to select LRG targets for the Sloan Digital Sky Survey (SDSS) and the SDSS-III / Baryon Oscillation Spectroscopic Survey (BOSS), as well as the 2dF- SDSS LRG and QSO survey (2SLAQ; Eisenstein et al., 2001; Cannon et al., 2006). However, this method becomes extremely difficult at greater distances as cosmic expansion redshifts the 4000 Å break into longer wavelength filters, therefore requiring long imaging exposure times to overcome the brightness of the night sky in the near-infrared (NIR). A new method is required in order to efficiently select LRGs at higher redshifts.

Old stellar populations exhibit global maxima in their SEDs at a rest-frame wavelength of 1.6 μm, corresponding to the minimum in the opacity of H- ions in their stellar atmospheres (John, 1988). Since the light measured from LRGs is predominantly produced by old stars, we expect this feature to dominate their overall SEDs. This enables efficient selection of LRGs at higher redshifts.

In this chapter, we demonstrate that a simple cut in optical-infrared color-color space

provides an efficient method for differentiating LRGs from other types of objects. The methods described here will be applied for selecting LRG targets for next generation spectroscopic surveys like The Extended Baryon Oscillation Spectroscopic Survey (eBOSS) and the Dark Energy Spectroscopic Instrument (DESI) Survey, and may be easily adapted to meet the needs of future prospects.

This chapter is organized as follows: In Section 2.2, we describe the construction of samples used to test LRG selection methods, including both the imaging and spectroscopic datasets used. In Section 2.3, we present a simple selection method for identifying LRGs based on optical and infrared photometry and analyze the efficiency of this selection method for a set of nominal selection cuts applied to our sample. In Section 2.4, we explore methods for optimizing the LRG selection algorithm by adjusting the parameters of our cuts in color-color space. In Section 2.5, we summarize our results and conclude with plans for future work. For this work, we assume a standard ΛCDM cosmology with H_0=100h km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$.

2.2 DATA

In this chapter, we make use of 5 cross-matched catalogs that cover two different regions of the sky, the Extended Groth Strip (EGS), with $214.0^\circ < \alpha < 215.70^\circ$ and $52.14^\circ < \delta < 53.22^\circ$, and the COSMOS field, with $149.41^\circ < \alpha < 150.82^\circ$and $1.49^\circ < \delta < 2.92^\circ$. These two regions have been surveyed by a variety of telescopes, providing photometry over a wide range of wavelengths. We have cross-matched objects based on their positions on the sky as recorded by each survey. In order to avoid duplicate matches, we match each object in the catalog with the lowest surface density to its nearest neighbors in the denser catalogs that are closer than 1.5 arcseconds. For DEEP 2 objects in the EGS, we use the cross-identifications provided by Matthews et al. (2013). Below, we briefly describe each of the catalogs used in this study.

2.2.1 Optical Photometry

Canada-France-Hawaii Telescope Legacy Survey (CFHT LS): CFHT LS consist of two parts. The Wide Survey covered $\sim$ 150 deg^2 divided over 4 fields with magnitude limits (50% completeness for point sources) of $u^\star \sim 26.0, g' \sim 26.5, r' \sim 25.9, i' \sim 25.7$, and $z' \sim 24.6$. The Deep Survey consists of 4 fields of 1 deg^2 area each, each with magnitude limits of $u^\star \sim 27.5, g' \sim 27.9, r' \sim 27.7, i' \sim 27.4$, and $z' \sim 26.2$. We use both the D2 Deep field which lies within the COSMOS region, and D3, which overlaps with EGS. We also use both the Wide survey and the Deep survey in the EGS.

We use the COSMOS *ugriz* magnitudes and their corresponding errors from the CFHT LS catalogs produced by Gwyn (2011a), which were created using the *MegaPipe* data pipeline at the Canadian Astronomy Data Centre.

CFHT LS uses MegaCam filters which are slightly redder than their SDSS counterparts. We convert CFHT LS photometry to SDSS pass-bands by inverting the filter relations given in Equations 1-5. The relations for the g, r, i and z-bands come from analyses by the SuperNova Legacy Survey (SNLS) group.[1] The relation for the u band is taken from the CFHT web pages.[2] [3] The transformed equations are:

$$u_{SDSS} = u_{Mega} + 0.181(u_{Mega} - g_{Mega}), \tag{2.1}$$

$$g_{SDSS} = g_{Mega} + 0.195(g_{Mega} - r_{Mega}), \tag{2.2}$$

$$r_{SDSS} = r_{Mega} + 0.011(g_{Mega} - r_{Mega}), \tag{2.3}$$

$$i_{SDSS} = i_{Mega} + 0.001(r_{Mega} - i_{Mega}), \text{and} \tag{2.4}$$

$$z_{SDSS} = z_{Mega} + 0.099(i_{Mega} - z_{Mega}), \tag{2.5}$$

where u_{Mega}, g_{Mega}, r_{Mega}, i_{Mega}, and z_{Mega} represent the *ugriz* magnitudes measured by CFHT LS and u_{SDSS}, g_{SDSS}, r_{SDSS}, i_{SDSS}, and z_{SDSS} are the standard SDSS magnitudes. The resulting SDSS-passband magnitudes are then corrected for Galactic extinction using

[1] http://www.astro.uvic.ca/ pritchet/SN/Calib/ColourTerms-2006Jun19/index.html

[2] http://cfht.hawaii.edu/Instruments/Imaging/MegaPrime /generalinformation.html

[3] http://www2.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/megapipe/ docs/filt.html

the dust map of Schlegel et al. (1998a), hereafter SFD. To calculate the extinction in a given band, A(λ), we interpolate the standard total-to-selective extinction ratios, i.e. A(λ)/E(B-V) from Table 6 of Schlegel et al. (1998a) for the effective wavelengths given in the filter list of CFHT LS.[4] We obtain E(B-V) values from the SFD dust map (Schlegel et al., 1998a) via the routine *dust_getval.pro* provided in the idlutils package.[5]

2.2.2 Infrared photometry

Wide-Field Infrared Survey Explorer (WISE) catalog: *WISE* completed a mid-infrared survey of the entire sky by July 2010 in four infrared channels, labeled *W1, W2, W3* and *W4*, centered at 3.4, 4.6, 12, and 22 μm, respectively. This was achieved using a 40 cm telescope with much higher sensitivity than previous infrared survey missions. WISE achieved 5σ point source sensitivities better than 0.08, 0.11, 1, and 6 mJy, corresponding to 19.1423, 18.7966, 16.4001, and 14.4547 AB magnitudes,[6] with angular resolutions of 6.1, 6.4, 6.5, and 12.0 arcseconds in the *W1, W2, W3* and *W4* channels, respectively (Wright et al., 2010a).

A detection by WISE is required for an object to be in our catalog. This restriction will have negligible effect on LRGs, since they are bright in the *W1*-band but greatly reduces the number of objects to which we must apply our selection cuts. Based on color-magnitude diagram, we observe that all $z < 20.5$ LRGs are detected in *W1* band at greater than 5-sigma. 3.4 micron (*W1*) magnitudes are taken from the publicly available WISE All-Sky Data Release catalog of Wright et al. (2010a). We convert these to the AB magnitude system and correct them for reddening using the SFD dust map (Schlegel et al., 1998a) and interpolate extinction ratios, much as above.

[4] http://www2.cadc-ccda.hia-iha.nrc-cnrc.gc.ca/en/megapipe/docs/filt.html

[5] http://www.sdss3.org/dr8/software/idlutils.php

[6] M_{AB} = - 2.5 $\times$ $\log_{10}(F_\nu/$ 3631 Jy)

2.2.3 Redshifts

COSMOS: The COSMOS photometric redshift ('photo-z') catalog from Ilbert et al. (2008a) is a magnitude-limited catalog with $I < 25$. This catalog provides photometric redshifts over the $\sim$ 2 deg^2 COSMOS field. The redshifts are computed using 30 bands covering the UV (GALEX), Visible-NIR (Subaru, CFHT, UKIRT) and mid-IR (Spitzer/IRAC). A χ^2 template-fitting method yields photo-z estimates which are calibrated with spectroscopic redshift measurements from VLT-VIMOS and Keck-DEIMOS. For details of photo-z determinations and accuracy, see Mobasher et al. (2007).

EGS: In the EGS we use the DEEP2 spectroscopic catalog. DEEP2 is a high-resolution redshift survey of $\sim$53,000 galaxies at redshifts $z \geq 0.7$ using the DEIMOS spectrograph at Keck Observatory (Newman et al., 2013). The survey covers an area of 2.8 deg^2 over four different fields. DEEP2 targeted galaxies brighter than $R_{AB} \sim 24.1$ with a spectral resolution of R ($= \Delta\lambda/\lambda$)$\sim$ 6000 and a central wavelength of 7800Å (Newman et al., 2013).

Since the DEEP2 catalogs only provide BRI photometry, Matthews et al. (2013) have created a catalog to supplement them with *ugriz* photometry from CFHTLS and SDSS. Each catalog is cross-matched by position on the sky in order to assign *ugriz* photometry to objects in the DEEP2 catalogs. We use the Matthews et al. (2013) catalog in the EGS field to obtain *ugriz* photometry. We correct this photometry for extinction as described above in Section 2.1.

All objects in our datasets are required to have reliable redshifts. This is important as this information is used in determining the rest-frame colors of galaxies. For the COSMOS field, we use the photometric redshifts, *zp_gal*, taken directly from Ilbert et al. (2008a). We don't consider objects with `zp_best`$= NULL$ as these are the objects in masked area. For the galaxies in the EGS, we make use of the heliocentric reference-frame spectroscopic redshift, *ZHELIO*, provided in the DEEP2 extended photometry catalog of Matthews et al. (2013). We also ensure that each galaxy in our sample has a securely measured redshift (i.e., we require redshift quality flags, $ZQUALITY$ of 3 or 4 in DEEP2).

2.2.4 Object type identification

COSMOS: In order to distinguish stars, galaxies, and X-ray sources within the COSMOS field from each other, we apply a variety of cuts based upon the photometric redshift estimates, *zp_best*, as well as the reduced chi-squared value associated with the separate star and galaxy template fits to each object's SED, *Chi_star* and *Chi_gal*.[7] The parameters we use have been provided by Ilbert et al. (2008a). To identify different objects, we use the criteria:

$$\textbf{X-ray sources} : zp\_best > 9, \tag{2.6}$$

$$\begin{aligned} \textbf{Galaxies} &: (Chi\_gal < Chi\_star) \\ &AND\ (0.011 < zp\_best < 9), \end{aligned} \tag{2.7}$$

$$\begin{aligned} \textbf{Stars} &: (chi\_gal > chi\_star) \\ &OR\ (chi\_gal < chi\_star) \\ AND\ &(zp\_best < 0.011\ OR\ zp\_best > 9)) \end{aligned} \tag{2.8}$$

These criteria distinguish galaxies from stars and X-ray sources. An object is flagged as a star if its SED is best fit with a stellar template or if it yields an extremely low redshift in case where a galaxy template is the better fit. An object identified as a galaxy should not only fit the galaxy template best but also yield a redshift between 0.011 and 9. Once the objects are identified, *zp_final* is our redshift indicator and its value is set to the photometric redshift, *zp_gal*, if the object is identified as a galaxy. The *zp_final* value is set to 0.0 for stars and -9.99 for X-ray sources.

EGS: We use the Hubble Space Telescope-Advanced Camera for Surveys (HST-ACS) general catalog for objects in the EGS field. The HST-ACS General Catalog is a photometric and morphological catalog created using publicly available data obtained with the ACS

[7] Variables defined the same way they appear in the catalogs.

instrument on the Hubble Space Telescope (HST). This provides a large sample of objects with reliable structure measurements. It includes approximately 470,000 sources originally observed in a variety of sky surveys, including the All-Wavelength Extended Groth Strip International Survey (AEGIS), COSMOS, GEMS, and GOODS (Griffith et al., 2012). A single Sersic model for each object is assumed for deriving quantitative structural parameters (e.g., surface brightness and effective radius).

Our goal is to be able to estimate the stellar contamination in the EGS. DEEP2 avoided targeting stars, so we can not assess this from the spectroscopic sample on its own. No such effort is required for the COSMOS field since the catalog of Ilbert et al. (2008a) contains both stars and galaxies. We use the same definition as Griffith et al. (2012) for identifying compact objects (presumed to be stars) based upon their larger surface brightness, mu_HI, and lower effective half-light radius, RE_Galfit_HI (for more details, see Figure 5 of Griffith et al. (2012)). Specifically, we identify objects with mu_HI $<$ 18.5 or (mu_HI $>$ 18.5 & RE_Galfit_HI $<$ 0.03) as stars, where RE_Galfit_HI is given in arc-seconds. Once the number of the stars is determined, we assume that the fraction of objects which are stars is uniform over the entire EGS. This gives us the total number of selected objects by our color-cut (galaxies, x-ray sources, and stars) which is then used to calculate the normalized Figure of Merit, see Figure 2.12. This step is necessary since the HST-ACS general catalog of Griffith et al. (2012) covers only a portion of the entire EGS.

2.3 METHOD OF LRG SELECTION

Our goal is to develop a method for selecting LRGs at high redshift, i.e., $z > 0.6$. One of the main challenges in LRG selection based on optical photometry alone is stellar contamination. The color overlap of stars with galaxies is illustrated in Figures 2.1 and 2.2, where contours depicting the density of stars are overlaid on the locations of galaxies (shown as dots) in g-r vs. r-i (Figure 2.1) and r-i vs. i-z (Figure 2.2) color-color plots. The strong overlap of these populations makes them difficult to separate cleanly.

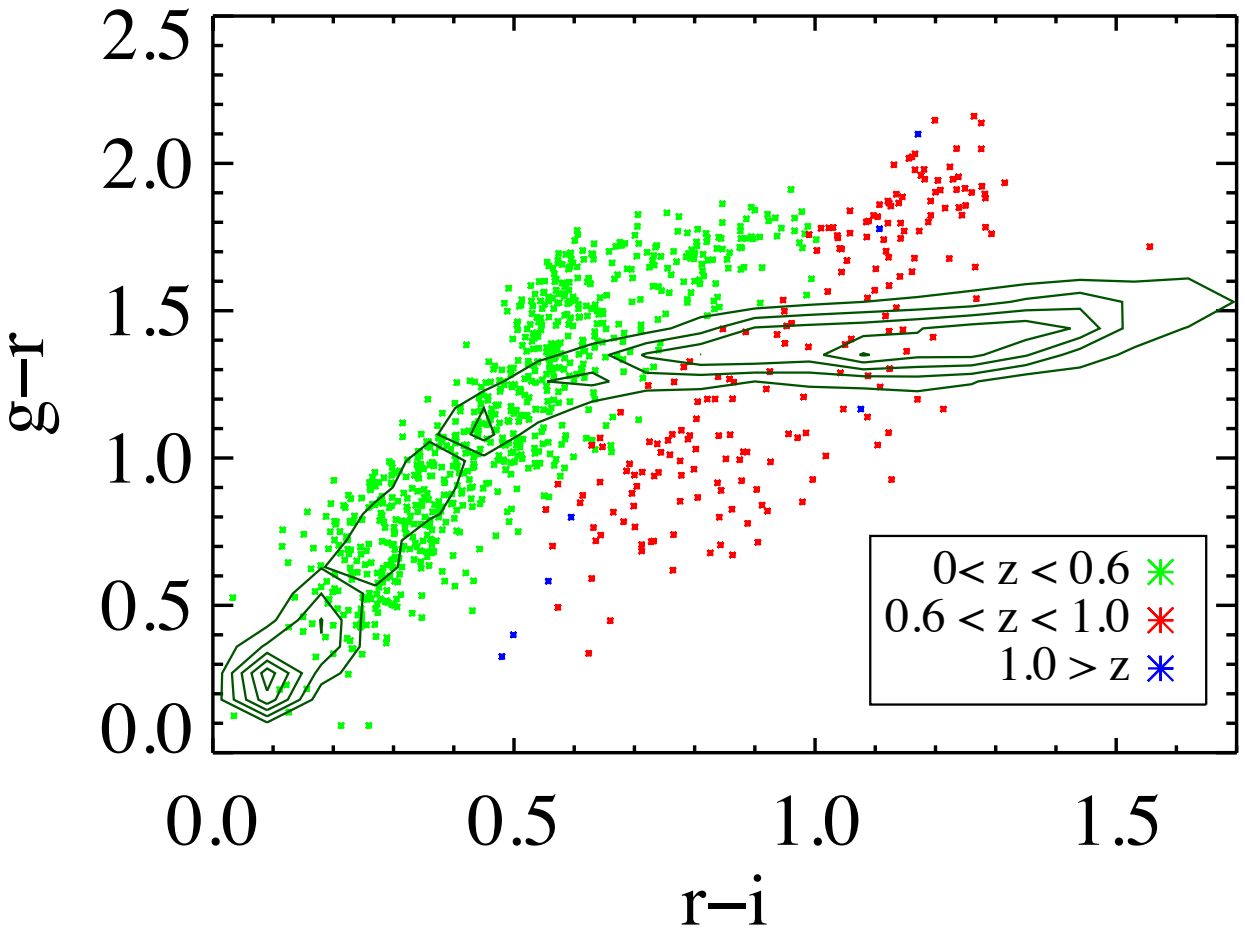


Figure 2.1: *g-r vs r-i optical color-color plot of galaxies (data points) observed by CFHT LS with COSMOS photometric redshifts. Star density contours are over-plotted to show the overlap with galaxies. This overlap makes LRG selection difficult at higher redshifts, requiring a new method of selecting them.*

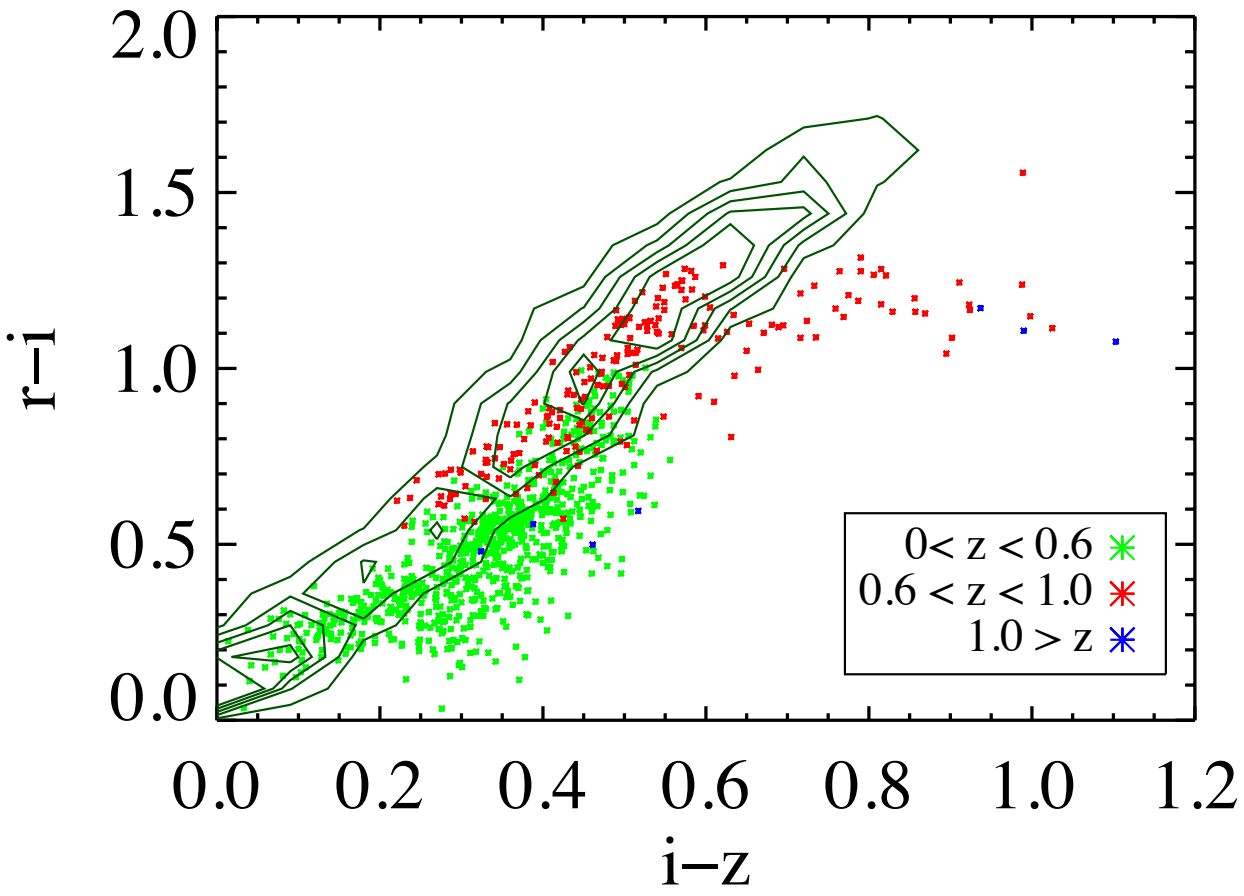


Figure 2.2: *r-i vs i-z optical color-color plot of galaxies (data points) observed by CFHT LS and the COSMOS survey, similar to Figure 2. Star density contours are overplotted to show the overlap with galaxies. Stars overlap almost entirely with high-redshift LRGs.*

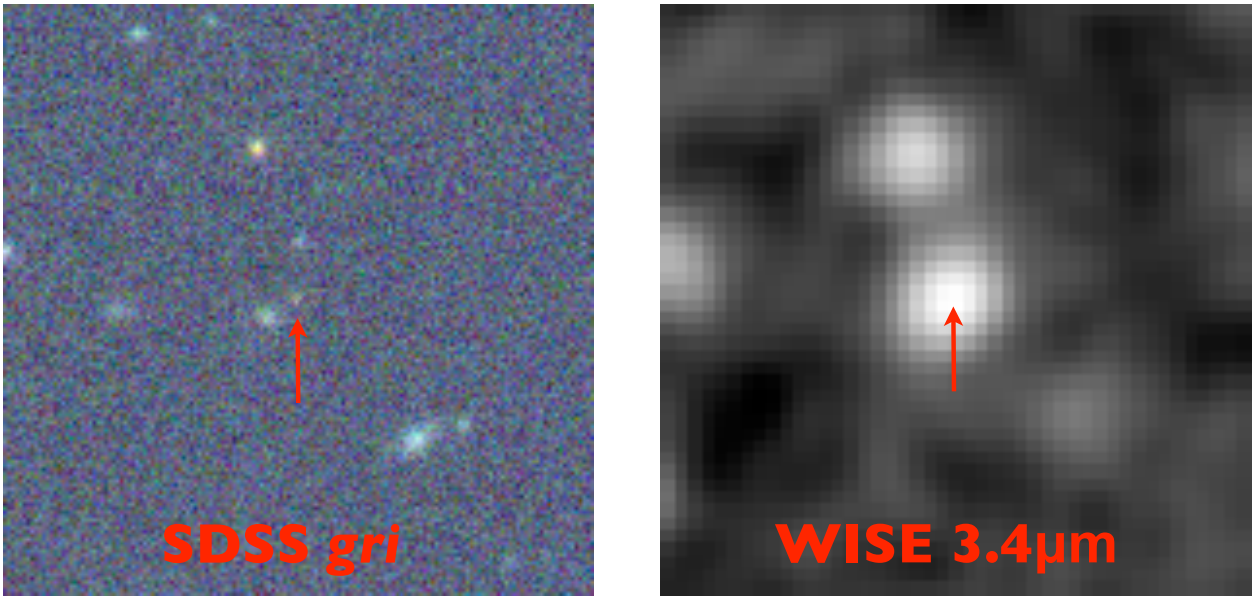


Figure 2.3: *1 arcminute square SDSS and WISE (3.4 micron) images of a $z \sim 1$ LRG. This object is at the 5 sigma detection limit in SDSS r and i but dominates the WISE image. Due to the redshifted '1.6 μm bump', LRGs at $z \sim 1$ look much brighter in the WISE 3.4 micron band than at optical wavelengths, providing a new method of selecting them.*

In this chapter, we present a new technique for identifying high-redshift LRGs which combines optical and infrared photometry. The lowest wavelength channel of imaging from the WISE satellite is centered around 3.4 μm. This overlaps with the red-shifted '1.6 μm bump' at redshifts of $z \sim 0.6 - 1$, causing LRGs at those redshifts to appear very bright in this band. This phenomenon is illustrated in Figure 2.3, where a typical LRG at $z \sim 1.0$, which is barely detected in a $1' \times 1'$ region of SDSS optical imaging, is the brightest object in the WISE NIR image of the same area. The relative brightness of LRGs in the WISE 3.4 μm band compared to the optical bands increases monotonically up to $z \sim 1$ and then declines past $z \sim 1.1$ (at which point optically bright LRGs become rare).

To match the expected spectroscopic depth of DESI LRGs, we restrict the dataset for all analyses in this chapter to those objects which have SDSS z-band magnitude $z < 20.5$. We now present a new technique which combines both optical and infrared photometry as a means of selecting galaxies that are intrinsically red and at high redshift while circumventing most stellar contamination. In Figure 2.4, we show both stars and galaxies in a

plot of *r-W1* color based on WISE and SDSS-passband photometry as a function of their SDSS *r-i* color. Here we can easily see that the two populations (stars, shown as green diamonds, and galaxies, shown as all other colored diamonds) exhibit a natural separation in the NIR-optical color space. In fact, the separation between the two populations grows as a function of galaxy redshift, allowing clean identification of the LRGs at higher redshifts, $z > 0.6$. Simultaneously, *r-i* color increases with increasing redshift as the 4000 Å break shifts redward, particularly for intrinsically red galaxies, allowing a selection specifically for intrinsically red, higher-redshift objects.

As a result, a simple cut in the optical-infrared color-color plot enables us to efficiently select LRGs at higher redshifts, rejecting bluer galaxies, lower-redshift objects, and stars. As a nominal scenario, we select all objects that have both $r\text{-}i > 0.98$ and $r\text{-}W1 > 2.0 \times (r\text{-}i)$, where r and i are extinction-corrected SDSS magnitudes and $W1$ is the magnitude in the WISE 3.4 micron pass-band on the AB system (Figure 2.4). We have determined these cuts through visual optimization by examining the populations in Figure 2.4.

Overall, our LRG color-cut selection has three free parameters: the minimum allowed *r-i* color (corresponding to the vertical line in Figure 2.4), and the slope and intercept of the line determining the minimum allowed *r-W1* color at a given *r-i* color (corresponding to the inclined line in Figure 2.4). The latter of these two criteria determines the degree to which stars are rejected from our sample of LRGs. The *r-i* cut will mostly affect the properties of the galaxies we select (e.g., their redshift distribution). We investigate the performance of this color-cut in Figure 2.6, 2.7, and 2.8. For an object to be classified as a high-redshift LRG, we generally require it to have both a rest-frame color $U - B \geq 1$ and a redshift $z > 0.6$, though we also consider other redshift thresholds. We use the `k-correct` package to obtain rest-frame $U - B$ color for all galaxies; see Blanton & Roweis (2007) for details.

To further justify our choice of redness threshold, $U - B \geq 1$, we have plotted the standard color-magnitude diagram for DEEP2 galaxies in Figure 2.5, clearly showing the red-sequence, blue sequence and the green valley. From Figure 2.6, we observe that with our nominal color-cut, 85.8% of the galaxies selected have rest-frame $U - B \geq 1$, indicating that they are intrinsically red in rest-frame color, adopting the same criterion employed by Gerke et al. (2007). The remaining galaxies selected are still relatively red and massive, but have

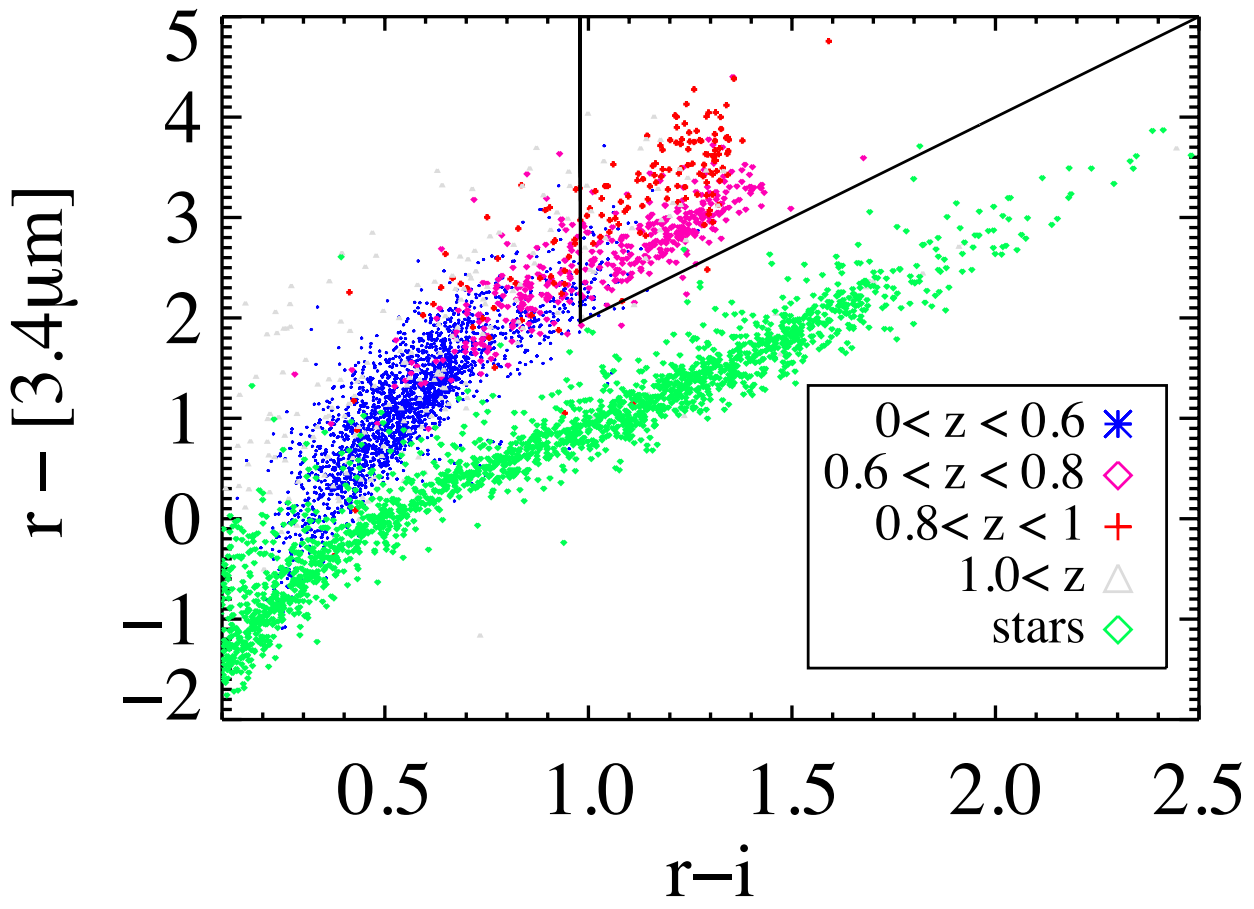


Figure 2.4: *Optical/Infrared color-color plot for galaxies observed by WISE, CFHT LS and the COSMOS survey. Blue symbols represent all the galaxies at $z < 0.6$. Pink diamonds represent galaxies at $z > 0.6$ while red crosses represent $z > 0.8$. Cyan triangles represent galaxies at higher redshift of $z > 1.0$. Stars are represented by green diamonds. The triangular region represents the LRG selection region used in Figures 2.6, 2.7 and 2.8.*

some ongoing star formation.

In Figure 2.7, we show a histogram of the redshifts measured for each of the objects selected, indicating that 77.6% of the galaxies selected by our nominal cuts fall within the redshift range of interest, i.e., $0.6 \geq z \geq 1.0$, and $< 1\%$ are stars.

2.4 OPTIMIZATION OF LRGS SELECTION

2.4.1 Optimization using Receiver Operating Characteristic

As a first attempt to optimize the efficiency of our selection method, we begin by varying the minimum allowed r-i color (corresponding to shifting the vertical line in Figure 2.4), while keeping all other parameters fixed. We interpret the results of varying this color criterion using a Receiver Operating Characteristic (ROC) plot, as shown in Figure 2.8. The ROC plot provides a visualization of the performance of a classification system; in our case, this quantifies our ability to segregate out the high-redshift LRGs in contrast to galaxies which are either blue or at lower redshift ('non-LRGs' for short). Each individual curve shows the result of using a different threshold on the minimum-allowed z to be a 'high-redshift' LRG; we only consider LRGs above the desired minimum redshift as our target population.

The y-axis of the ROC plot represents the True Positive Rate (TPR), also known as the 'sensitivity'. The TPR is defined as the fraction of all true high-redshift LRGs in the underlying sample that are within a given color cut. One of our main goals is to maximize the TPR. Of course, if the minimum-allowed r-i color was shifted so blue as to select all galaxies this would be achieved by definition. However, there is a cost associated with doing this; we would select many galaxies that are not LRGs. This misidentification is quantified as the False Positive Rate (FPR) which is plotted on the x-axis of the ROC plot. The FPR is the fraction of all non-LRGs (in our case, all $z < 20.5$, WISE-selected objects that are blue ($U - B < 1$) and/or below the desired redshift) that are placed in the LRG sample by a given color cut. Our goal is to minimize this quantity when varying the color cuts used to

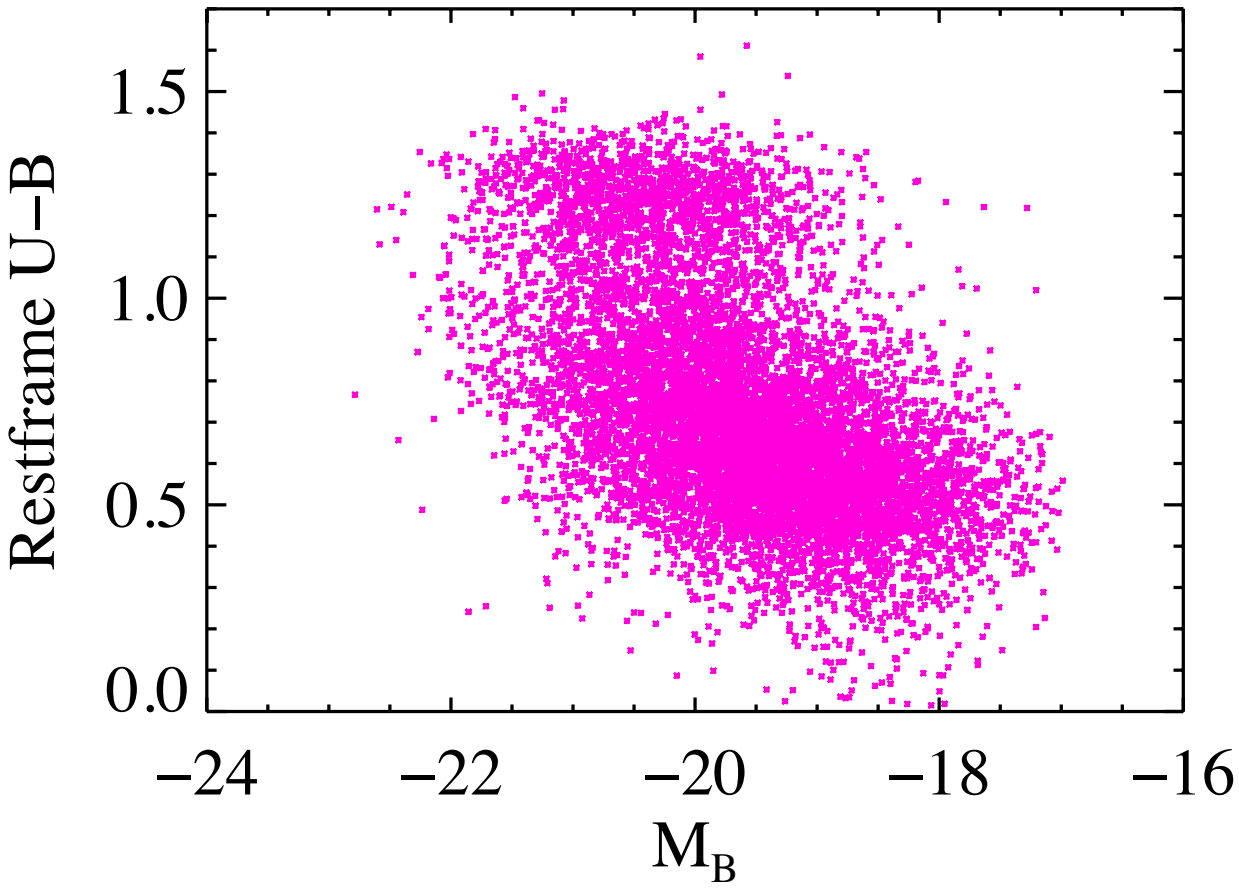


Figure 2.5: *Standard color-magnitude diagram of DEEP2 DR4 galaxies in redshift range* $0.5 < z < 1.0$. *Restframe U-B color is plotted on the y-axis, and the absolute B-band magnitude is plotted on the x-axis. We clearly see two different overdensities, one around U-B* $\sim$ *0.4 - 0.7 and another around U-B* $\sim$ *1.25. The galaxies with U-B* $\geq$ *1.0 are commonly referred to as the red-sequence galaxies while galaxies having U-B* $<$ *1.0 form the blue cloud. The low density region in between these overdensities is generally referred to as the green valley. Restframe U-B colors and absolute magnitudes are obtained using the* `k-correct` *package which assumes* $h = 1$*; see Blanton & Roweis (2007) for details. Choosing a different value of* h *will not change U-B, just* M_B.

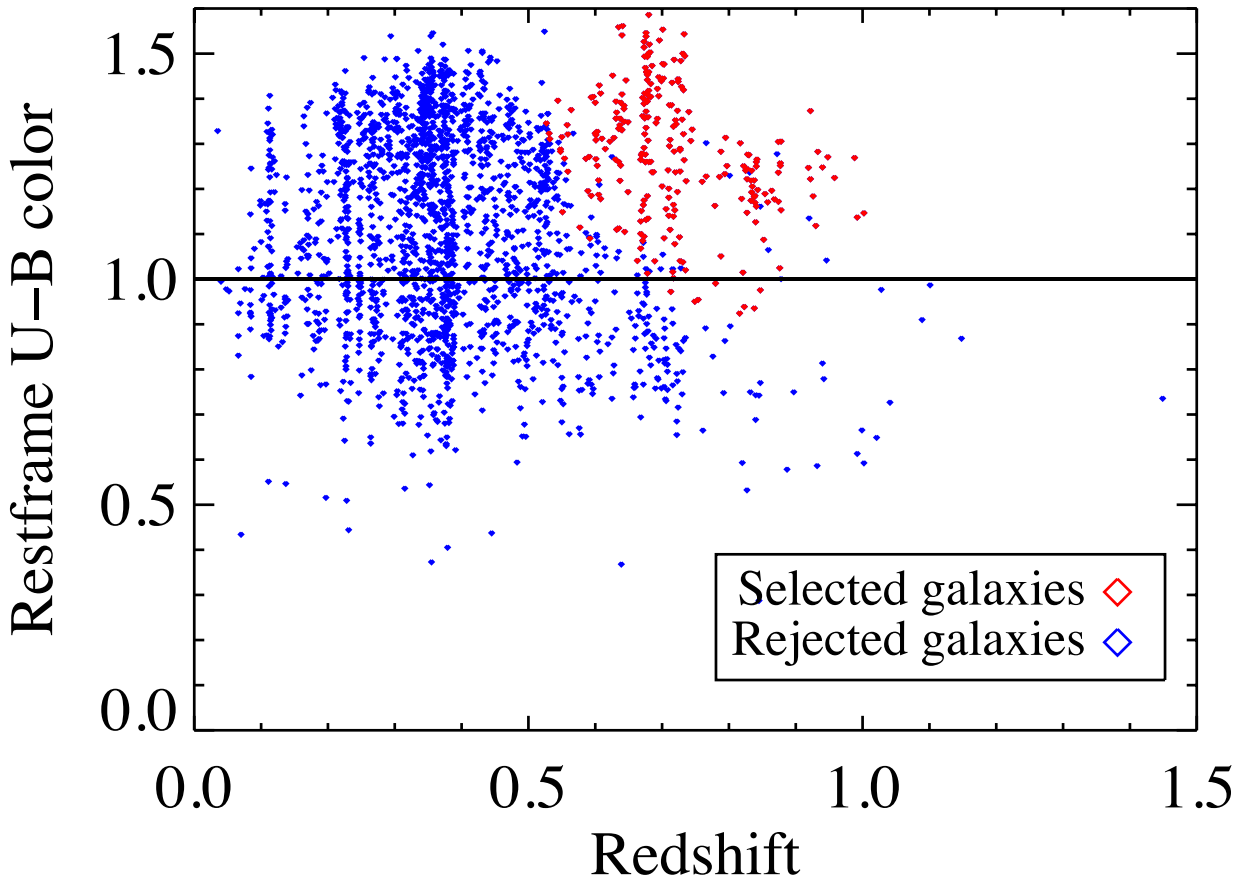


Figure 2.6: *Rest-frame U-B colors of WISE-detected, $z < 20.5$ galaxies as a function of redshift. Red dots represent the galaxies from the LRG selection region of Figure 2.4). Blue dots represent the galaxies excluded by this color selection. The red sequence corresponds well to those galaxies with U-B ≥ 1.0. Of the galaxies selected by our color cut, 85.8% have U-B greater than 1.0; most selected objects that miss this cut are only slightly bluer than the red sequence.*

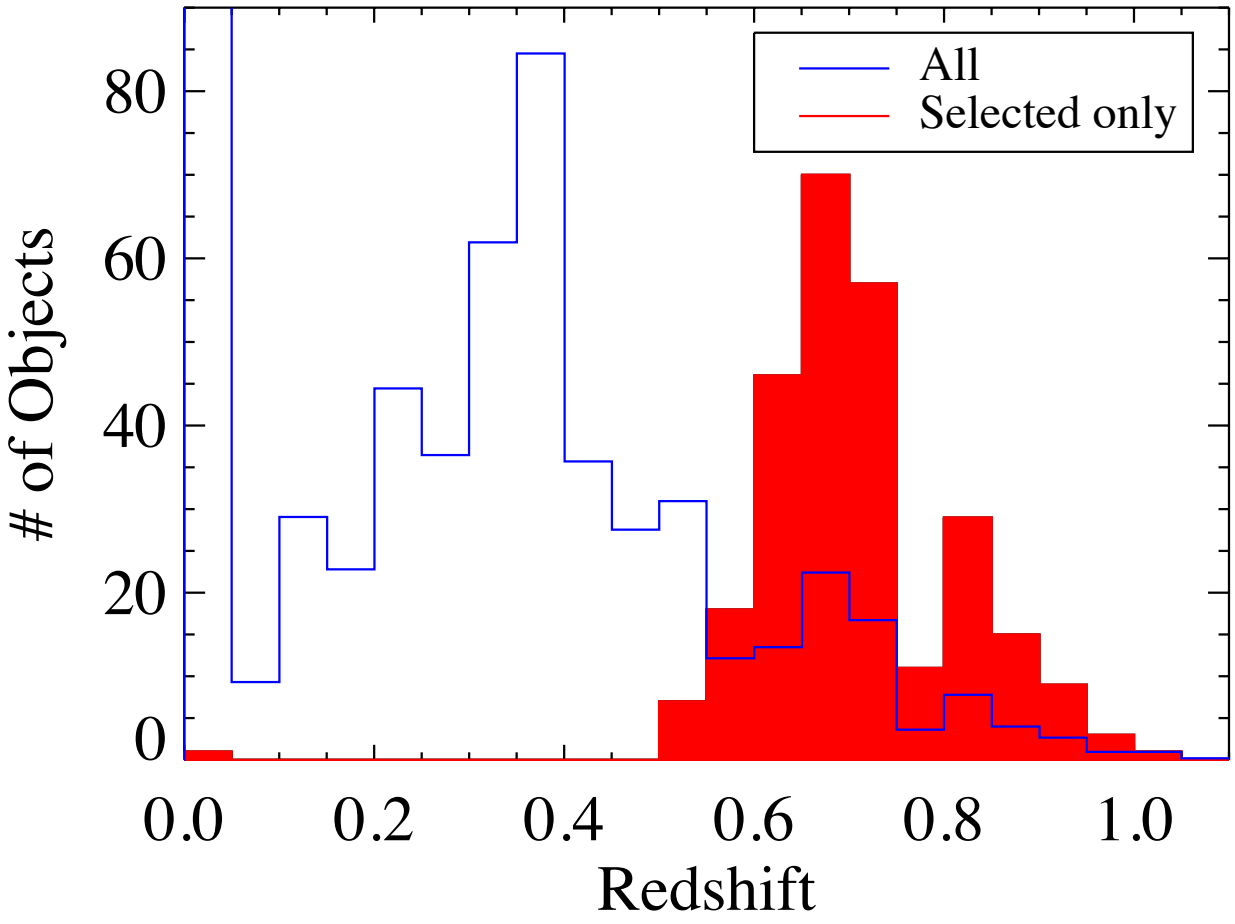


Figure 2.7: *Redshift histogram of WISE-detected, $z < 20.5$ galaxies. The blue histogram depicts the redshift distribution of the full WISE-detected, $z < 20.5$ sample, with an arbitrary renormalization applied. The red histogram represents galaxies from the LRG selection region of Figure 2.4. The selected objects predominantly fall in the redshift range $0.6 < z < 1$, with a modest 0.63% contamination by stars (appearing at $z \sim 0$).*

pick our sample while at the same time maximizing the TPR.

One common way of assessing the performance of a selection algorithm is the Area Under Curve (AUC) diagnostic (Hanley & McNeil, 1982). This is calculated by integrating the ROC curve over all FPRs. Here, we assess the efficiency of selecting LRGs using 3 different redshift thresholds, $z \geq 0.55$, 0.6, or 0.65. We have varied the minimum allowed r-i color over the range in values from 0.8 to 1.2, and calculated the TPR and the FPR for each selection. Figure 2.8 shows that, with our chosen cuts, we are able to attain a TPR of $85 - 95\%$ (depending on the choice of the minimum allowed 'high' redshift, threshold z) while at the same time keeping the FPR below $\sim$3%. Based on the AUC, we conclude that our selection algorithm performs best for a threshold redshift of $z \geq 0.6$ (corresponding to the blue curve in Figure 2.8), as it encompasses the maximum area.

2.4.2 Optimization using Figure Of Merit for large scale structure studies

To optimize the efficiency of our methods further, we attempt to account for the contribution to cosmological analyses from those non-LRG objects which are selected. While LRGs are the prime targets we are after, those galaxies which are not red and have a redshift of $z \geq 0.6$ (hereafter, 'high-z blue galaxies'), still provide useful information. To assess this, we define a Figure Of Merit (FOM) as

$$FOM = a \times n_{LRGs} + b \times n_{high-z\ blue\ gals}, \tag{2.9}$$

where a and b are constants weighting the targets, and n_{LRGs} and $n_{high-z\ blue\ gals}$ represent the number density (number per unit area) of LRGs and high-redshift blue galaxies, respectively, for a given set of color cut criteria. Since LRGs are our prime targets, we assign $a = 1.0$ and $b < a$. For the purpose of this chapter, we adopt $b = 0.75$ as an example; a more ideal weighting would set b/a according to the relative contribution of each class of objects towards, e.g., the uncertainty in the BAO scale. We assume that stars, X-ray sources (which tend to be at higher z), and galaxies at $z \leq 0.6$ contribute nothing towards the FOM. We want to optimize not only the total FOM but also the FOM per object which we will refer

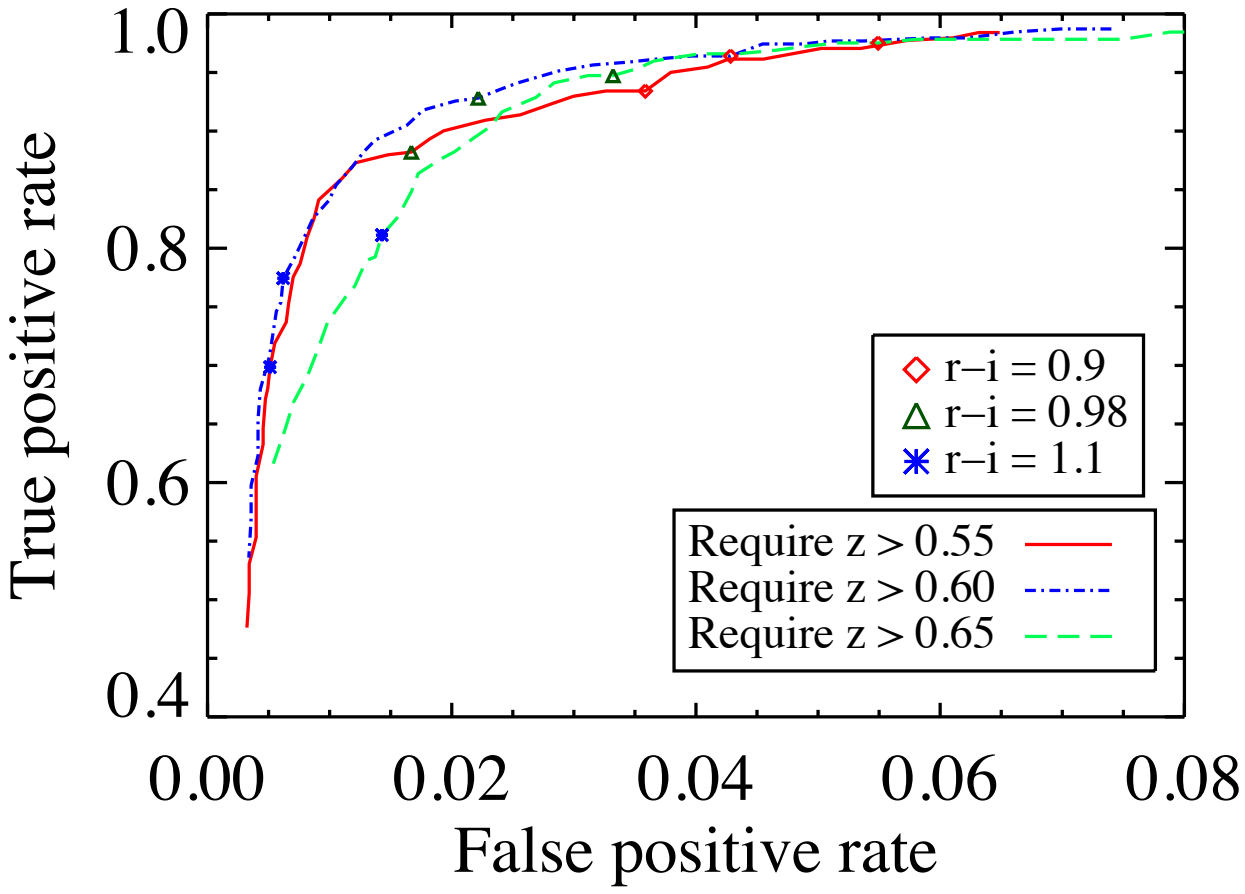


Figure 2.8: *The Receiver Operating Characteristic plot for changes in the minimum allowed r-i for selecting LRGs over the range* $0.8 < r\text{-}i < 1.2$. *For each curve, we only consider intrinsically red objects that have redshifts above a different specified minimum redshift, z, to be high-z LRGs. The brown triangles represent the performance at the threshold of* $r\text{-}i = 0.98$. *Similarly, the red diamonds and the blue stars represent the performance at the threshold of* $r\text{-}i = 0.9$ *and* $r\text{-}i = 1.1$, *respectively. Based on the commonly-used Area Under Curve criterion, our color cuts perform slightly better for LRGs with minimum redshift of* $z \sim 0.6$ *than 0.55 or 0.65.*

to as the Normalized FOM. The total FOM represents the total constraining power of the whole sample and will increase even when we select lower-value high-z blue galaxies. On the other hand, the Normalized FOM will increase only when LRGs make up a higher fraction of the sample and, therefore, can be more useful for optimization. Both should be considered; selecting all the objects regardless of color would yield a large total FOM but little value per object; while an extremely restrictive selection could have normalized FOM$\sim$ 1, but have little total constraining power due to the small number of objects chosen. We then create a 3-dimensional grid to tabulate the FOM for each possible combination of the minimum allowed r-i and the slope and intercept of the line marking the minimum allowed r-*W1* for a given r-i color. Figure 2.8 shows that our nominal cut of r-$i = 0.98$ performs quiet well on this metric.

To simplify the search space, we next fix our r-i cut at the optimal value and analyze the FOM when varying the two parameters associated with the r-*W1* cut. In Figure 2.9, we plot the FOM as a function of the slope and the intercept of the r-*W1* threshold line. We overplot the contours of constant FOM to highlight the linear nature. It is clear that the FOM depends primarily on a fixed combination of these parameters. As a result, the optimal value of the slope of the line, S_{r-W1}, given a value the intercept, i_{r-W1}, can be obtained from the relation:

$$S_{r-W1} = 2.0 - 0.4 \times i_{r-W1}. \tag{2.10}$$

Although the exact form of this optimal relation is dependent on the value of b in Equation 2.9, the existence of this correlation is significant, as it reduces our selection algorithm from a 3-parameter to a 2-parameter problem. We find that similar linear correlations occur for different values of b in Equation 2.9 (e.g. $b = 0.25$ or 0.5). In contrast to FOM, the normalized FOM depends little on the slope/intercept in the relevant parameter range, so in this case, our decisions are driven by FOM.

Based on this approximation, we define a new variable v:

$$v = (2.0 - S_{r-W1}) - 0.4 \times i_{r-W1}. \tag{2.11}$$

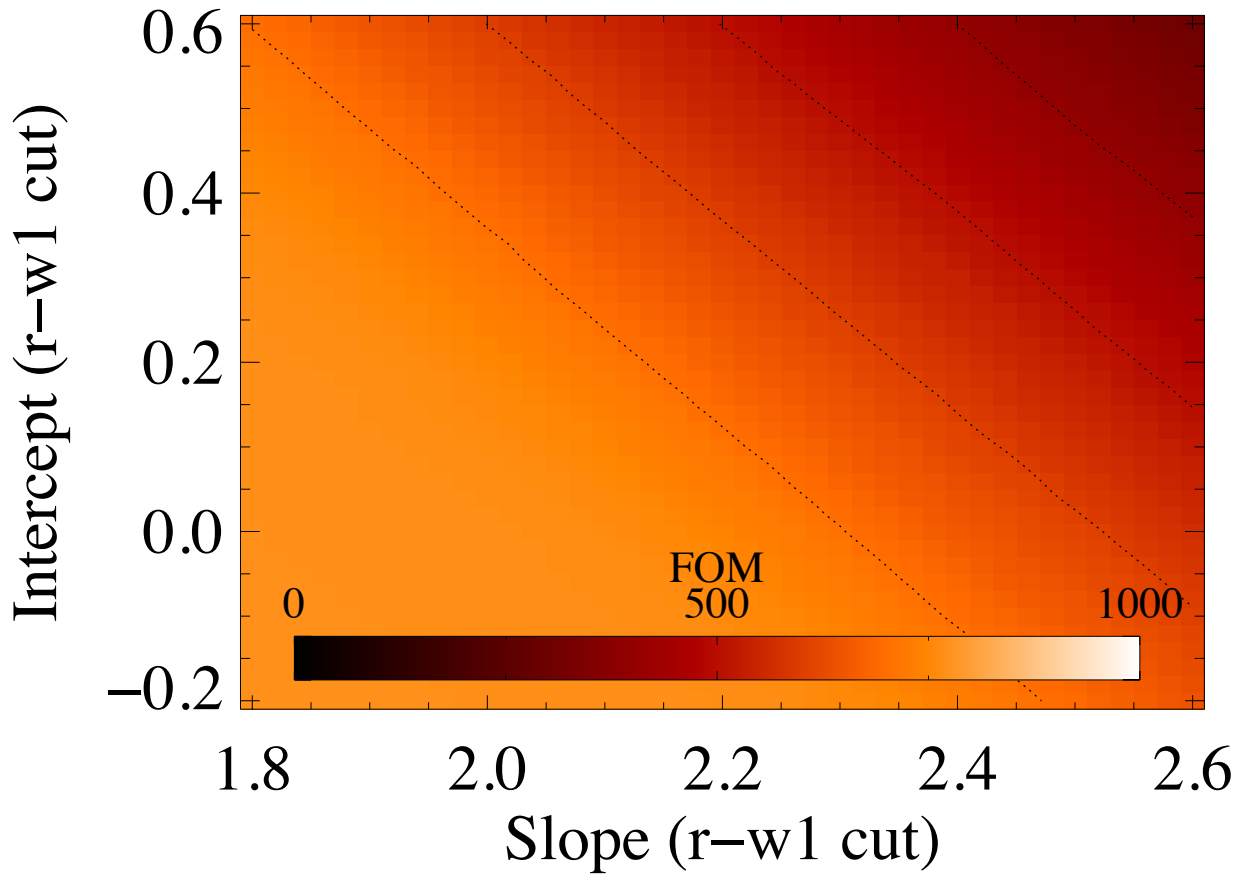


Figure 2.9: *The LRG selection figure of merit (see Equation 2.9) as a function of the slope and intercept of the r-W1 cut, using a fixed r-i threshold of 0.98. The visible pattern and the overplotted contours of constant FOM show that these parameters can be effectively replaced by a single linear combination of the two, as in Equation 2.10.*

This is defined such that v=0 at our nominal parameter values for the *r-W1* cut. Next, we create a 2-dimensional grid to tabulate the FOM at each possible combination of the two parameters in our model, the minimum allowed *r-i* and v. This grid is then analyzed to determine the parameter values which maximize the FOM. In Figure 2.10, we have plotted the maximum of the FOM over all i_{r-W1} values as a function of the *r-i* color cut. The maximum FOM decreases monotonically as the threshold *r-i* color moves redward (corresponding to moving the vertical line in Figure 2.4 to the right). This result is consistent with our expectation, since we select a decreasing number of both LRGs and high-z blue galaxies as the *r-i* cut is moved redward.

In Figure 2.11, we have shown the FOM and normalized FOM as a function of v assuming $r\text{-}i = 0.98$. We have also overplotted the fraction of objects selected as a consistency check. The flatness of the curves in Figure 2.11 can be understood with the help of Figure 2.4, as well as Equations 2.10 and 2.11. As v becomes positive, the intercept of the *r-W1* threshold line becomes negative. This region lies within the empty region separating stars and galaxies in Figure 2.4. Hence, we do not see any significant change in any of the quantities plotted. As v is further increased, the *r-W1* cut starts including stars into our sample. This causes an abrupt increase in the fraction of objects selected and a corresponding abrupt decrease in the FOM normalized by the number of selected objects. However, the total FOM remains mostly unaffected as stars do not contribute to it.

Figure 2.8 illustrates that $r\text{-}i = 0.98$ is a well-optimized cut for our purposes. Any decrease in this parameter causes the FPR to increase significantly without any significant gain in the TPR. Overall, we conclude that our nominal cuts of $r\text{-}i > 0.98$ and *r-W1* $> 2.0 \times$ *(r-i)*, are well optimized and are the final cuts used in this selection. We find that when varying our threshold redshift marginally, e.g. $z > 0.55$ or $z > 0.65$, our analysis yields similar results for the optimal value of selection parameters.

The analyses described so far all rely on our COSMOS field sample. It is worthwhile to test whether adding more data from different regions of the sky, which will reduce sample/cosmic variance, can improve our optimization. As explained in Section 2, in the EGS, we have obtained DEEP2 extended photometry from the catalog of Matthews et al. (2013). We repeat the same analysis as was done for the COSMOS field to estimate the FOM for our

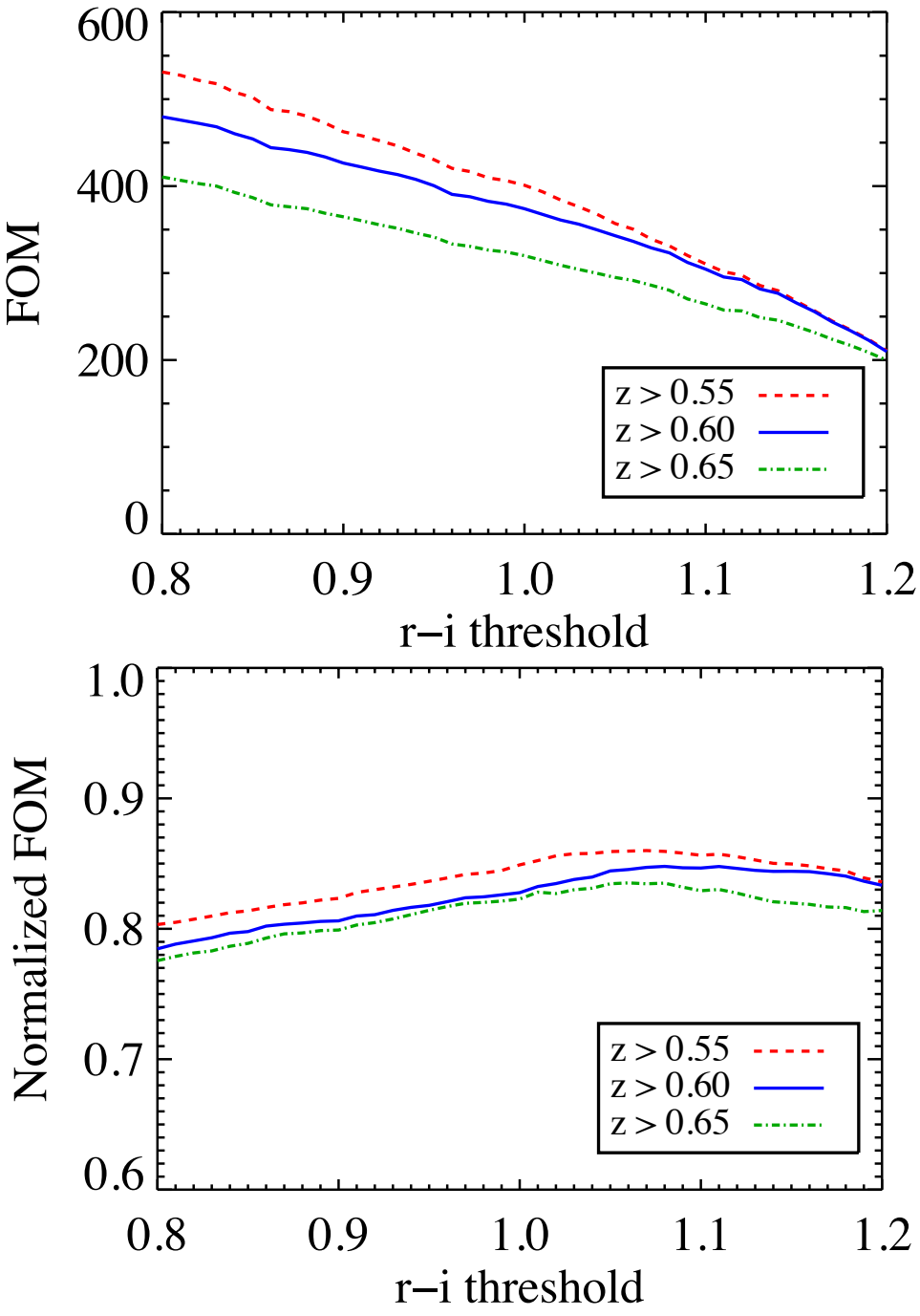


Figure 2.10: *The maximum FOM for a given r-i color threshold for $z < 20.5$ galaxies, using data in the COSMOS region. Each individual curve shows the result of using a different choice for the minimum-allowed z for a high redshift LRG. In the top panel, the monotonic decrease in FOM as the threshold r-i color moves redward (corresponding to moving the vertical line in Figure 2.4 to the right) is consistent with our expectations, since we select a decreasing number of both LRGs and high-z blue galaxies as the r-i cut is moved to the right. In the bottom panel, we analyze the normalized FOM as the threshold r-i color moves redward. For r-i < 1.05, as the color cut moves redder, it selects a higher fraction of high-z LRGs, increasing the purity of the sample and hence the normalized FoM monotonically.*

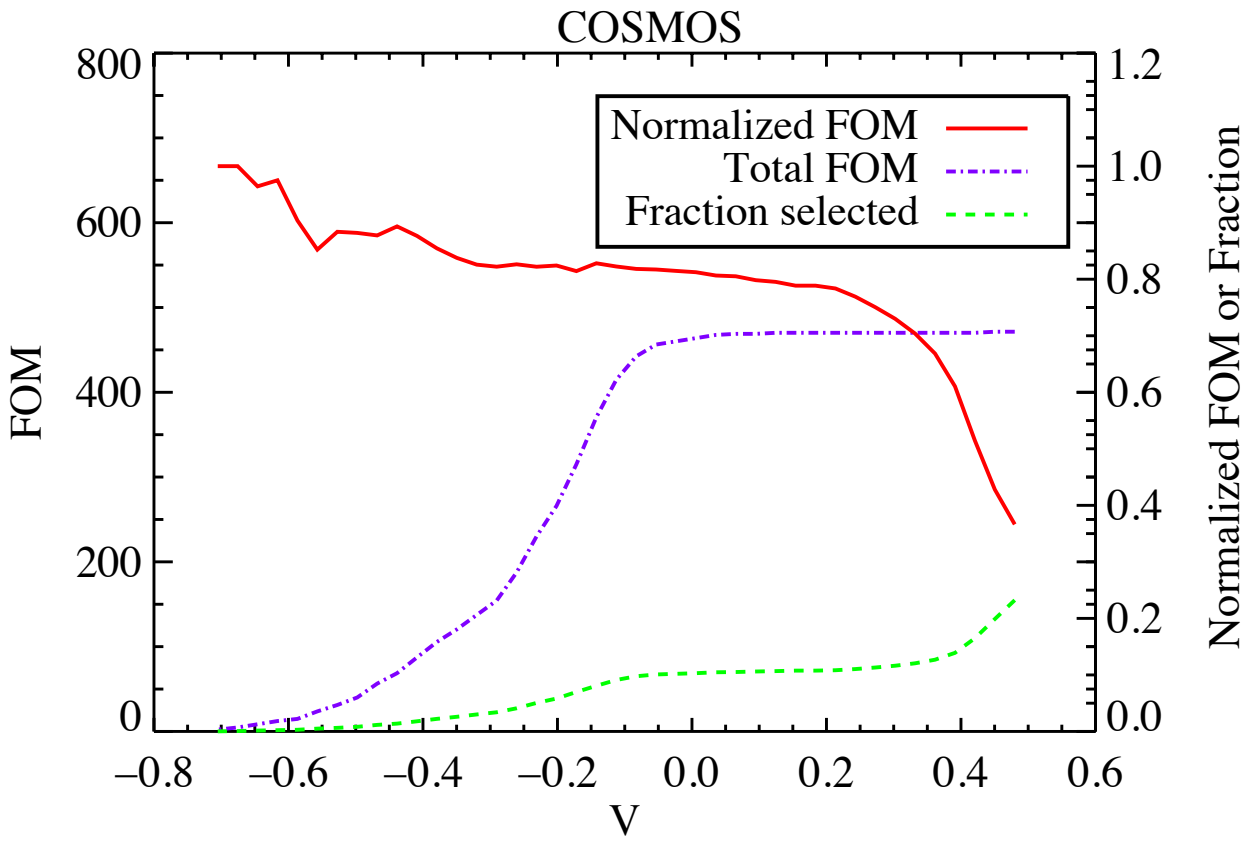


Figure 2.11: *The FOM at r-i = 0.98 for z< 20.5 galaxies as a function of the combined slope/intercept parameter (v) in the COSMOS region. Only galaxies above a threshold redshift of z > 0.6 are counted in the FOM. Each individual curve shows a different quantity. The second y-axis (on the right) is used for plotting both the Normalized FOM (FOM normalized by the total number of objects selected by a given color cut) and the fraction of all z < 20.5 objects that are targeted. Values of v around 0 are roughly optimal for both the total and normalized FOM.*

2-parameter model. However, to estimate the total number of objects selected by a given color cut, we need to estimate the stellar contamination. Since DEEP2 avoided targeting stars (Newman et al., 2013), this is done separately using the HST-ACS general catalog of Griffith et al. (2012), as described in Section 2.4. We otherwise repeat the same analysis as done for the COSMOS field to estimate the stellar contamination for a given color cut in 2D parameter space. The FOM in the EGS region shows a very similar behavior as in the COSMOS field as can be seen in Figure 2.10 and 2.11, indicating that sample/cosmic variance is not a major issue.

In Figure 2.12, we show the behavior of the FOM averaged over both the EGS + COSMOS fields. The plot shows a very consistent behavior, enabling us to conclude that our baseline color selection, $r\text{-}i > 0.98$ and $r\text{-}W1 > 2.0 \times (r\text{-}i)$ is indeed well-suited for selecting $z > 0.6$ LRGs using this method.

2.5 CONCLUSIONS AND FUTURE WORK

We have found a reliable and efficient method of identifying and selecting LRGs at higher redshifts by combining optical and infrared photometry. We have explored a variety of methods for optimizing our color cuts, given a particular set of rest-frame color and redshift requirements. With these optimization procedures, we can, for instance, tune the redshift range to select LRGs as required by different surveys.

These methods have now been used to assemble large samples of LRGs. More than 10,000 $z < 20$, SDSS+WISE selected LRGs were targeted by a BOSS Ancillary program in 2012-2013 (SDSS DR12, in prep.). This will not only provide a good check on our selection methods but will also greatly increase the sample that we can use to optimize the selection process further. We have also selected LRGs based on similar methods, but using colors derived only from SDSS i,z and *WISE W1* (i.e., using $i\text{-}W1$ and $i\text{-}z$ colors). Selection in these redder bands helps for targeting higher-z LRGs, but they are not as efficient as the combination of r, i, and $W1$ in star-galaxy separation.

We have created a sample of LRGs over the entire SDSS footprint which has been

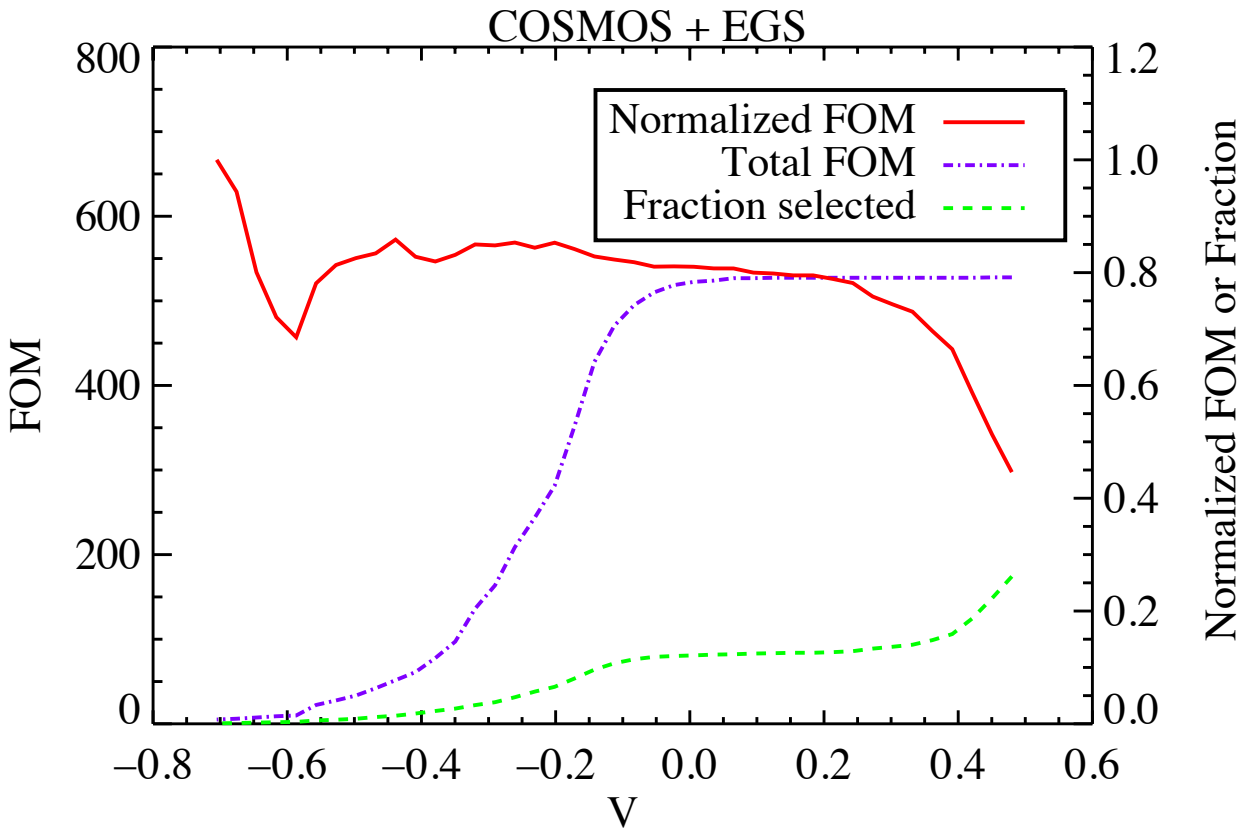


Figure 2.12: *The FOM for r-i = 0.98 as a function of the combined slope/intercept parameter (v), averaged over both the COSMOS and the EGS fields. Only galaxies above a threshold redshift of $z > 0.6$ are countedd in the FOM. Similar to Figure 2.11, values of v around 0 prove to be roughly optimal for both the total and normalized FOM.*

used in selecting targets for a second BOSS ancillary survey, the SDSS Extended Quasars, Emission line galaxy, and LRGs Survey (SEQUELS). In SEQUELS, we have targeted $\sim$ 70,000 $z < 20$ LRGs selected by one of two color cuts: One which utilizes *r*-*i*, *r*-*W1*, and a minimum value of *i*-*z*, and a second which uses only *i*-*z* and *i*-*W1* colors. The work of analyzing the resulting spectra is in progress and will be reported in future publications. The same methods are being used for selecting LRG targets for eBOSS, which began observations in Fall 2014. The results from these selection algorithms have been described in Chapter 3 (also see the companion paper, Prakash et al. (2016)).

We are also investigating even deeper selections of LRGs using *r*, *z*, and *W1* photometry for the proposed DESI survey (see DESI Conceptual Design Report).[8] Optical/IR LRG selections have proved to be effective in our tests, and will provide a cornerstone sample for BAO surveys through the next decade.

[8] http://desi.lbl.gov/cdr/

3.0 THE SDSS-IV EXTENDED BARYON OSCILLATION SPECTROSCOPIC SURVEY: LUMINOUS RED GALAXY TARGET SELECTION

The contents of this chapter have been published in Prakash, Licquia, Newman, et al., 2016 June 8, *The Astrophysical Journal Supplement Series*, Volume 224, Number 2

3.1 INTRODUCTION

Surveys like the Sloan Digital Sky Survey (SDSS; York et al., 2000a) and the *SDSS-III*/Baryon Oscillation Spectroscopic Survey (*BOSS*), as well as the 2dF-SDSS LRG and QSO survey (e.g. Eisenstein et al., 2001; Cannon et al., 2006) have successfully targeted LRGs for BAO detection. In combination, *SDSS-I*, *SDSS-II* and *SDSS-III* targeted LRGs at $z \lesssim 0.7$ to a magnitude limit of $i < 19.9$ and $i_{fiber2} < 21.5$ (Eisenstein et al., 2001, 2005, 2011; Dawson et al., 2013a). The methods used to select LRGs for these studies are limited in redshift range as a result of using optical photometry alone for selection. Identifying LRGs with shallow optical photometry becomes prohibitively difficult at higher redshifts as the 4000 Å break passes into the near-infrared and colors overlap strongly with M stars.

New multi-wavelength imaging is now available which allows high-redshift LRGs to be selected much more efficiently than optical-only imaging would make possible. In particular, optical-infrared (optical-IR) colors provide a powerful diagnostic for separating galaxies and stars(Prakash et al., 2015), as well as a diagnostic of redshift. As a result, infrared observations from satellites such as the Wide-field Infrared Survey Explorer (*WISE*; Wright et al., 2010a) provides additional information for targeting LRGs in regions of optical color space

that would otherwise be heavily contaminated by stars.

Increasing our current sample of LRGs to higher redshifts will allow measurements of the Baryon Acoustic Oscillation (BAO) feature, and hence of the expansion rate of the universe (Seo & Eisenstein, 2003; Lin & Mohr, 2003; Ross et al., 2008), during the era when accelerated expansion began. An optical + *WISE* selection makes it possible to target LRGs in the redshift range $0.6 \lesssim z \lesssim 1$ efficiently (Prakash et al., 2015); with spectroscopy of these targets, we can obtain stronger constraints on the BAO scale at these redshifts. At even higher redshift, other tracers such as quasi-stellar objects (quasars) and Emission Line Galaxies (ELGs) can be used to provide further complementary probes of the BAO scale. In combination, these target classes can provide powerful constraints on the evolution of cosmic acceleration across a wide range of redshifts. This led to the conception of a new survey, the *extended Baryon Oscillation Spectroscopic Survey* (*eBOSS* Dawson et al., 2016) as part of *SDSS-IV*.

The LRG component of *eBOSS* will obtain spectra for ∼375,000 objects. Approximately 265,000 of these are expected to be LRGs in the redshift range $0.6 < z < 1.0$, with a median redshift of $z \sim 0.7$. The main goal of this spectroscopic campaign is to produce more precise measurements of the BAO signal at $0.6 < z < 1.0$, thus extending probes of the BAO scale using LRGs beyond the *BOSS* redshift range. *eBOSS* LRGs are also expected to yield a 4% measurement of Redshift Space Distortions (RSD) which will allow improved tests of General Relativity at these redshifts (e.g., Beutler et al., 2014a,b; Samushia et al., 2014).

Altogether, *SDSS-IV*/*eBOSS* will produce a spectroscopic sample of both galaxies and quasars over a volume that is 10 times larger than the final *SDSS-III BOSS* sample, although at lower target density. This sample will enable a wide range of scientific studies beyond a BAO measurement. For example, the resulting sample of hundreds of thousands of LRGs extending to $z = 1$ will be useful for a variety of studies of the evolution of the brightest elliptical galaxies, including measurements of luminosity functions, mass functions, size evolution, and galaxy-galaxy lensing.

In this chapter, we describe the algorithm used to select LRG targets for the *eBOSS* survey. Further technical details about *eBOSS* can be found in companion papers on Quasar

selection (Myers et al., 2015), ELG selection (Comparat et al., 2015a), survey strategy (Dawson et al., 2016), and the *Tractor* analysis of *WISE* data (Lang et al., 2014).

The chapter is organized as follows. In Section 2, we outline the goals of *eBOSS* and the requirements placed on the LRG sample to meet these goals. The parent imaging data used for *eBOSS* LRG target selection is outlined in Section 3. In Section4, we describe our new method of LRG selection and supporting tests for this method that were conducted during *BOSS*. In Section 5, we describe the *eBOSS* LRG targeting algorithms and the meaning of the relevant targeting bits, while Section 6 uses the latest results from *eBOSS* to test the target selection algorithm. An important criterion for any large-scale structure survey is sufficient homogeneity to facilitate modeling of the distribution of the tracer population, i.e., the 'mask' of the survey. In Section 7, we use the full *eBOSS* target sample to characterize the homogeneity of *eBOSS* LRGs. We present conclusions and future implications for *eBOSS* LRGs in Section 8.

Unless stated otherwise, all magnitudes and fluxes in this chapter are corrected for extinction using the dust maps of Schlegel et al. (1998a), hereafter SFD, and are expressed in the AB system (Oke & Gunn, 1983a). The SDSS photometry has been demonstrated to have colors that are within 3% of being on an AB system Schlafly & Finkbeiner (2011a). We use a standard ΛCDM cosmology with H_0=100h km s^{-1} Mpc^{-1}, $h = 0.7$, $\Omega_{\rm M} = 0.3$, and $\Omega_\Lambda = 0.7$, which is broadly consistent with the recent results from *Planck* (Planck Collaboration et al., 2014).

3.2 COSMOLOGICAL GOALS OF EBOSS AND IMPLICATIONS FOR LRG TARGET SELECTION

3.2.1 Overall Goals for the Luminous Red Galaxy Sample

The primary scientific goals of the *eBOSS* LRG survey are to constrain the scale of the BAO to 1% accuracy over the redshift regime $0.6 < z < 1.0$. This requires selecting a statistically uniform set of galaxies with the desired physical properties for which spectroscopic redshifts

can be efficiently measured. The density of selected LRGs must not strongly correlate with either tracers of potential imaging systematics (e.g., variations in the depth of the imaging) or with astrophysical systematics such as Galactic extinction and stellar density.

3.2.2 Target Requirements for LRGs

As explained in Dawson et al. (2016), a density of 50 deg^{-2} spectroscopic fibers are allocated to *eBOSS* LRGs and a density of 40 deg^{-2} LRGs with redshifts $0.6 < z < 1.0$ is required, over the projected 7,500 deg^{-2} survey footprint, to meet the eBOSS scientific goals (see Dawson et al. (2016) for more details). If one were to consider *eBOSS* in isolation, this corresponds to a requirement that 80% of *eBOSS* LRG targets result in a spectroscopically confirmed galaxy with $0.6 < z < 1.0$. However, given that *BOSS* observed a density of 12 deg^{-2} LRGs (assuming half of the *BOSS* galaxies with $z > 0.6$ are LRGs, consistent with Ross et al. (2014), eBOSS can obtain its required number density of LRGs if it observes 28 deg^{-2} additional LRGs with $0.6 < z < 1.0$. Additionally, we require the redshifts be accurate to better than 300 km s^{-1} RMS and robust such that the fraction of catastrophic redshift errors (exceeding 1000 km s^{-1}) is $< 1\%$ in cases where the redshifts are believed to be secure. The construction of a sample designed to fulfill these requirements is described in Sections 3.4 and 3.5.

A further requirement to obtain robust BAO measurements is that the density of selected LRGs must not strongly correlate with either tracers of potential imaging systematics (e.g., variations in the depth of the imaging) or with astrophysical systematics such as Galactic extinction and stellar density.*BOSS* has shown that fluctuations associated with surveys artifacts can be handled effectively via weighting schemes provided the amplitude of fluctuations is relatively small (Ross et al., 2012a). To facilitate weighting schemes in future clustering studies, we require that that fluctuations in the expected target density as a function of potential imaging systematics, stellar density, and Galactic extinction be less than 15% (total variation around mean density). We require density differences due to imaging zero point variations in any single band to be below 15% as well. Tests of the homogeneity of the LRG target sample are presented in Section 3.7.

3.3 PARENT IMAGING FOR TARGET SELECTION

3.3.1 Updated calibrations of *SDSS* imaging

All *eBOSS* LRG targets rely on imaging from the *SDSS-I/II/III*. *SDSS* photometry was obtained by the *SDSS* telescope (Gunn et al., 2006) using its wide-field imaging camera (Gunn et al., 1998) in the *ugriz* system (Fukugita et al., 1996a). *SDSS-I/II* primarily obtained imaging over the $\sim 8400\,\mathrm{deg}^2$ "Legacy" area, $\sim 90\%$ of which was in the North Galactic Cap (NGC). This imaging was released as part of *SDSS* Data Release 7 (DR7; Abazajian et al., 2009). The legacy imaging area of the *SDSS* was expanded by $\sim 2500\,\mathrm{deg}^2$ in the South Galactic Cap (SGC) as part of *DR8* (Aihara et al., 2011). The *SDSS-III/BOSS* survey used this DR8 imaging for target selection over $\sim 7600\,\mathrm{deg}^2$ in the NGC and $\sim 3200\,\mathrm{deg}^2$ in the SGC (Dawson et al., 2013a). LRG targets for *eBOSS* have been selected over the same footprint covered by *BOSS*; however, ultimately *eBOSS* will obtain spectroscopy for LRGs over a roughly $7500\,\mathrm{deg}^2$ subset of this *BOSS* area, utilizing 50% of the dark time for 6 years in *SDSS-IV*. In this available time, it would not be possible to observe the full extragalactic footprint available from SDSS imaging.

Although conducted over the same *area* as *BOSS*, *eBOSS* target selection takes advantage of updated calibrations of the *SDSS* imaging. Schlafly et al. (2012) have applied the "uber-calibration" technique of Padmanabhan et al. (2008) to imaging from the *Pan-STARRS* survey (Kaiser et al., 2010), achieving an improved global calibration compared to *SDSS DR8*. The improvements in the photometric accuracy is very modest compared to DR8, typically less than 0.5%. More importantly, *BOSS* was still acquiring imaging during its program, and therefore there was no single photometric solution for the full footprint. In contrast, *eBOSS* is targeting from a single photometric solution. Targeting for *eBOSS* is conducted using *SDSS* imaging that is calibrated using the Schlafly et al. (2012) *Pan-STARRS* solution. We will refer to this as the "updated" photometry below.

Specifically, targets are selected using the updated *SDSS* photometry stored in the `calib_obj` files, the basic imaging catalog files used in the *SDSS-III* data model.[1] The

[1] e.g., `http://data.sdss3.org/datamodel/files/PHOTO_SWEEP\/RERUN/calibObj.html`

updated Pan-STARRS-calibrated photometry will be made available as part of a future *SDSS* Data Release. The magnitudes provided in these files are Pogson magnitudes (Jones, 1968) rather than the asinh magnitudes used for some *SDSS* data releases (Lupton et al., 1999). We use `Model Magnitudes` for all colors and fluxes used in selection. The `Model Magnitudes` are obtained by first determining what type of model (exponential or deVaucouleurs) best fits the object image in the 'canonical' band (typically r, but other bands may be used if they have higher signal-to-noise), and then using the model fit from the canonical band (convolved with the appropriate PSF) to obtain fluxes in each filter. Additionally, we also apply flux limits based upon an object's `fiber2mag` values; i.e., the total flux within a 2" diameter of the object center, corresponding to the aperture of a *BOSS* spectroscopic fiber (Smee et al., 2013a), after convolving the imaging data to achieve a standard 2" seeing.

3.3.2 WISE

The *eBOSS* LRG target selection algorithm also relies on infrared photometry from the Wide-Field Infrared Survey Explorer (*WISE*; Wright et al., 2010a). *WISE* observed the full sky in four infrared channels centered at 3.4, 4.6, 12, and 22 microns, which we refer to as W1, W2, W3, and W4, respectively. For *eBOSS* LRGs, we use the W1 band only. *WISE* magnitudes are commonly measured in the Vega system, but we convert to the AB system for LRG selection.[2] Over the course of its primary mission and the 'NEOWISE post-cryo' continuation, *WISE* completed two full scans of the sky in the W1 and W2 bands. Over 99% of the sky has 23 or more exposures in W1 and W2, and the median coverage is 33 exposures. We use the 'unWISE' forced photometry from Lang et al. (2014), which photometered custom coadds of the *WISE* imaging at the positions of all *SDSS* primary sources. Using forced photometry allows accurate flux measurements to be obtained even for significantly blended sources, including objects below the significance threshold for *WISE* only detections. Since the *WISE* W1 point-spread function is relatively broad (6.1 arc-seconds FWHM, ∼4 times larger than typical *SDSS* seeing), many sources are blended and forced photometry presents substantial advantages. Additionally, forced photometry allows us to leverage the relatively

[2] $W1_{\rm AB} = W1_{\rm Vega} + 2.699$

deep *SDSS* photometry to measure fluxes of *WISE* sources that are otherwise below the detection threshold. The same canonical morphological model is used in fitting photometry of the optical SDSS and infrared WISE images, therefore consistently measuring colors across all bands. Using unWISE photometry instead of the Wright et al. (2010a) *WISE* catalog increases the size of the resulting *eBOSS* LRG sample by $\sim 10\%$.

3.4 SELECTION OF HIGH-Z LRGS

Our overall goal is to cleanly select a sample of LRGs at redshifts beyond 0.6. In this redshift regime, however, optical photometry alone becomes insufficient for discriminating these high-z objects from foreground stars in our galaxy because both LRGs and red stars occupy the same region in optical color-color space. It is also not feasible to separate stars from galaxies reliably based on morphological information, as the S/N of *SDSS* photometry for these objects is low. Prakash et al. (2015) presented a new technique which eliminates almost all stellar contamination by combining both optical and infrared imaging data and applying a simple cut in optical-infrared color-color space. This takes advantage of the prominent 1.6 micron 'bump' in the spectral energy distributions of LRGs and other objects with old stellar populations (John, 1988), which results from the minimum in the opacity of H^- ions. The lowest wavelength channel of the *WISE* satellite is centered at 3.4 microns, almost perfectly in sync with the bump at $z \sim 1$. In that chapter, the authors presented a variety of optimization tests based on CFHTLS photometry, DEEP2 spectroscopic redshifts and COSMOS photometric redshifts. This method has been adapted here to meet *eBOSS* requirements specifically.

Figure 3.1 shows both stars and galaxies in a plot of r-*W1* verses r-i color, where *W1* indicates the magnitude of a source in the *WISE* 3.4 micron pass-band (on the AB system) and r and i indicate *SDSS* model magnitudes in the appropriate passband. Stars separate increasingly from the galaxy population in near-IR/optical color space as redshift increases, allowing clean discrimination of galaxies at $z > 0.6$ from stars. Simultaneously, r-i color increases with increasing redshift (particularly for intrinsically red galaxies) as the 4000 Å

break shifts redward, allowing a selection specifically for higher-redshift objects. While the combination of optical and IR imaging provides an excellent means of removing stellar contamination from an LRG target sample, this approach also means that we are limited to objects that are detected by *both SDSS* and *WISE*. The detections are performed on the optical *SDSS* images only. *WISE*-only detections have not been utilized in the *eBOSS* target catalogs.

As a basic color selection for characterizing potential *eBOSS* LRG targets, we select all objects that satisfy the criteria

$$r\text{-}i > 0.98, \text{and} \tag{3.1}$$

$$r\text{-}W1 > 2.0 \times (r - i), \tag{3.2}$$

where all magnitudes are corrected for Galactic extinction. These cuts were determined by examining the location of objects of known redshift and rest-frame color in color-color space, as in Figure 3.1. The clear separation seen between the locuses of stars and galaxies in Figure 3.1 is not so clear when a similar plot is made using *SDSS* photometry because of the latter's lower signal-to-noise ratio; i.e., the gap between the two populations is partially filled in due to objects with noisy measurements. The selection cuts above were optimized by assessing a figure of merit (FOM) which is a linear combination of fraction LRGs and low redshifts galaxies selected by a given color selection and normalized by the total number of selected objects (see Equation 9 and Figure 11 of Prakash et al. (2015)). Hence, we are primarily optimizing for the purity of the sample, rather than its completeness. Further details on the motivation for this selection and various tests on optimization can be found in Prakash et al. (2015).

To test this new selection technique, we targeted 10,000 objects satisfying this selection in a *BOSS* ancillary program in 2012-2013 (see the Appendix of Alam et al., 2015). Selection was limited to objects with $z_{Model} < 20$; 98% of the spectra yielded secure redshift measurements. These redshift estimates were found to be reproducible when observed multiple times. An additional 5,000 LRGs were selected by relaxing the r-i color requirement to r-$i > 0.85$ in order to estimate the number of LRGs missed by the color cuts in Equation 3.1. The

distributions of observed colors as a function of redshift for the resulting sample of 15,000 LRGs is presented in Figure 3.2.

Our method of combining optical and infrared photometry for this selection is unique; however, the specific choice of color cuts is not. We are able to cleanly select similar samples of LRGs by using different color combinations; e.g., *r*-*W1* and *r*-*z*, or *i*-*W1* and *i*-*z*. As can be seen in Figure 3.2, incorporating multiple colors can improve the efficiency of identifying true LRGs in the redshift range of interest by rejecting lower-redshift objects. We tested two parallel selection algorithms with different color selections in another *BOSS* ancillary program, *SEQUELS*, to select the algorithm which is best poised to meet *eBOSS* requirements, as described in Section B.0.1. The purity of the *eBOSS* sample (i.e., the level of contamination by stars and low-z galaxies) will be discussed further in Section 3.6.

3.5 THE *EBOSS* LRG TARGET SELECTION ALGORITHM

In this section, we describe in detail the final selection algorithm for *SDSS-IV*/*eBOSS* LRGs. At the redshifts of the LRGs ($z > 0.6$), the 4000 Å break moves into the *SDSS* i-band. As some objects are too faint to be detected by SDSS imaging in the r band, flux measurements can occasionally be negative; by making color cuts in flux space rather than magnitude space, this poses no problems. However, for convenience, we describe the selection algorithm and flux limits in terms of extinction-corrected AB magnitudes and colors here.

To summarize our selection methods: We first employ photometric processing flags to eliminate those objects with problematic imaging.[4] To ensure robust selection while maintaining a sufficient signal-to-noise ratio in *eBOSS* spectra, we also apply a variety of flux limits. Finally, to maximize the fraction of targets that are in fact high-redshift LRGs, we apply several color cuts. In the following sub-sections, we detail all the selections used for creating the *eBOSS* LRG sample.

[4] https://www.sdss3.org/dr8/algorithms/photo_flags_recommend.php

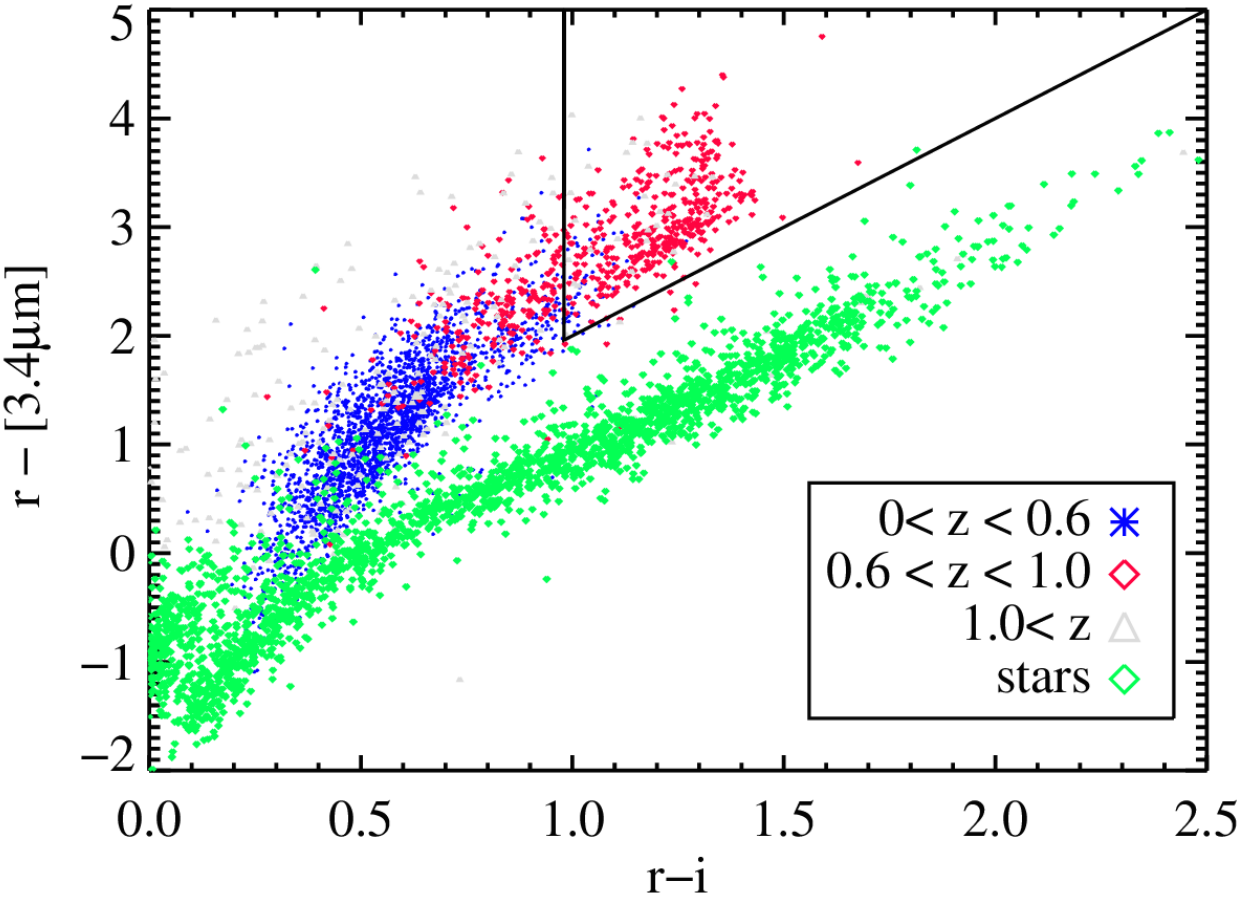


Figure 3.1: Optical/infrared color-color plot for galaxies observed by WISE and CFHT LS with photometric redshifts from the COSMOS survey. Blue symbols represent galaxies with photometric redshifts of $z < 0.6$, red diamonds represent galaxies at $0.6 < z < 1.0$, and cyan triangles represent galaxies at $z > 1.0$. Stars are represented by green diamonds. The triangular area depicts the broad selection presented in Equations 3.1 and 3.2. Photometric redshifts are taken from the COSMOS photo-z catalog of Ilbert et al. (2008b) and optical photometry is from the catalog of Gwyn (2011b), transformed to SDSS passbands. The conversion relation can be found on the CFHT LS webpage.[3]

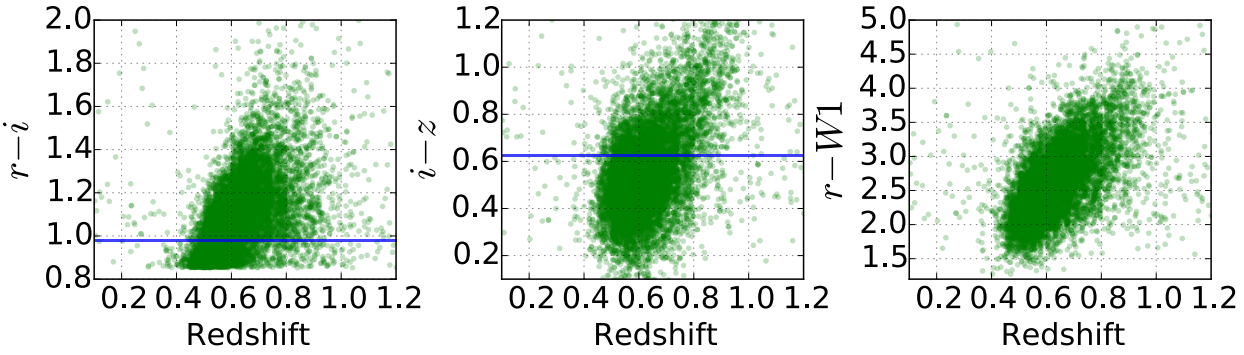


Figure 3.2: Plots of $r-i$, $i-z$ and $r-W1$ color as a function of redshift for 15,000 LRGs targeted via a *BOSS* ancillary program (see the Appendix of Alam et al., 2015), which utilized broader selection criteria than those used for *eBOSS*. The blue lines represent the cuts applied as part of the *eBOSS* target selection algorithm. Selecting objects with r-i > 0.98, i-z > 0.625 and $r-w1 > 2*(r-i)$ rejects a significant number of $z < 0.6$ galaxies while missing relatively few $z > 0.6$ LRGs.

3.5.1 Photometric flags for the LRG sample

Since many of the *SDSS* imaging runs overlap on the sky, an object may be observed twice or more (Stoughton et al., 2002). Only one observation is designated as the `primary` observation of the object during the `resolve` process. Hence, to exclude duplicate objects we enforce the following logical condition on the `RESOLVE_STATUS` bit-mask:

$$(Resolve\_status \ \& \ Survey\_primary) \neq 0. \tag{3.3}$$

3.5.2 Magnitude limits

The median 5-σ depth for photometric observations of point sources in the *SDSS* is $u = 22.15$, $g = 23.13$, $r = 22.70$, $i = 22.20$, $z = 20.71$ (Dawson et al., 2016). Additionally, we require a detection of the flux in the *W1* forced photometry for an object to be targeted. Keeping these requirements in mind, we apply the following flux limits to the entire sample:

$$MODEL\_IVAR_{r,i,z} \neq 0, \tag{3.4}$$

$$z_{Fiber2} \leq 21.7, \tag{3.5}$$

$$19.9 \leq i_{Model} \leq 21.8, \tag{3.6}$$

$$z_{Model} \leq 19.95, \tag{3.7}$$

$$W1_{vega} \neq 0, \text{and} \tag{3.8}$$

$$W1_{AB} \leq 20.299, \tag{3.9}$$

where `MODEL_IVAR` are the inverse variances on the model fluxes in r, i, and z bands. Equation 3.4 implies that errors in flux measurement are physically meaningful. The application of Equation 3.5 serves to maintain a sufficiently high signal-to-noise ratio of the *eBOSS* spectra. This cut is similar in spirit to the i_{Fiber2} cut that was used for the *BOSS* CMASS galaxy sample (Eisenstein et al., 2011). We apply the lower limit defined in Equation 3.6 in order to avoid targeting $i < 19.9$ *BOSS* CMASS galaxies, which generally lie at lower redshifts and have been observed previously. $W1_{vega}$ being nonzero implies that the photometry is reliable, while Equation 3.9 ensures that *WISE* flux measurements have a signal-to-noise ratio greater than 5 (Wright et al., 2010a). The i and z faint magnitude limits are set to achieve the required target density of $\sim$60 targets deg^{-2} matching the *eBOSS* fiber allocation for LRGs (Dawson et al. 2015), while maximizing the brightness of targets. LRGs are given lower priority for selection than the other main target class, QSOs, and hence a non-negligible fraction cannot be targeted due to fiber collisions. As a result, we must select 60 targets per square degrees to end up with 50 LRG targets per square degree placed on fibers.

3.5.3 Color Selection

We use the *r-W1* (optical-infrared) color for separating LRGs from stars.[5] The optical colors of galaxies are used to ensure that the targeted objects are intrinsically red and lie in the

[5]Note that we do not explicitly use any morphological cuts, but rather separate stars and galaxies based only on their colors.

desired redshift range. We thus apply the following three selection criteria:

$$r - i > 0.98, \tag{3.10}$$

$$r - W1 > 2.0 \times (r - i), \text{and} \tag{3.11}$$

$$i - z > 0.625. \tag{3.12}$$

Equations 3.10 and 3.11 represent the basic LRG color selection discussed at the beginning of Section 3.4 and are identical to Equations 3.1 and 3.2, the color curs used in initial tests of LRG selection. We use Equation 3.12 to reduce contamination from $z < 0.6$ galaxies.

The overall *eBOSS* LRG selection algorithm is shown schematically as a flow chart in Figure 3.3. The details of this algorithm were optimized based upon a pilot survey, the *Sloan Extended Quasar, ELG and LRG Survey* (*SEQUELS*), which is summarized in the appendix of Alam et al. (2015); the *SEQUELS* LRG selection algorithm is detailed in an Appendix (see Section B.0.1).

In addition to the LRGs targeted by Equations 3.3–3.12, we target a small number of objects, ∼200 over the 10,000 deg^2 SDSS imaging area, via a different but related algorithm. These objects have $i_{Model} \geq 21.8$ and are designated `LRG_IDROP`. These are not significant for BAO studies but constitute a separate sample designed to identify rare objects at extremely high redshifts. Further details are provided in an Appendix to this chapter (see Section A).

3.6 TESTS OF THE TARGET SELECTION ALGORITHM

In this section, we assess the results of our target selection methods using the current *eBOSS* data. We use the automated spectral classification, redshift determination, and parameter measurement pipelines of *SDSS-III BOSS* which are described in Bolton et al. (2012), to reduce and analyze spectra of *eBOSS* targets. To assess the true redshifts of LRG sample, we have conducted a visual inspection of a subset of *eBOSS* spectra, employing the `idlspec2d` package for this purpose.[6]

[6]http://www.sdss3.org/dr8/software/products.php

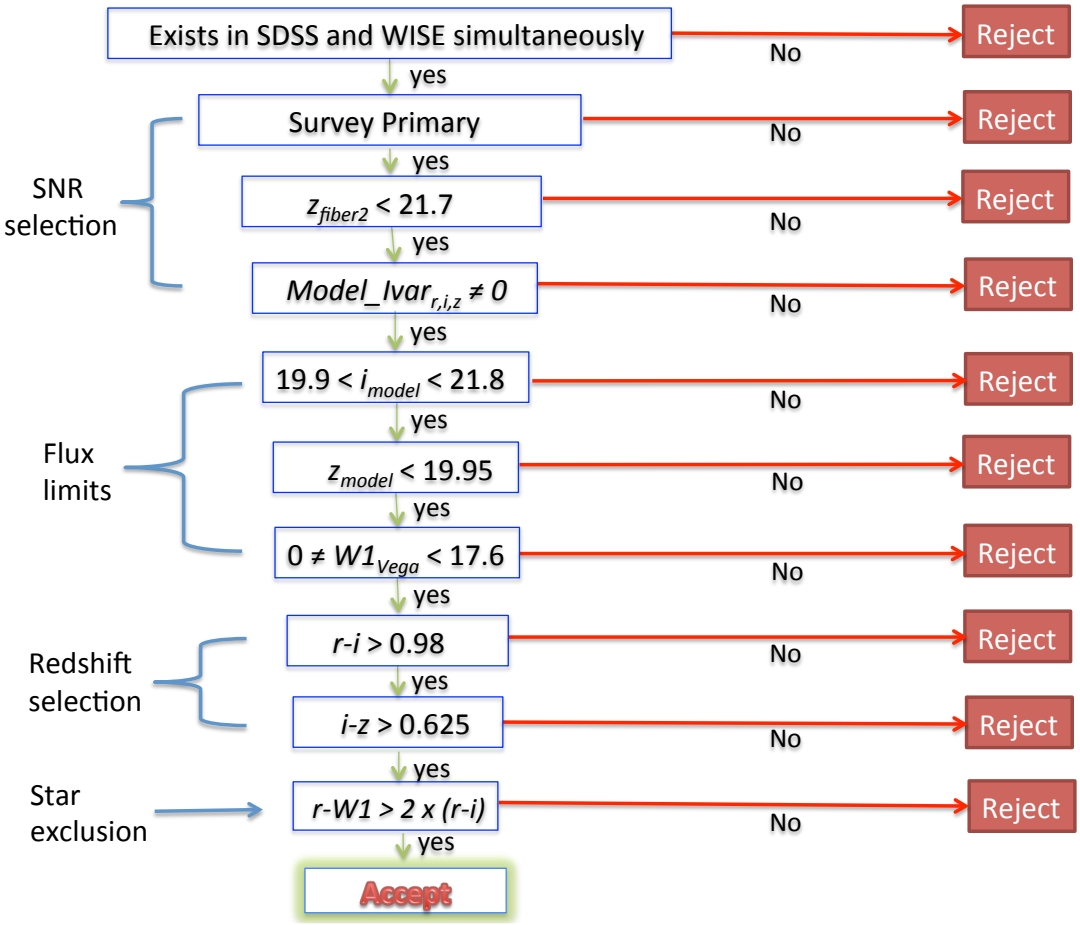


Figure 3.3: Schematic flow chart for the *eBOSS* LRG target selection algorithm. All quantities are corrected for Galactic extinction. Refer to the text for a full description of all of the quantities shown in this figure.

Specifically, we present results based on 2,557 LRG candidates from eight plates that were visually inspected to assess the quality of spectra and robustness of redshift measurements by a team of *eBOSS* members. Each plate was inspected by multiple individuals to cross-check the results. Visual inspectors selected what they believed to be the best estimate of the correct redshift for each spectrum, as well as assessing the security of that redshift according to a simple four level confidence metric, `z_conf` (confidence of inspector in the measured redshift). Targets are assigned `z_conf` values of 0 to 3, with 2 and 3 corresponding to measurements which were believed to be robust. A value $z_{\rm conf} = 1$ denotes a spectrum that is ambiguously classified, i.e., where more than one of the chi-squared minima correspond to models which are a possible fit, while `z_conf` $= 0$ is used for objects where it is not possible to classify the objects and establish their redshift. These objects are considered unreliable and not used in the calculations of redshift distributions or related quantities.

In the remainder of this section, we briefly present the expected basic characteristics of the *eBOSS* LRG sample (e.g., its redshift distribution, spectral quality, and redshift success) derived from this sample with visual inspections. We also test the efficiency of our target selection algorithm against the science requirements for the *eBOSS* LRG sample as described in Section 3.2.

Two redshift distributions are presented in Table 3.1. The more conservative estimate (the one with a higher rate of "Poor spectra") assumes that only objects given `z_conf` > 1 have been assigned a correct redshift. The less conservative estimate includes all objects with `z_conf` > 0; this is a relevant scenario, since it is likely that a great majority of `z_conf` $= 1$ redshifts are correct, but will inevitably include at least some incorrect redshifts. It is likely that the true distribution lies between these two bounds. It is expected that pipeline improvements now underway will enable at least some redshifts currently assigned `z_conf` < 2 to be recovered automatically in the future.

As can be seen in the Table 3.1, even with the conservative scenario (`z_conf` > 1), the *SDSS* spectral pipeline generates a secure redshift solution for $\sim 89\%$ of the LRG candidates visually inspected. However, the fit determined to be correct via visual inspection sometimes does not correspond to the minimum chi-squared solution from the pipeline,

but rather an alternative chi-squared minimum.[7] Pipeline improvements now under way (which include both improved two-dimensional extractions and reductions in the freedom of template+polynomial fitting) are expected to improve the automated redshift-finding, so this figure should be a floor to the actual performance of *eBOSS* LRGs.

The remaining $\sim 11-12\%$ of the LRG targets without a secure redshift determination typically have spectra with low signal-to-noise ratios. An additional $\sim 9\%$ of the LRG targets are found to be stars. These two factors (low S/N and stellar contaminants), in combination, make it impossible for this sample to meet the eBOSS-only LRG requirement that that 80% of all targets be LRGs within the range $0.6 \lesssim z \lesssim 1.0$, even before the redshift distribution of the galaxies is considered. In the end, 68–72% of all LRG targets are in fact galaxies with definitive redshift measurements that lie in the desired regime. For detailed discussion of the pipeline results, visual inspections, templates and sources of redshift failures, see Dawson et al. (2016).

In Figure 3.4, we present the overall redshift distribution ($N(z)$) of the visually-inspected *eBOSS* LRGs. Although we fail to meet the requirement of 80% efficiency at targeting $0.6 < z < 1.0$ LRGs, our target selection algorithm still exceeds the median redshift requirement, which is calculated only for actual galaxies (and hence includes only non-stellar targets with robust redshift measurements). In Figure 3.5, we show examples of LRG spectra across the redshift range of interest for *eBOSS*. There is an excellent match between the measured SEDs and the templates, confirming the robustness of these redshift measurements. The *eBOSS* LRG sample can be augmented with $z > 0.6$ *BOSS* CMASS LRGs to meet our requirements on the total number of LRG redshifts within the range $0.6 < z < 1.0$; as a result, we still expect to achieve a 1% measurement of the LRG BAO scale at $z \sim 0.7$, even though the LRG sample falls short of its requirements.

[7] The *SDSS* pipeline generates a set of possible fits; cf. Bolton et al. (2012).

Table 3.1. Redshift distribution of *eBOSS* LRGs based upon visual inspection of spectra.

	LRGs `z_conf` > 0	**LRGs** `z_conf` > 1
Poor spectra	4.0	6.7
Stellar	5.3	5.3
Galaxy	N/A	N/A
$0.0 < z < 0.5$	0.6	0.6
$0.5 < z < 0.6$	6.2	5.9
$0.6 < z < 0.7$	**15.2**	**14.8**
$0.7 < z < 0.8$	**15.3**	**14.7**
$0.8 < z < 0.9$	**9.4**	**8.7**
$0.9 < z < 1.0$	**3.2**	**2.7**
$1.0 < z < 1.2$	0.6	0.5
Targets	60	60
Total Tracers	43.1	41.0

Note. — Redshift distribution of *eBOSS* LRGs, based upon results for a sample of 2,557 visually inspected spectra. The surface densities are presented in units of $\deg^{-2}$, normalizing to the total surface density of the parent sample for these spectra. Entries highlighted in bold font denote the subset of the sample that lies in the redshift range used to assess the high-level science requirements for the LRG sample.

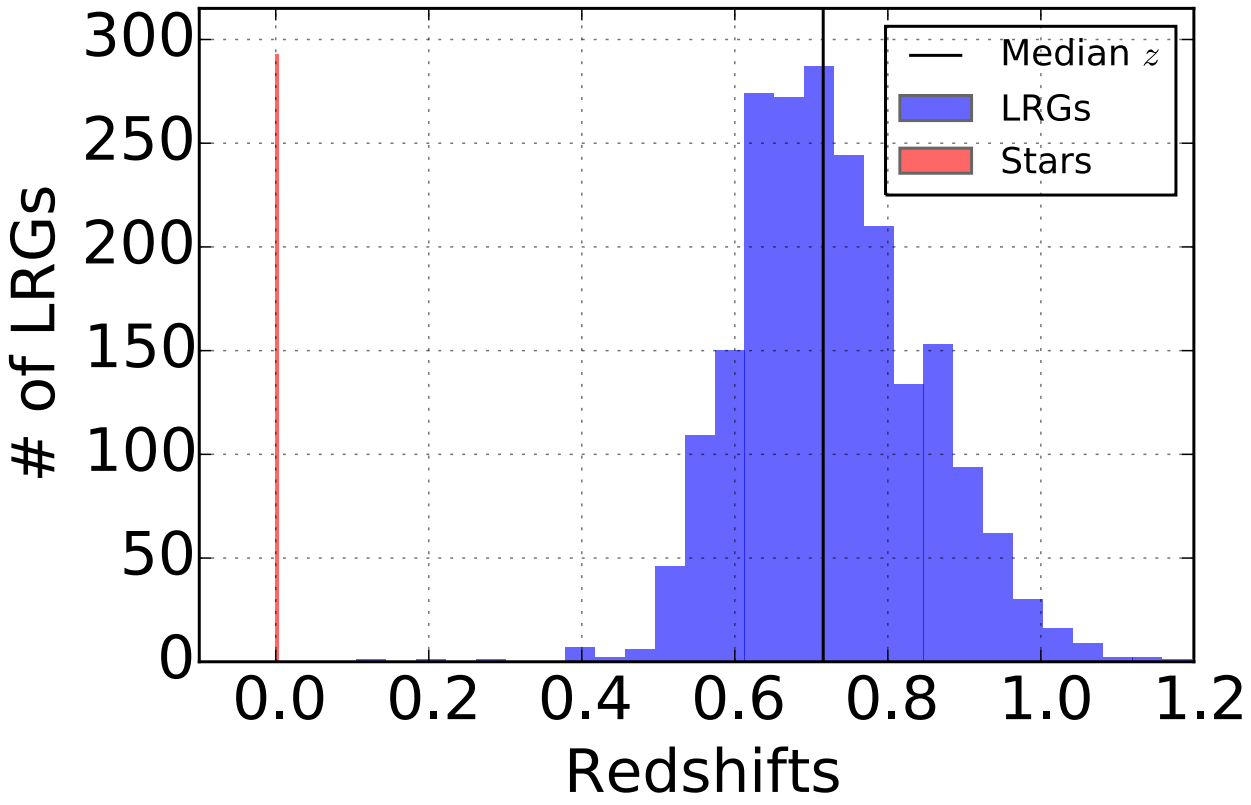


Figure 3.4: Redshift histogram of 2,119 visually inspected LRGs (blue bar) observed with *eBOSS*. The median redshift of confirmed galaxies is 0.712 (black line), with 9% stellar contamination (red bar). We use only objects with secure redshifts (`z_conf` > 1) here.

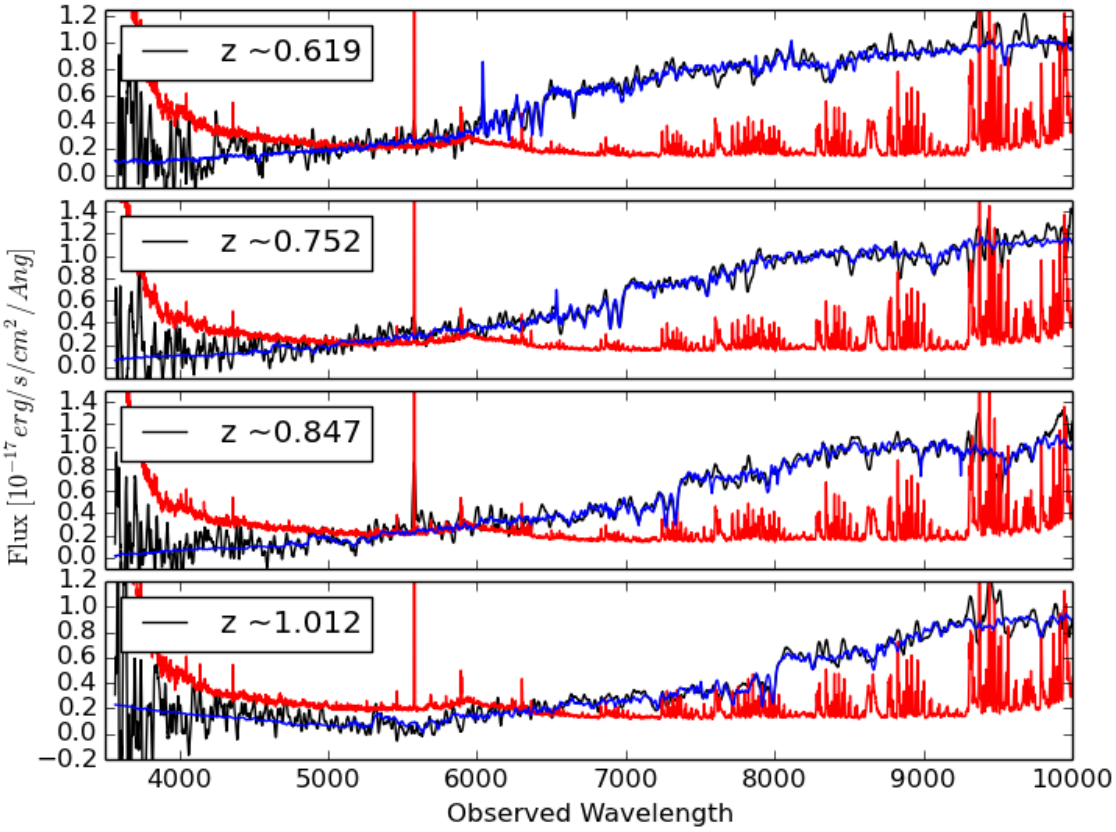


Figure 3.5: Representative spectra of galaxies from the *eBOSS* LRG sample, smoothed with a 21 pixel boxcar kernel. Shown are four LRGs covering the entire redshift regime of $0.6 \lesssim z \lesssim 1$. Flux errors are plotted in red while the template model fits are in blue. Black curves depict the observed spectra.

3.7 TESTS OF HOMOGENEITY AND IMPLICATIONS FOR LARGE-SCALE CLUSTERING MEASUREMENTS

As discussed in Section 3.2.1, we require that the target sample be highly uniform to prevent non-cosmological signals from contaminating clustering measurements. Exploring systematics that can affect the inferred clustering of targets is often considered only when survey data is used for science analyses. We instead have investigated these issues while exploring target selection methods, enabling more informed decisions regarding survey strategy. For instance, foreknowledge of which areas of the survey may pose problems for controlling clustering measurements potentially allows the survey footprint to be shifted.

We assess the uniformity of the target sample by comparing the observed density of targets to maps of local imaging conditions and Galactic structure. We apply a regression analysis of surface density against a broad set of tracers of potential systematics; the intention is similar to, e.g., Scranton et al. (2002); Ross et al. (2011a); Ho et al. (2012); Leistedt et al. (2013); Giannantonio et al. (2014), but unlike those works, we simultaneously fit for the impact of a wide variety of systematics rather than correlating against one at a time. This has the advantage of producing a model of systematic-affected density that will provide accurate predictions for the combined effects of all the systematics considered, even if the input systematic maps are covariant with each other (as, for instance, stellar density and dust extinction must inevitably be).

We focus on systematics associated with imaging data characteristics or with known astrophysical effects such as dust extinction and stellar density. Using the results of the regression analysis (described below) we assemble maps of the observed density and the predicted density. We identify regions within our footprint where the total span of target density fluctuation is less than 15%, and consider the portion of sky with larger variations to be contaminated at an unacceptable level; this criterion is based on prior experience with the level of systematics that may be corrected reliably in BOSS Ross et al. (2011a). We note that fluctuations in density within the final $0.6 < z < 1.0$ LRG catalog are likely to be smaller than this, as once spectra are obtained, stars and redshift outliers can be removed; such objects are naturally expected to be less homogenous over the SDSS survey area than

the true LRGs.

3.7.1 Homogeneity of *eBOSS* LRG targets

To begin, we identify a broad set of imaging parameters that could affect *eBOSS* target selection:

1. `W1covmedian`: The median number of single-exposure frames per pixel in the WISE W1 band.
2. `moon_lev`: The fraction of frames that were contaminated with scattered moon light in the WISE W1 band.
3. `W1median`: The median of accumulated flux per pixel in the WISE W1 band measured in units of DN (data number).[8].
4. **Galactic Latitude**: used as a proxy for stellar contamination.
5. **Galactic extinction**: We use r band extinction, as given by SFD.
6. **FWHM** in the *SDSS* z-band: We use FWHM as an estimate of the 'seeing' or imaging quality for the SDSS imaging.
7. **SKYFLUX** in the *SDSS* z-band: the background sky level affects the detection of faint objects is more difficult in the brighter regions of the sky .

We create maps of the WISE systematics over the entire *SDSS* footprint using the metadata tables associated with the Atlas images and source tables provided by WISE survey team; `W1covmedian`, `W1median`, and `moon_lev` are all quantities in these tables.[9] We use the seeing and the sky-background in the z band since the *eBOSS* LRG selection algorithm is flux-limited in that bandpass filter. Due to the scan strategy of SDSS, the seeing and sky background in other SDSS bands should correlate strongly with this quantity, making the use of multiple filters' quantities redundant.

Next, we break the sky up into equal-area pixels of 0.36 deg^2 and weight all pixels equally. The observed density, SD_{obs}, in each pixel can be expressed as a combination of a mean level, the impact of all of the systematics, and random noise:

[8] The accumulated photons in each pixel are represented by a number in units of DN
[9] http://wise2.ipac.caltech.edu/docs/release/allsky/expsup\/sec2_4f.html

$$SD_{obs} = S_0 + \sum_{i=1}^{7} S_i \times x_i + \epsilon, \tag{3.13}$$

where S_0 is the constant term representing the mean density of objects in each pixel, S_i are the coefficients for the values of each individual source of potential systematics fluctuations in that pixel (x_i), and ϵ represents the combined effect of Poisson noise (or shot noise) and sample/cosmic variance in that pixel. For larger pixels such that the mean pixel target density is $\sim$15 or more, the Poisson noise can be approximated as a Gaussian. Under these conditions, multi-linear regression provides an effective means of determining the unknown coefficients, S_0 and S_i. We derive a best-fit model based on minimizing the value of reduced-χ^2 (χ^2 per degree of freedom). We have explored larger or smaller pixelizations and find that our results are unchanged.

The coefficients obtained from this multi-linear regression are then used in combination with the maps of potential systematics to predict the target density across the whole footprint, producing a statistic that we will refer to as the *Predicted Surface Density* or PSD. We also define a *Residual Surface Density*, or $Residual\_SD_j$, for any particular systematic as the difference between SD_{obs} and the $Reduced\_PSD_j$ (which is calculated by omitting the j'th systematic term in calculating the PSD). This quantity should be linear in systematic j with a slope corresponding to S_j if our linear regression model is appropriate to the problem. To summarize our formalism:

$$PSD = S_0 + \sum_{i=1}^{n} S_i \times x_i, \tag{3.14}$$

$$Reduced\_PSD_j = PSD - S_j \times x_j, \text{and} \tag{3.15}$$

$$Residual\_SD_j = SD_{obs} - (PSD - S_j \times x_j), . \tag{3.16}$$

where the j index indicates a single systematic of interest.

3.7.2 Predicted Surface Density for *eBOSS* LRG targets

The PSD is highly useful for testing the uniformity of the target sample across the whole footprint, enabling comparisons to survey requirements. We find that the effects of systematics produce significantly different best-fit models (in terms of both the mean density and the coefficients for each systematic) in the areas of *SDSS* imaging around the Northern Galactic Cap (NGC) and the Southern Galactic Cap (SGC). However, for both the regions considered independently, multi-linear regression provides an acceptable best-fit model. Hence we analyze these regions separately.

The resulting regression fits are shown in Figure 3.6. In these plots, we plot the $Residual\_SD$ for each individual systematic which was been left out in calculating $Residual\_SD_j$. The data points plotted are averages over 4000 sky pixels in the NGC or 2000 sky-pixels in the SGC; the error bars represent the standard error on the mean for each point. The straight lines represent the prediction from the regression model for the impact of the systematic indicated on the x-axis, x_j (cf. Eqn. 3.13); i.e., we plot $y = S_j \times x_j$.

3.7.3 Analysis of regression results

Our regression analysis allows us to determine what fraction of the survey footprint satisfies the requirement of less than 15% total variation in target density (point 6 in Section 3.2.2). This 15% window is not necessarily symmetric around the mean, so we fix its limits such that the footprint area satisfying the requirement is maximized. The windows containing regions with PSD variation $< 15\%$ are overplotted on the histograms of predicted density in Figure 3.7. In the NGC, $\sim 97\%$ of the imaging area meets the *eBOSS* survey requirements for homogeneity. However, in the SGC, only $\sim 82\%$ of the area meets these requirements. At worst, these fluctuations will require that 8% of the total 7500 deg^2 eBOSS area is masked. However, these fluctuations may be reduced once spectroscopic redshifts are obtained; we will perform a similar analysis on the final spectroscopic sample in later work.

Differences in the observed number density between the NGC and SGC were found for the *BOSS* `CMASS` and `LOWZ` samples, and were analyzed in depth by Ross et al. (2012a). These differences matched the photometric offsets between the two regions determined by

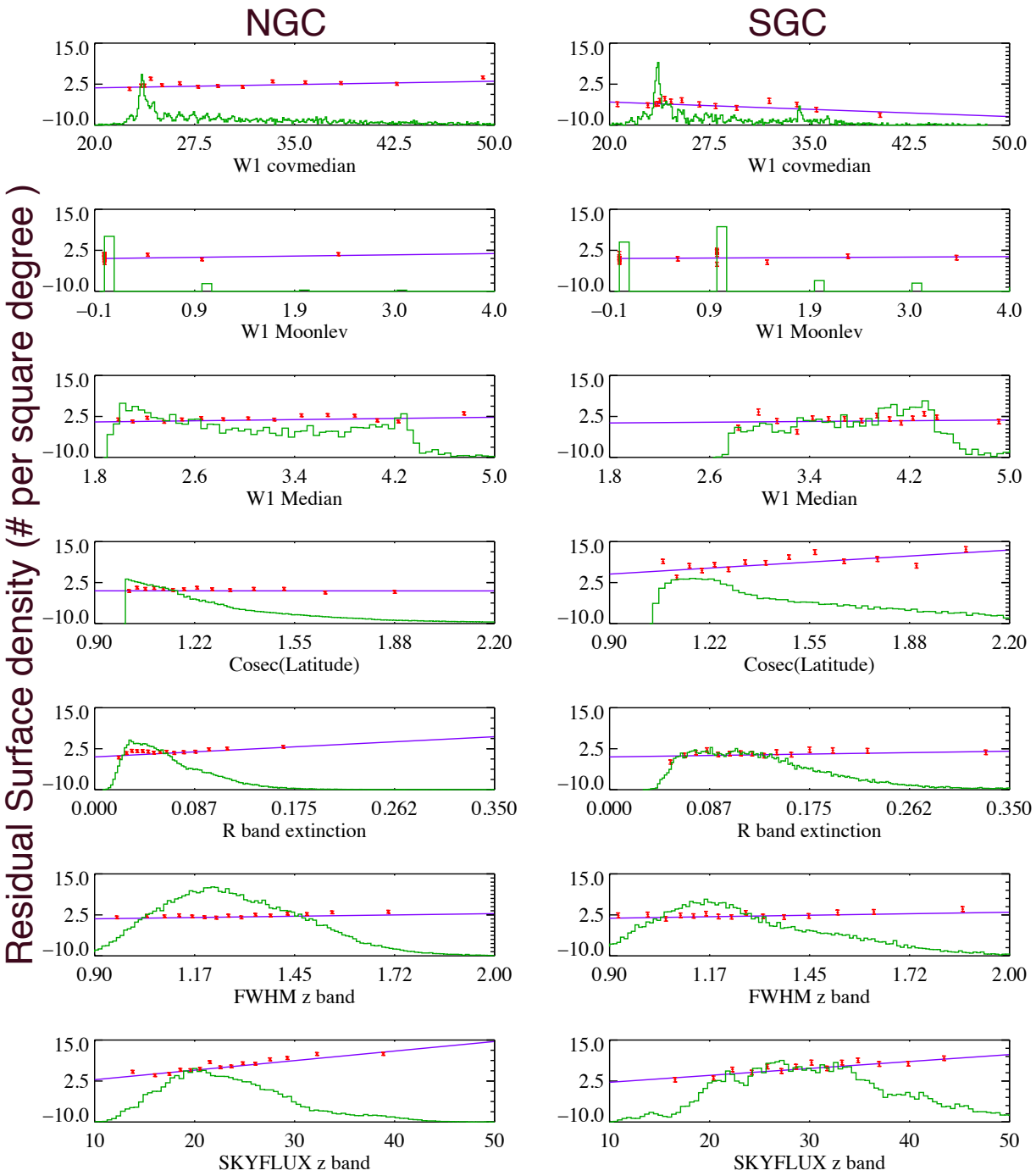


Figure 3.6: The residual surface density (number of targets deg^{-2}) from the regression model with a single systematic omitted as a function of the systematic left out, along with the corresponding predictions from the regression model, for all systematic maps considered. Individual points have been averaged over 4000 sky pixels in the Northern Galactic Cap (NGC) or 2000 pixels in the Southern Galactic Cap (SGC). The straight line shows the prediction of the regression model for the impact of the systematic indicated on the x-axis (cf. equation 3.13). The overplotted histograms show the distribution of pixel values, and correspond to the y axis at the right side of each plot. The left-hand column of plots are for pixels in the NGC, while the right-hand column of plots are for the SGC. A linear model appears to be appropriate for all systematics considered.

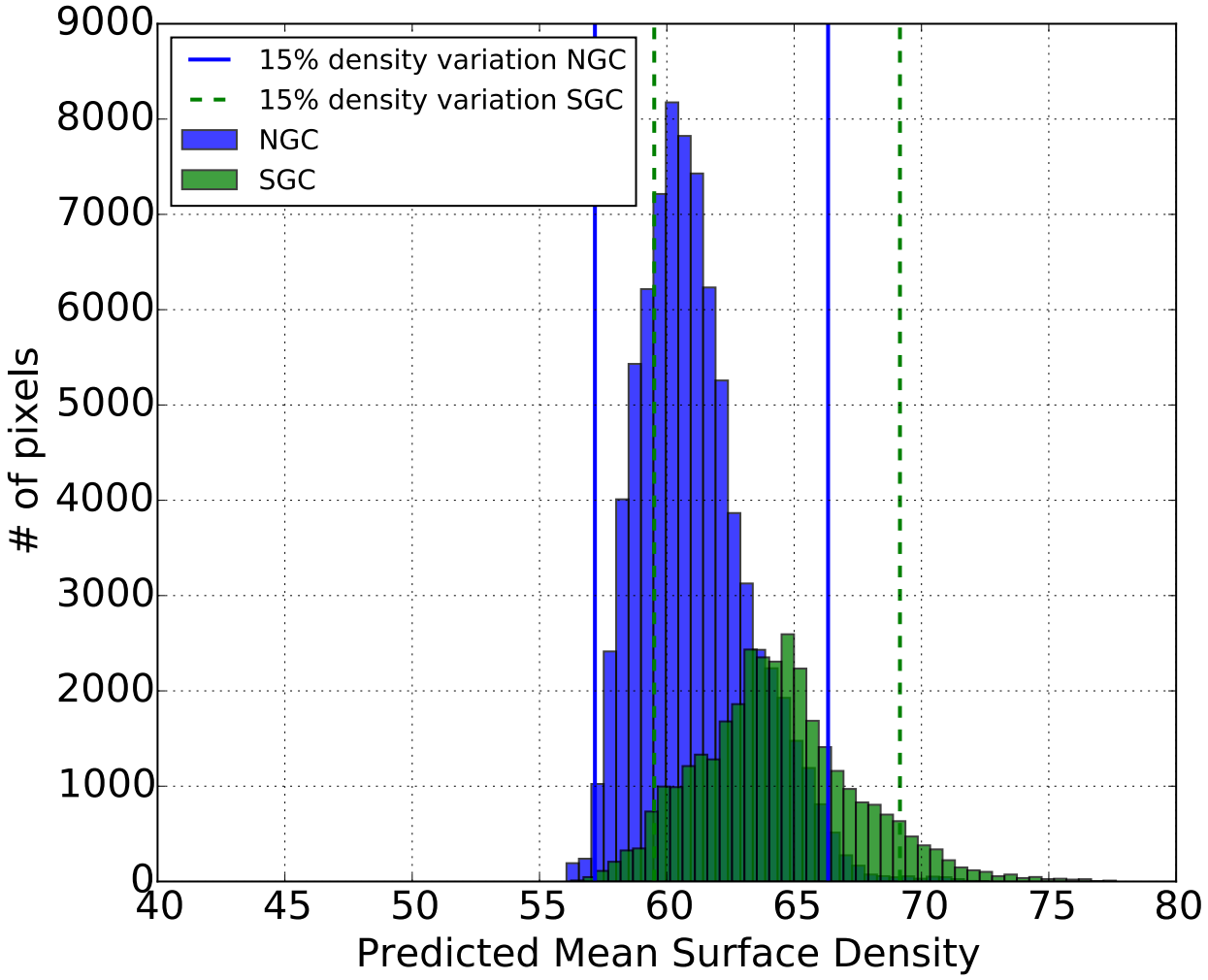


Figure 3.7: Histogram of the surface density predicted by the regression models described in Section 3.7.1. The blue bars represent the NGC, with solid blue lines depicting the 15% window within which samples are expected to be sufficiently homogeneous for robust large-scale-structure measurements. Similarly, the green bars represent the density in the SGC, with dotted green lines depicting the 15% window. We find that $\sim 97\%$ of the NGC footprint with *SDSS* imaging meets the homogeneity requirements of *eBOSS* (see Section 3.2.2). However, in the SGC, only $\sim 82\%$ of the possible *eBOSS* footprint meets these requirements.

Schlafly & Finkbeiner (2011a). These offsets have been incorporated into the re-calibrated photometry used for *eBOSS*; any difference in target density between the regions is therefore due to still-unknown differences between the two regions. This issue will require further investigation in future *eBOSS* studies.

Based on the regression model, we can assess which systematics are most strongly affecting target selection. We find that all of the potential WISE imaging systematics have relatively weak effects on the density of selected targets. This can be seen from the flatness of *Residual_SD* for these parameters in Figure 3.6. The most significant effects are associated with dust extinction, stellar contamination, and the *SDSS* sky background level, as seen from the steep slopes in Figure 3.6. It is unclear whether dust or stellar contamination is more fundamentally responsible for variations in density, since the two correlate with each other strongly. Given the variation in coefficients, it is likely that the same phenomenon is being ascribed more to dust in the NGC and to Galactic latitude in the SGC, and those differences in coefficient are not truly significant. Fortunately, the regression model will still predict the correct density from covariant variables such as these, regardless of which covariate is actually responsible.

We depict the observed surface density, the predicted surface density, and the mask of the survey across the whole footprint of *SDSS* in Figure 3.8.

3.7.4 Impact of zero point variations

We next assess the expected level of variation in target density due to errors in zero-point calibrations, which can then be compared to the targeting requirements. We investigate this by determining the fractional derivative in the number of targets selected (N) as we shift all magnitudes in a given band (m) by a constant amount – i.e., we calculate $\frac{1}{N}dN/dm$ – and then assess what impact this sensitivity has on target density. We find that zero-point errors of 0.01 magnitude in the *r*, *i*, *z*, and *W1* bands causes fractional changes of 2.26%, 2.5%, 6.24%, and 0.6%, respectively, in the target density of the LRG sample. Finkbeiner et al. (2014) estimate that the 1σ zero point uncertainties (σ_{zp}) after recalibration of SDSS are 7, 7, and 8 millimagnitudes in the *SDSS* *r*, *i*, and *z* bands respectively, while *WISE* calibration

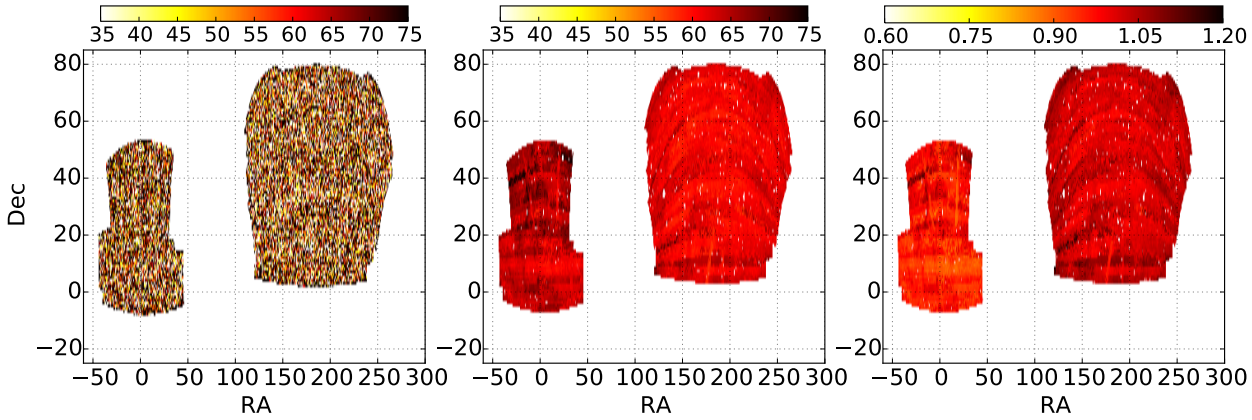


Figure 3.8: Observed surface density map of *eBOSS* LRGs over the area of the *SDSS* imaging footprint used to derive targets for the *BOSS* survey. *eBOSS* will target more than ∼375,000 LRGs over a ∼7500 deg^2 subset of this area, corresponding to a surface density of ∼50 deg^{-2}. The color scale in this panel is dominated by Poisson noise and sample/cosmic variance. The middle panel shows a similar plot based on the predicted density, while the third panel shows the regions which would be masked to reach *eBOSS* homogeneity requirements. Pixels shown with a shade that is fainter or darker than the typical are those which fail to meet the 15% target density variation requirement of Section 3.2.2.

uncertainties in the W1 band are approximately 0.016 mag (Jarrett et al., 2011a).

Assuming that zero point errors will be Gaussian-distributed, 95% of all points on the sky will be within $\pm 2\sigma$ of the mean zero point. Hence, the total fractional variation in density over that area will be $\frac{4}{N} \times \mid \Delta N/\Delta m \mid \times \sigma_{zp}$. We present the results of this calculation in Table 3.2 and Figure 3.9. For all bands but z, the impact of zero point variations on the density of LRG targets will be minimal. However, the estimated level of z band zero point uncertainty is sufficiently large that more than 13% of the *eBOSS* area will go beyond the 15% target density variation requirement. The impact of this variation and strategies for mitigating it will be explored in future *eBOSS* papers.

Table 3.2. Summary of variations in target density due to errors in *SDSS* imaging zero points.

Bands	Derivative of fractional density	RMS zero-point error	95% range of variation in fractional density
	$\mid \frac{1}{N}\Delta N/\Delta m \mid$	(σ_{zp})	$4\times \mid \Delta\bar{N}/\Delta Mag \mid \times\sigma_{zp}$
SDSS r	2.26	7×10^{-3}	0.063
SDSS i	2.5	7×10^{-3}	0.070
SDSS z	6.24	8×10^{-3}	0.199
*WISE W*1	0.60	16×10^{-3}	0.038

Note. — The impact of zero point uncertainties on the density of targets. The *eBOSS* LRG sample meets the requirement that density variations due to zero-point errors be less than 15% in the *SDSS* r and i bands, but fails to meet that criterion in the z band with current calibrations.

3.8 CONCLUSIONS

The LRG component of *SDSS-IV*/*eBOSS* will obtain spectroscopy of a sample of over 375,000 potential intrinsically luminous early-type galaxies at $z > 0.6$. Based on the initial set of *eBOSS* data, we find that the efficiency of this selection for selecting $0.6 < z < 1.0$ spectroscopically confirmed LRGs with secure redshift measurements is $\sim 68 - 72\%$. Although this is lower than the required level for eBOSS to achieve its science goals in isolation, once augmented with BOSS LRGs the required galaxy density should be attained. One reason the success rate does not approach 80% is that 9% of LRG targets prove to be stars, a result of noisy *SDSS* photometry. The sample is flux-limited to keep the selection algorithm robust, as well as to maintain a sufficient signal-to-noise ratio to enable the resulting LRG spectra to provide secure redshift measurements. The LRG sample is uniform and homogenous over $\sim 92\%$ of the *BOSS* survey footprint, showing little or no dependence on imaging systematics and flux calibrations. The remaining $\sim 8\%$ of the footprint will have to be assessed carefully for systematic effects before being included in cosmology measurements if they fall

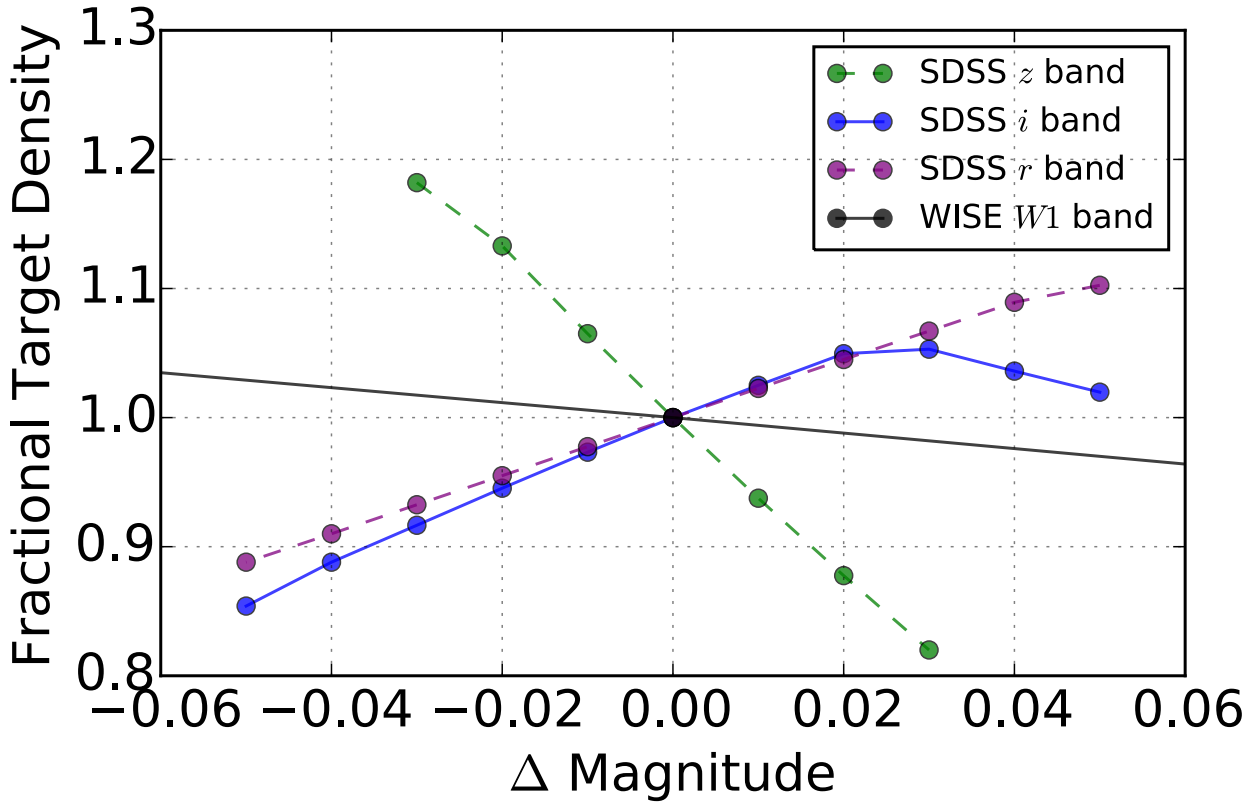


Figure 3.9: Change in target density as a function of an overall shift in all magnitudes in either the *SDSS* r, i, z, or the *WISE W1* band. Given the current level of zero point uncertainty in SDSS and *WISE* photometry, the LRG target selection is only sensitive to the uncertainty in the zero point of the z band.

within the *eBOSS* footprint.

The primary science drivers of the *eBOSS* LRG sample are to study the large-scale structure of the universe out to $z \sim 1$. With careful control of incompletenesses and selection effects, the *eBOSS* LRG algorithm will also provide a large sample for galaxy evolution studies of giant elliptical galaxies. The *SDSS-IV/eBOSS* LRGs will cover a volume either not probed, or not probed at high density, by *SDSS-III/BOSS*, and will enable both BAO and RSD measurements with a highly uniform set of luminous, early-type galaxies. The *SDSS-IV/eBOSS* LRG sample will provide a powerful extension of *SDSS-III/BOSS* for the study of structure and galaxy evolution at high redshifts.

4.0 THE SDSS-IV EXTENDED BARYON OSCILLATION SPECTROSCOPIC SURVEY: QUASAR TARGET ANALYSIS

The contents of this chapter have been published in Myers, Palanque-Delabrouille, Prakash, et al., 2015 December 2, *The Astrophysical Journal Supplement Series*, Volume 221, Number 2

4.1 INTRODUCTION

Quasars are the most luminous ($L_{bol} \sim 10^{47-48}$ erg s^{-1}) non-transient objects in the universe, and are expected to be powered by accretion of matter onto supermassive black holes at their center. Almost half a century has elapsed since quasars were discovered as bright, bluer than main sequence stars, unresolved, extragalactic sources in optical imaging (Schmidt, 1963). Ever since, a number of imaging surveys have employed a UV-excess (UVX) criterion, visible in simple optical color cuts, to provide algorithms for selecting quasars (e.g., Sandage & Luyten, 1969; Green et al., 1986; Boyle et al., 1990). This approach, which mainly targets quasars at redshifts around $0.5 < z < 2.5$, has greatly increased the number of spectroscopically confirmed quasars through subsequent spectroscopic surveys, like the 2dF QSO Redshift Survey (Croom et al., 2005), and the 2dF-SDSS LRG and QSO Survey (Croom et al., 2009).

UVX approach was further modified to enable quasar targeting over all of optical color space beyond the stellar locus, and rather than just the blue side (e.g. Warren et al., 1987; Kennefick et al., 1995; Newberg & Yanny, 1997). This extended the selection of large numbers of quasars to $z > 2.5$. This methodology was effectively employed by the *Sloan Digital Sky*

Survey (*SDSS*; York et al., 2000b) on imaging taken using a new *ugriz* filter system (Fukugita et al., 1996b). *SDSS* eventually spectroscopically confirmed an unprecedentedly large sample of over 100,000 quasars (Richards et al., 2002; Schneider et al., 2010) as part of the *SDSS-I* and *II* surveys.

Detecting BAO features using quasars as direct tracers of matter density has been a challenge due to their low space density. For this reason, most of the surveys use quasars to study the distribution of neutral hydrogen in the universe at high redshifts, usually called Lyman-α forest. *BOSS* spectroscopically identified $\sim$ 170,000 new quasars of redshift $2.1 \leq z < 3.5$ to a depth of $g < 22$ with a goal to study the Lyman-α forest (Dawson et al., 2013b).

Mid-IR colors provide a powerful mechanism for separating stars from galaxies. This is also true for quasars that otherwise resemble stars in optical color space (e.g. Stern et al., 2012; Assef et al., 2013; Yan et al., 2013). This enables us to use Wide-field Infrared Survey Explorer (*WISE*; Wright et al., 2010b) data to extract additional information for targeting quasars. Due to the addition of new and extensive multi-wavelength and multi-epoch imaging it has been possible to target many quasars which may have been missed by previous surveys.

In combination, *SDSS-I/II/III* targeted quasars at $2.1 \lesssim z \lesssim 4$ to a magnitude limit of $g < 22$ or $r < 21.85$ (Ross et al., 2012b) and quasars at all redshifts to $i < 19.1$[1] (Richards et al., 2002). There remains an obvious, highly populated discovery space using *SDSS* imaging data—namely, $z < 2.1$ quasars fainter than $i = 19.1$.

The potential of *SDSS* and other imaging for targeting new quasars provides an exciting opportunity to use BAO as direct tracers of the underlying matter density and to measure the expansion of the Universe (Seo & Eisenstein, 2003; Lin & Mohr, 2003). No strong BAO constraint currently exists in the redshift range $1 \lesssim z \lesssim 2$, and BAO measurements at yet higher redshift remain a particularly potent constraint on the evolution of the angular diameter distance, $d_A(z)$ and of the Hubble Parameter, $H(z)$ (Aubourg et al., 2014). These factors led to the conception of *eBOSS*(Dawson et al., 2016) as part of *SDSS-IV*.

As mentioned before, it has been difficult to detect BAO features using quasars as direct tracers due to their low space density. *eBOSS* overcomes this challenge by surveying

[1]In addition, smaller dedicated programs affiliated with *SDSS* have targeted higher redshift quasars to fainter limits

quasars over a huge volume, corresponding to 7,500 $\deg^2$ of sky. The quasar component of *eBOSS* will attempt to statistically target and measure redshifts for $\sim$ 500,000 quasars at $0.8 < z < 2.2$ (including spectroscopically confirmed quasars from *SDSS-I/II*, which will not need to be retargeted). We refer to this homogeneous tracer sample as the *eBOSS CORE* quasars. In contrast to *BOSS*, *eBOSS* will open up the $i > 19.1$, $z < 2.2$ parameter space to directly use quasars themselves as cosmological tracers.

In total, at the conclusion of *eBOSS*, the *SDSS* surveys will have spectroscopically confirmed more than 800,000 quasars. The scope of the science that can be conducted with a large sample of quasars across a range of redshifts has been shown to be vast. In addition to higher-redshift studies, *SDSS-IV/eBOSS* will produce a $z < 2.2$ sample of quasars about six times larger than the final *SDSS-II* quasar catalog (Schneider et al., 2010) and will further benefit from upgrades conducted for *SDSS-III* (such as larger wavelength coverage for spectra; see Smee et al., 2013b, for extensive details of upgrades). Many high-impact projects that used the original *SDSS-I/II* quasar samples can therefore potentially be revisited using much larger samples with *eBOSS*, such as composite quasar spectra, rare types of quasars, and precision studies of the quasar luminosity function (e.g. Vanden Berk et al., 2001; Inada et al., 2003; McLure & Dunlop, 2004; Hennawi et al., 2006; Richards et al., 2006; York et al., 2006; Netzer & Trakhtenbrot, 2007; Kaspi et al., 2007; Shen et al., 2008; Boroson & Lauer, 2009).

In this chapter, we present a detailed analysis of quasar sample of the *SDSS-IV/eBOSS* survey. The details of quasar selection is beyond the scope of this thesis and is presented in detail in Myers et al. (2015). Further technical details about *eBOSS* can be found in our companion papers which include an overview of *eBOSS* (Dawson et al., 2015) and discussions of targeting for Luminous Red Galaxies (see Chapters 2 and 3), and Emission Line Galaxies (Comparat et al., 2015b). *eBOSS* will run concurrently with two surveys; the SPectroscopic IDentification of ERosita Sources survey (*SPIDERS*) and the Time Domain Spectroscopic Survey (*TDSS*; Morganson et al., 2015). These associated surveys are further outlined in our companion overview paper (Dawson et al., 2015).

In Section 4.2, we discuss how forecasts for BAO constraints at different redshifts drive targeting goals for *eBOSS* quasars. The parent imaging used for *eBOSS* quasar target

selection is outlined in Section 4.3. An important criterion for any large-scale structure survey is sufficient homogeneity to facilitate modeling of the distribution of the tracer population—the "mask" of the survey. In Section 4.4, we use the full *eBOSS* quasar target sample to characterize the homogeneity of *eBOSS* quasar selection. In Section 4.5, we provide our overall conclusions regarding *eBOSS* quasar targeting.

Unless we state otherwise, all magnitudes and fluxes in this chapter are corrected for Galactic extinction using the dust maps of Schlegel et al. (1998b). Specifically, we use the correction based upon the recalibration of the *SDSS* reddening coefficients measured by Schlafly & Finkbeiner (2011b). For *WISE* we adopt the reddening coefficients from Fitzpatrick (1999). The SDSS photometry has been demonstrated to have colors that are within 3% (Schlafly & Finkbeiner, 2011b) of being on the AB system (Oke & Gunn, 1983b). *WISE* is calibrated to be on the Vega system. We use a cosmology of ($\Omega_{\rm m}$, Ω_Λ, $h \equiv H_0/100\,{\rm km\,s^{-1}\,Mpc^{-1}}$) = (0.315, 0.685, 0.67) consistent with recent results from *Planck* (Planck Collaboration et al., 2014).

4.2 COSMOLOGICAL GOALS OF *EBOSS* AND IMPLICATIONS FOR QUASAR SAMPLE

4.2.1 CORE quasars

eBOSS quasar survey aims to study the scale of the BAO in two redshift regimes; $z \sim 1.5$ using the clustering of quasars, and $z \sim 2.5$ using high redshift quasars to illuminate the intervening Lyman-α Forest. The first approach requires a sample of statistically selected quasars in the redshift range $0.9 < z < 2.2$, which we will refer to as "CORE quasars" and quasars selected at $z > 2.1$, which are referred to as "Lyman-α quasars". In this thesis we focus on the CORE quasars only. The selection of CORE quasars must be statistically uniform and reproducible. In fact, the full redshift range of the CORE sample will extend well beyond $0.9 < z < 2.2$, and many CORE quasars can thus be utilized as Lyman-α quasars.

4.2.2 Requirements for CORE quasar sample

As explained in Dawson et al. (2016), a density of 90 deg^{-2} spectroscopic fibers are allocated to *eBOSS* quasars and a density of 58 deg^{-2} quasars with redshifts $0.9 < z < 2.2$ is required over the projected 7,500 deg^{-2} survey footprint, to meet the eBOSS scientific goals (see Dawson et al. (2016) for more details). Additionally, we require the redshifts be accurate to better than 300 km s^{-1} RMS and robust such that the fraction of catastrophic redshift errors (exceeding 3000 km s^{-1}) is $< 1\%$ in cases where the redshifts are believed to be secure. The construction of the quasar sample designed to fulfill these requirements is driven by instrument capabilities and a 2% measurement of the BAO distance scale (G. Zhao et al. 2016, in preparation).

A further requirement to obtain robust BAO measurements is that the density of selected quasars must not strongly correlate with either tracers of potential imaging systematics (e.g., variations in the depth of the imaging) or with astrophysical systematics such as Galactic extinction and stellar density. *BOSS* has shown that fluctuations associated with surveys artifacts can be handled effectively via weighting schemes provided the amplitude of fluctuations is relatively small (Ross et al., 2012a). To facilitate weighting schemes in future clustering studies, we require that fluctuations in the expected target density as a function of potential imaging systematics, stellar density, and Galactic extinction be less than 15% (total variation around mean density). We require density differences due to imaging zero point variations in any single band to be below 15% as well. Tests of the homogeneity of the quasar target sample are presented in Section 4.4.

4.3 PARENT IMAGING FOR TARGET SELECTION

4.3.1 Updated calibrations of *SDSS* imaging

The parent imaging from SDSS and WISE for quasar selection is identical to the one for LRGs presented in 3.3 of Chapter 3. In addition, we also use the Palomar Transient Factory (PTF) photometry. We present it here for the sake of completeness of this chapter and convenience

of the readers. All *eBOSS* quasar targets rely on imaging from the *SDSS-I/II/III*. *SDSS* photometry was obtained by the *SDSS* telescope (Gunn et al., 2006) using its wide-field imaging camera (Gunn et al., 1998) in the *ugriz* system (Fukugita et al., 1996a). *SDSS-I/II* primarily obtained imaging over the $\sim 8400\,\mathrm{deg}^2$ "Legacy" area, $\sim 90\%$ of which was in the North Galactic Cap (NGC). This imaging was released as part of *SDSS* Data Release 7 (DR7; Abazajian et al., 2009). The legacy imaging area of the *SDSS* was expanded by $\sim 2500\,\mathrm{deg}^2$ in the South Galactic Cap (SGC) as part of *DR8* (Aihara et al., 2011). The *SDSS-III/BOSS* survey used this DR8 imaging for target selection over $\sim 7600\,\mathrm{deg}^2$ in the NGC and $\sim 3200\,\mathrm{deg}^2$ in the SGC (Dawson et al., 2013a). Quasar targets for *eBOSS* have been selected over the same footprint covered by *BOSS*; however, ultimately *eBOSS* will obtain spectroscopy for quasars over a roughly $7500\,\mathrm{deg}^2$ subset of this *BOSS* area, utilizing 50% of the dark time for 6 years in *SDSS-IV*. In this available time, it would not be possible to observe the full extragalactic footprint available from SDSS imaging.

Although conducted over the same *area* as *BOSS*, *eBOSS* target selection takes advantage of updated calibrations of the *SDSS* imaging. Schlafly et al. (2012) have applied the "uber-calibration" technique of Padmanabhan et al. (2008) to imaging from the *Pan-STARRS* survey (Kaiser et al., 2010), achieving an improved global calibration compared to *SDSS DR8*. The improvements in the photometric accuracy is very modest compared to DR8, typically less than 0.5%. More importantly, *BOSS* was still acquiring imaging during its program, and therefore there was no single photometric solution for the full footprint. In contrast, *eBOSS* is targeting from a single photometric solution. Targeting for *eBOSS* is conducted using *SDSS* imaging that is calibrated using the Schlafly et al. (2012) *Pan-STARRS* solution. We will refer to this as the "updated" photometry below.

Specifically, targets are selected using the updated *SDSS* photometry stored in the `calib_obj` files, the basic imaging catalog files used in the *SDSS-III* data model.[2] The updated Pan-STARRS-calibrated photometry will be made available as part of a future *SDSS* Data Release. The magnitudes provided in these files are Pogson magnitudes (Jones, 1968) rather than the asinh magnitudes used for some *SDSS* data releases (Lupton et al., 1999). We use `Model Magnitudes` for all colors and fluxes used in selection. The `Model Magnitudes`

[2] e.g., `http://data.sdss3.org/datamodel/files/PHOTO_SWEEP\/RERUN/calibObj.html`

are obtained by first determining what type of model (exponential or deVaucouleurs) best fits the object image in the 'canonical' band (typically r, but other bands may be used if they have higher signal-to-noise), and then using the model fit from the canonical band (convolved with the appropriate PSF) to obtain fluxes in each filter. Additionally, we also apply flux limits based upon an object's `fiber2mag` values; i.e., the total flux within a 2" diameter of the object center, corresponding to the aperture of a *BOSS* spectroscopic fiber (Smee et al., 2013a), after convolving the imaging data to achieve a standard 2" seeing.

4.3.2 WISE

The *eBOSS* quasar targets also relie on infrared photometry from the Wide-Field Infrared Survey Explorer (*WISE*; Wright et al., 2010a). *WISE* observed the full sky in four infrared channels centered at 3.4, 4.6, 12, and 22 microns, which we refer to as W1, W2, W3, and W4, respectively. For *eBOSS* LRGs, we use the W1 band only. *WISE* magnitudes are commonly measured in the Vega system, but we convert to the AB system for LRG selection.[3] Over the course of its primary mission and the 'NEOWISE post-cryo' continuation, *WISE* completed two full scans of the sky in the W1 and W2 bands. Over 99% of the sky has 23 or more exposures in W1 and W2, and the median coverage is 33 exposures. We use the 'unWISE' forced photometry from Lang et al. (2014), which photometered custom coadds of the *WISE* imaging at the positions of all *SDSS* primary sources. Using forced photometry allows accurate flux measurements to be obtained even for significantly blended sources, including objects below the significance threshold for *WISE* only detections. Since the *WISE* W1 point-spread function is relatively broad (6.1 arc-seconds FWHM, ∼4 times larger than typical *SDSS* seeing), many sources are blended and forced photometry presents substantial advantages. Additionally, forced photometry allows us to leverage the relatively deep *SDSS* photometry to measure fluxes of *WISE* sources that are otherwise below the detection threshold. The same canonical morphological model is used in fitting photometry of the optical SDSS and infrared WISE images, therefore consistently measuring colors across all bands. Using unWISE photometry instead of the Wright et al. (2010a) *WISE*

[3] $W1_{\rm AB} = W1_{\rm Vega} + 2.699$

catalog increases the size of the resulting *eBOSS* quasar sample since many of our targets have *WISE* fluxes below the "official" *WISE* catalog detection limits.

4.3.3 The Palomar Transient Factory (PTF)

The Palomar Transient Factory[4] (*PTF*) is a wide-field photometric survey aimed at a systematic exploration of the optical transient sky via repeated imaging over 20,000 $\deg^2$ in the Northern Hemisphere (Rau et al., 2009; Law et al., 2009). The PTF image processing is presented in Laher et al. (2014), while the photometric calibration, system and filters are discussed in Ofek et al. (2012). In February 2013, the next phase of the program, *iPTF* (intermediate *PTF*), began. Both surveys use the CFHT12K mosaic camera, mounted on the 1.2 m Samuel Oschin Telescope at Palomar Observatory. The camera has an 8.1 $\deg^2$ field of view and $1''$ sampling. Because one detector (CCD03) is non-functional, the usable field of view is reduced to 7.26 $\deg^2$. Observations are mostly performed in the Mould-R broad-band filter, with some in the SDSS g-filter. Under median seeing conditions, the images are obtained with $2.0''$ FWHM, and reach 5σ limiting AB magnitudes of $m_R \simeq 20.6$ and $m_{g'} \simeq 21.3$ in 60-second exposures. The cadence varies between fields, and can produce one measurement every five nights in regions of the sky dedicated to supernova searches. Four years of *PTF* survey operations have yielded a coverage of $\sim 90\%$ of the *eBOSS* footprint. Two automated data processing pipelines are used in parallel in the search for transients; a near-real-time image subtraction pipeline at Lawrence Berkeley National Laboratory (LBNL), and a database populated on timescales of a few days at the Infrared Processing and Analysis Center (IPAC). The *eBOSS* analysis uses the individual calibrated frames available from IPAC (Laher et al., 2014).

We have developed a customized pipeline based on the SWarp (Bertin et al., 2002) and SCAMP (Bertin, 2006) public packages to build coadded *PTF* images on a timescale adapted to quasar targeting—i.e., typically 1 to 4 epochs per year depending on the cadence and total exposure time within each field. Using the same algorithms, a full stack is also constructed by coadding all available images. This full stack is complete at 3σ to $g \sim 22.0$,

[4]See http://irsa.ipac.caltech.edu/Missions/ptf.html for the public *PTF* data

and has over 50% completeness to quasars at $g \sim 22.5$. The full stack is used to extract a catalog of *PTF* sources from each of the coadded *PTF* images. The light-curves (flux as a function of time) for all of these *PTF* sources are measured.

4.4 TESTS OF THE HOMOGENEITY OF THE CORE QUASAR SAMPLE

In order to perform clustering measurements to characterize the BAO scale, it is necessary to mimic the angular distribution imposed by the target selection. This survey "mask" is often expressed as a random catalog, or control sample, that mimics the characteristics of the targeted population but in the absence of any clustering. At its simplest, this process involves uniformly distributing random points over the footprint of the target imaging. This simple approach, however, is rarely adequate because survey systematics such as seeing, sky brightness, Galactic extinction etc. alter the target density in a complex manner. A related issue is that zero-point calibrations in *SDSS* imaging can vary across the survey, also producing non-cosmological variations in target density.

4.4.1 Overall characteristics of *eBOSS* quasars

Beyond the cosmological goals of *eBOSS*, the quasar sample produced by *SDSS-IV* should be unparalleled, exceeding the depth and numbers of any previous quasar sample. As there is likely to be significant interest in the nature of *eBOSS* for quasar science, quasars observed as part of *SEQUELS* are broadly characterized in this section. Because *SEQUELS* observations were conducted in tandem with *BOSS*, some quasars that would not normally receive a fiber in *eBOSS* because of existing *BOSS* spectroscopy did receive a *SEQUELS* fiber. We treat such objects as if they had the `DO_NOT_OBSERVE` bit set by correctly incorporating (non-*SEQUELS*) redshifts and classifications.

The overall expected quasar numbers for *eBOSS* can be estimated from the number densities. Assuming a minimum *eBOSS* area of $7500\,\mathrm{deg}^2$, *eBOSS* should, conservatively, comprise at least 500,000 spectroscopically confirmed $0.9 < z < 2.2$ quasars selected in a

uniform manner with which to pursue quasar clustering studies such as the BAO scale, and at least 500,000 total *new* quasars (at any redshift) that have never before been spectroscopically identified and characterized. Overall, at the completion of *eBOSS*, the *SDSS* surveys will have provided unique spectra of over 800,000 total quasars, including *SDSS* areas outside of the *eBOSS* footprint as well as new quasars observed by the *TDSS* and *SPIDERS* surveys.

4.4.2 Target density fluctuations due to systematics

Previous studies of large-scale galaxy clustering over the *SDSS* footprint (e.g., Ross et al., 2011b) have demonstrated that systematics that produce target density variations at a level of ∼15% or less can be controlled for by weighting the random catalog by a model of the effect of that systematic. Beyond the 15% level, systematics become more difficult to "weight" for, perhaps because some major systematics are covariant. When the effect of systematics exceeds the 15% level, that area of the survey may have to be excised from clustering analyses.

As part of *eBOSS* target selection, a set of regression tests have been devised to study how possible systematics in *SDSS* and *WISE* imaging may affect target density—and whether such effects are below the ∼ 15% level that could be modeled with a suitable weighting scheme. The slate of systematics, which represents a reasonable (but not necessarily exhaustive) list of quantities that could bias *eBOSS* target density, is further detailed in Chapter 3 (see also the companion paper Prakash et al. (2016)). Relevant to the *WISE* imaging; the systematics include the median numbers of exposures per pixel, the fraction of exposures contaminated by the Moon, and the total flux per pixel, all in the W1 band (`W1covmedian, moon_lev, W1median`). Relevant to the *SDSS* imaging; the systematics include the FWHM and background sky-level in *SDSS* z-band, which are used to track the quality of the seeing and the sky brightness. Additional systematics include Galactic latitude (to map the density of possible contaminating stars) and Galactic dust (extinction in the r-band is used to represent this systematic).

The adopted regression technique is also detailed in Chapter 3 (see also the companion paper Prakash et al. (2016)). Briefly, the potential *eBOSS* imaging footprint is deconstructed into equal-area pixels of 0.36 deg^2. The *eBOSS* CORE quasar target density and the mean

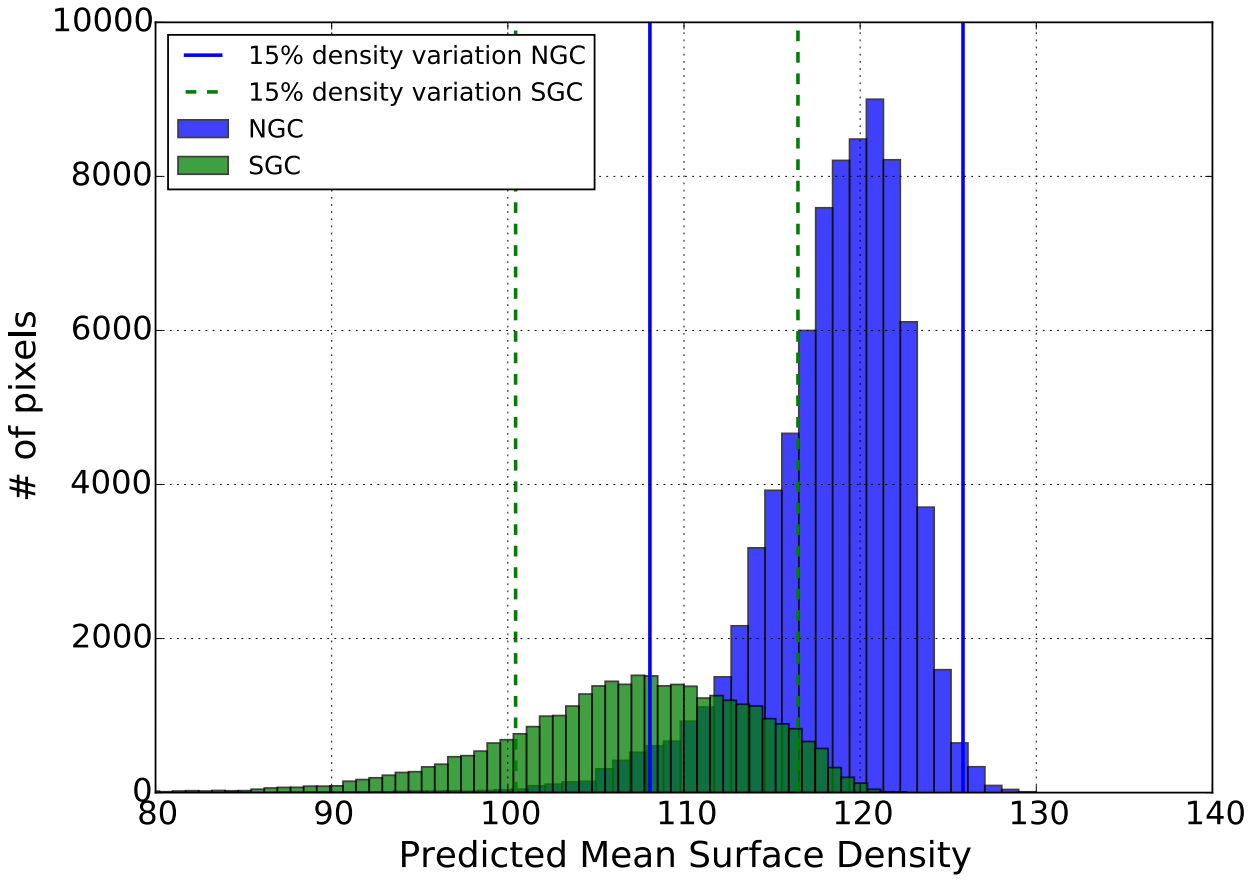


Figure 4.1: Histograms of the surface density of CORE quasar targets predicted by the regression models described in Section 4.4.2 (the "PSD"). The blue histogram represents the NGC, with solid blue lines depicting the window within which angular fluctuations in quasar target density meet the $\leq 15\%$ requirement of Section 4.2.2. The green histogram and dotted green lines depict the same quantities for the SGC. The histograms demonstrate that $\sim 97\%$ ($\sim 77\%$) of the NGC (SGC) footprint meets the homogeneity requirements of *eBOSS* (see Section 4.2.2). The PSD and the fractional deviation from the mean PSD in each pixel are depicted as a sky map in Figure 4.2.

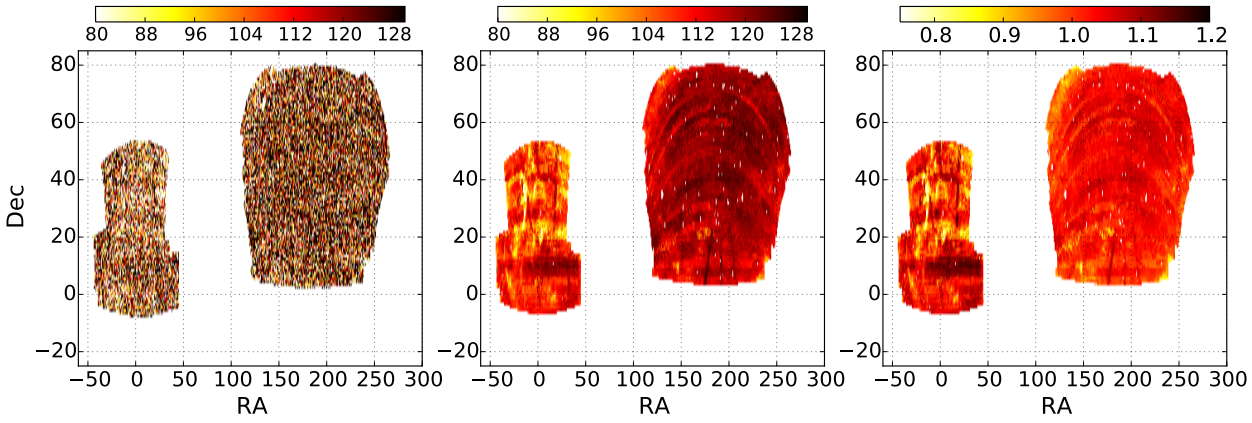


Figure 4.2: Actual and theoretical maps of *eBOSS* CORE quasar targets in J2000 Equatorial Coordinates (degrees). The left-hand panel shows the observed surface density sky-map of targets over the *BOSS* footprint. *eBOSS* will target quasars over a ∼7500 deg^2 subset of this area. As CORE quasar targets are relatively scarce (~ 115 deg^{-2}) fluctuations in this map are dominated by Poisson noise and sample variance. The central panel shows the theoretical map of CORE quasar target density predicted by the linear regression from imaging systematics (the PSD described in Section 4.4.2). The color bars above the left-hand and central panels represent target densities in deg^{-2}. The right-hand panel rescales the map in the central panel so that it is expressed as a fractional deviation from the mean (i.e. the color-bar above this panel represents the quantity PSD/⟨PSD⟩).

value of each systematic is determined for each of these pixels. The observed surface density ($\mathrm{SD_{obs}}$) of *eBOSS* CORE quasar targets in each pixel can be expressed as a linear model of systematics

$$\mathrm{SD_{obs}} = S_0 + \sum_{i=1}^{7} S_i x_i + \epsilon, \tag{4.1}$$

where S_0 is the mean target density across the pixels, S_i is the weight accorded to fluctuations in target density (x_i) due to systematic i, and ϵ is the combined effect of noise and variance, which is approximated as a Gaussian. Multi-linear regression is used to determine S_0 and S_i by minimizing the value of reduced χ^2 across the pixels. This regression is conducted separately in each Galactic hemisphere, such that different coefficients are derived for the NGC and SGC regions of the *SDSS* imaging.

Once the coefficients of the linear regression model for systematics have been established, a statistic designated the *Predicted Surface Density* or "PSD" is computed. The PSD

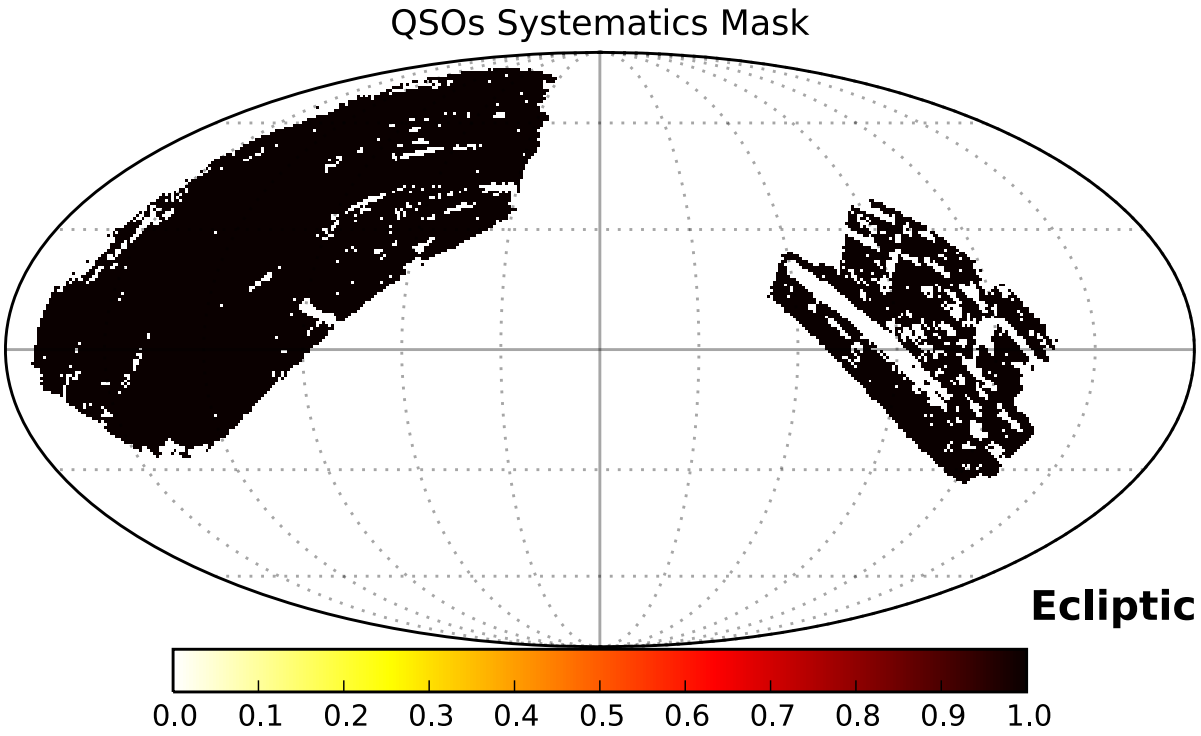


Figure 4.3: *eBOSS* Quasar survey mask based on theoretical map. The map is made using healpix pixelization scheme (NSIDE=512). The black color represents the regions showing less than 15% variation in the density around mean. The regions showing more than 15% variation in target density is excluded from LSS studies.

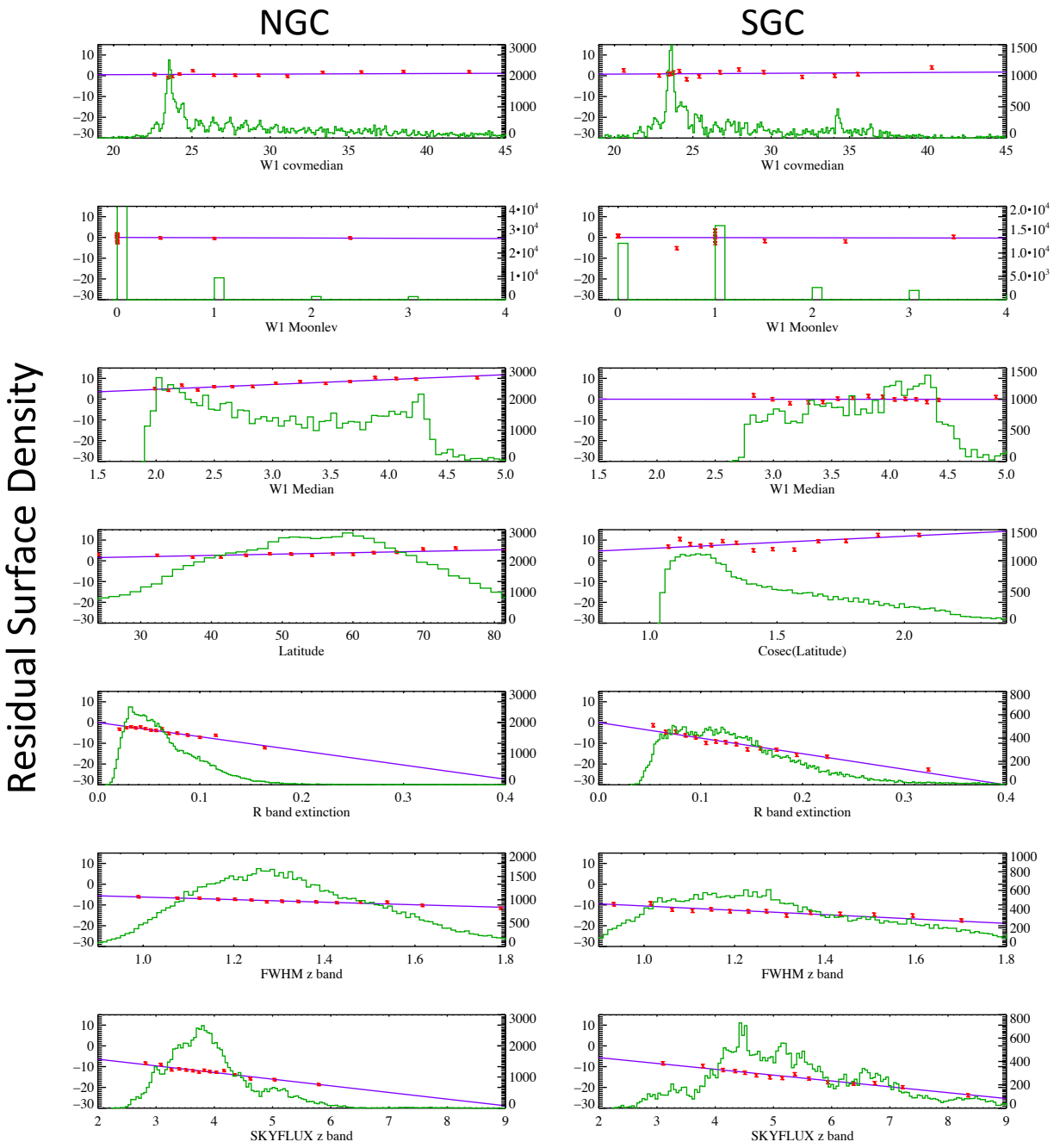


Figure 4.4: Systematics distributions and linear regression surface density models for *eBOSS* CORE quasar targets. Each row of panels corresponds to one of the systematics outlined in Section 4.4.2 ("Latitude" refers to Galactic latitude). The left-hand (right-hand) column of panels displays results for these systematics for the NGC (SGC). The green histograms depict the distribution of pixels as a function of the mean value of each systematic in each pixel. The number of pixels is quantified on the right-hand axis of each plot. The red data points and blue lines depict, instead, measures of the Residual_SD (Eqn. 4.3), which is quantified on the left-hand axis of each plot. The points are the measured values of the Residual_SD averaged over 4000 sky pixels in the NGC or 2000 pixels in the SGC. The error bars depict the standard error on the mean across the pixels. The lines show the best-fit regression models. A linear regression model appears to be an adequate description of how each displayed systematic affects *eBOSS* CORE quasar target density.

is obtained by using S_0 and S_i to calculate what the *eBOSS* CORE quasar density *should be* in a given pixel if the linear regression model is an adequate description

$$\mathrm{PSD} = S_0 + \sum_{i=1}^{7} S_i x_i \quad . \tag{4.2}$$

Figure 4.1 presents a histogram of the CORE quasar PSD as predicted from the derived linear regression model coefficients across all of the systematics. We further present the mask of *eBOSS* quasar survey in Figure 4.3. A total of 96.7% of the *SDSS* imaging footprint in the NGC[5] fluctuates in CORE quasar PSD at less than 15%. The corresponding fraction is 76.7% in the SGC footprint.

Figure 4.2 illustrates these deviations *on the sky* using a map of the PSD statistic, which serves to illustrate the most problematic areas of the *SDSS* footprint for *eBOSS*. The right-hand panel of Figure 4.2 approximates the "mask" that will be necessary to ameliorate the effects of systematics on clustering measurements that use *eBOSS* CORE quasars. The effective area or random catalog in each region of the *eBOSS* footprint can be re-weighted by the values displayed in the right-hand panel of Figure 4.2, although regions that deviate by more than 15% from expectation may need to be excised from the survey in order to reach the target density variation requirement of Section 4.2.2. The central panel of Figure 4.2 is a particularly clear illustration of why the PSD is regressed separately in the NGC and SGC regions—the NGC appears to be more robust to systematics than the SGC. To determine whether a linear regression adequately models the effect of systematics on the target density of *eBOSS* CORE quasars, the statistics designated the Reduced_PSD$_j$ and the Residual_PSD$_j$ as described in Chapter 3 can be calculated. The Reduced_PSD$_j$ is derived from the PSD by omitting the j'th systematic term when calculating the PSD—in order to represent the deviation from the PSD caused by each systematic. The *difference* between the PSD and the observed sky density of targets, called the *Residual Surface Density*, or "Residual_SD," is then calculated. If a linear model is an appropriate representation of the regression of a given systematic, then the Residual_PSD should be well-represented by a model with a slope of S_j. Formally:

[5] only the area that could be useful for *eBOSS* targeting, due to scheduling constraints, is considered (see Dawson et al., 2015, and Figure 4.2)

$$\text{Reduced\_PSD}_j = \text{PSD} - S_j \times x_j \quad ,$$
$$\text{Residual\_SD}_j = SD_{\text{obs}} - \text{Reduced\_PSD}_j \quad . \tag{4.3}$$

Figure 4.4 shows how the CORE quasar Residual_SD varies as a function of each of the individual systematics, together with the underlying distributions of those systematics. In general, a linear regression seems to be adequate for modeling variations in CORE quasar target density. Figure 4.4 suggests that sky brightness, and, in particular, Galactic extinction, are the main culprits in causing variations in *eBOSS* CORE quasar target density. The SGC has a 68% range of r-band extinction of 0.075 to 0.19 with a median of 0.12, whereas the NGC has a 68% range of r-band extinction of 0.032 to 0.10, with a median of only 0.057. The corresponding numbers for z-band sky flux are 4.1 to 6.8 with a median of 5.1 in the SGC and 3.3 to 4.6 with a median of 3.8 in the NGC. The higher median and wider range of values of these systematics in the SGC are likely responsible for both the suppressed density of SGC targets and the larger RMS in predicted surface density that can be seen in Figure 4.2. These systematics will act to reduce the effective depth of an exposure and hence to increase the error on the fluxes of a test object being assigned a quasar probability by the *XDQSOz* method. In effect, as the flux errors for a test object increase, the formal probability that the object is a quasar is reduced, and fewer objects are then assigned PQSO($z > 0.9$) > 0.2 by *XDQSOz*.

Overall, the *eBOSS* quasar sample outlined in this chapter for clustering measurements is expected to be robust against systematics across essentially the entire NGC and across about three-quarters of the SGC. This statement may be pessimistic, as *eBOSS* does not attempt to restrict the CORE quasar redshift range to $0.9 < z < 2.2$ in advance of spectroscopic confirmation. Quasars at $z > 2.2$ are closer to the stellar locus in optical color space, so the target density of quasars at $z > 2.2$ may fluctuate more due to systematics than at $z < 2.2$. Weighting for systematics as a function of quasar redshift is a possible avenue for further improving *eBOSS* clustering measurements once target redshifts have been confirmed by spectroscopy. The final *eBOSS* footprint is yet to be derived but in the worst-case scenario that the entire SGC has to be observed, only $\sim$ 86.7% of *eBOSS* will meet the

Table 4.1: Results of how zero-point fluctuations affect target density

	$N^{-1}(\Delta N/\Delta \mathrm{m})$	zero-point error	fluctuation
	(1)	(2)	(3)
u	0.544	13×10^{-3}	2.8%
g	0.856	9×10^{-3}	3.1%
r	0.514	7×10^{-3}	1.4%
i	0.475	7×10^{-3}	1.3%
z	0.061	8×10^{-3}	0.2%
W	0.223	20×10^{-3}	1.8%

Note. — (1) Fractional deviation in target density that results from a ± 0.01 mag scatter in each band; (2) Zero-point RMS error in each band in magnitudes. Values for the *SDSS* are taken from D. Finkbeiner et al. (2016, in preparation). Values for the *WISE* stack are estimated from Jarrett et al. (2011b); (3) 95% ($\pm 2\sigma$) values in target density fluctuation corresponding to $100\% \times 4 \times$ [zero-point error] $\times$ $[N^{-1}(\Delta N/\Delta \mathrm{m})]$

requirements of Section 4.2.2. This fraction of useful area is almost exactly offset by the expected excess of *eBOSS* CORE quasars. *eBOSS* will confirm $(0.95 \times 72.0 =)$ 68.4 deg^{-2} $0.9 < z < 2.2$ quasars. Serendipitously, 68.4 deg$^{-2} \times 0.867 =$ 59.3 deg^{-2}, exceeding the requirement of 58 deg^{-2} $0.9 < z < 2.2$ quasars noted in Section 4.2.2.

4.4.3 Target density fluctuations due to zero-point variations

A further requirement of *eBOSS* is that fluctuations in target density due to shifting zero-point calibrations across the *SDSS* imaging footprint are well-controlled. Similar to Section 4.4.2, such fluctuations need to be kept below the 15% level (see also Section 4.2.2)[6]. To study how changes in zero-point affect the density of *eBOSS* CORE quasar targets, each of the bands used in the *eBOSS* CORE quasar selection is offset by ± 0.01 mags (i.e. scaled by 1% in flux) and the resulting fractional changes in target density are determined after

[6]This 15% limit is on the two-tailed distribution (i.e. between the peaks due to a positive and a negative fluctuation in zero-point)

re-running the target selection pipeline. Each *SDSS* band is tested individually. As the *WISE* bands are only incorporated into *eBOSS* CORE quasar target selection, both W1 and W2 are simultaneously shifted by ± 0.01 mags and the result is reported as a single band (henceforth denoted W).

The resulting fractional fluctuations in target density from these offsets ($N^{-1}[\Delta N/\Delta \mathrm{m}]$) can then be multiplied by the zero-point RMS error expected for the imaging calibrations used by *eBOSS* (see Section 4.3) to determine the expected RMS variation in number density due to zero-point calibrations shifting across the *eBOSS* footprint. We adopt the zero-point errors in $[u, g, r, i, z]$ of [13, 9, 7, 7, 8] mmag RMS from D. Finkbeiner et al. (2016, in preparation) and conservatively estimate a zero-point error of 20 mmag RMS for the W stack (see Jarrett et al., 2011b). Assuming that the zero-point errors can be modeled using a Gaussian distribution, 95% of CORE quasar targets in *eBOSS* will be within $\pm 2\sigma$ of the expected RMS variation. In other words, 95% fractional variance in target density can be interpreted as meaning that 95% of the area of the sky is expected to be described by fluctuations of $\pm 2\sigma$. Thus, the overall 95% fractional variance in target density due to zero-point errors can be expressed (as a percentage) as $100\% \times 4 \times$ [zero-point error] $\times$ $[N^{-1}(\Delta N/\Delta \mathrm{m})]$. Table 4.1 displays the results of this analysis, which indicate that g-band is the least robust to zero-point variations when selecting *eBOSS* CORE quasars. Even g-band, however, causes a (2σ) variation of only 3%, far less than the 15% limit outlined in Section 4.2.2. *eBOSS* CORE quasar target selection is thus completely robust to zero-point errors.

4.5 CONCLUSIONS AND SUMMARY

The SDSS-IV will include the *eBOSS*, a project with an ambitious goal of using galaxies and quasars to measure the BAO scale across a range of redshifts. This chapter details the properties of a sample of quasars, referred to as the *eBOSS* "CORE" sample, that can provide the first 2% constraints on the BAO scale at redshifts $0.9 < z < 2.2$ through clustering measurements. Not all such sources, however, are targeted for spectroscopy in

eBOSS. The *eBOSS* survey does not place a fiber on any target that has an existing good spectrum from earlier iterations of the *SDSS*.

Ultimately, *eBOSS* will uniformly target in excess of 500,000 quasars in the redshift range $0.9 < z < 2.2$, exceeding previous such clustering samples by a factor of more than ten. Samples of new spectroscopically confirmed quasars across all redshifts in *eBOSS* will exceed 500,000 quasars, which will be at least three times larger than all previous samples across the *eBOSS* footprint combined. At the conclusion of *eBOSS*, in excess of 800,000 confirmed quasars should have spectra from some iteration of the *SDSS*. In essence, *eBOSS* is *the* next-generation quasar survey, and, in the wake of 20 years of observations from *SDSS-I, II, III* and *IV*, *eBOSS* will usher in the era of million-fold spectroscopic quasar samples.

5.0 SDSS-IV/EBOSS: THE CLUSTERING OF LUMINOUS RED GALAXIES USING PHOTOMETRIC REDSHIFTS

5.1 INTRODUCTION

The spatial distribution of galaxies is a powerful cosmological probe for studying structure formation and evolution in the universe. The observed galaxy density is expected to have a simple relationship to the overall density of matter (primarily dark matter) density. The BAO feature which is embedded in the spatial distribution of galaxies can be used as a standard ruler for cosmology (Eisenstein & Hu, 1998). The distinguishing feature of the BAO appears in galaxy clustering at large scales as a peak in the two-point autocorrelation function at a comoving length of roughly 100 h^{-1}Mpc. The BAO scale may also be observed via the locations of the first peak in the power spectrum of the CMB, corresponding to the redshift of hydrogen recombination at $z \sim 1100$ (Peebles & Yu, 1970; Sunyaev & Zeldovich, 1980). The measurement of the expansion of the universe using the apparent size of a standard ruler at different redshifts probes the energy content of the universe and can be used to study the properties of the dark energy (Hu, 2005; Eisenstein et al., 2005).

The comoving scale of the BAO feature is determined by the matter and radiation densities at hydrogen recombination. This is due to that fact that the speed of a sound wave in a plasma is determined by the ratio of baryons and photons. As mentioned in Chapter 2, this feature was first detected in the SDSS Luminous Red Galaxy sample (Eisenstein et al., 2005). This has inspired astronomers to push LRG sample selection to higher redshifts, as well as exploring other suitable tracers to make 3D maps of the universe. Measuring the apparent size of the BAO features at different redshifts opens up the possibility of directly estimating the angular diameter distance as a function of redshift (Eisenstein & Hu, 1998;

Seo & Eisenstein, 2003; Lin & Mohr, 2003; Ross et al., 2008).

Traditionally, spectroscopic redshifts are used to measure galaxy clustering. However, acquiring spectroscopic redshifts is a time-consuming and expensive process even with modern multi-fiber spectrographs. Additionally, the ultimate accuracy of distance estimates from spectroscopy is limited by peculiar velocities, which can reach 1000 km/s, and lack of knowledge of cosmological parameters. Photometric redshifts, first developed in 1960s (Baum, 1962), are an estimate of the redshifts of galaxies using their photometry alone. This implies using the flux (or brightness) of the objects as seen through different wavelength channels (or filters). This technique relies on the astronomical objects having a strong feature in their spectrum that can be detected by the relatively crude broad-band filters. The technique was replaced by spectroscopic redshifts which use spectroscopy to measure the shift in spectral features from their laboratory positions. However, large sky surveys in the 1990s highlighted the necessity of photometric redshifts as the limited telescope time restricted the spectroscopic follow-up of objects to a very small fraction.

Large multi-band imaging surveys enable us to estimate more accurate photometric redshifts than it was possible before. These photometric redshifts can be improved further using advanced machine learning algorithms applied to large spectroscopic redshift training samples. Although photometric redshifts are less accurate than spectroscopic redshifts, they are significantly easier to obtain, and for a constant amount of time, one can image both wider areas and deeper magnitude limits than would be possible with spectroscopy, allowing one to probe both larger scales and larger volumes.

The aim of this chapter is to demonstrate the practicality of using photometric redshifts for large-scale clustering measurements with real data. We start with the photometrically selected sample of *eBOSS* LRGs which is described in detail in Chapter 3. SDSS and WISE imaging augmented with the spectroscopic training sample of ∼100,000 LRGs spectroscopically confirmed in the first two years of *eBOSS* observations make an accurate estimation of photometric redshifts possible. We then measure the angular clustering in redshift bins of $\Delta z = 0.05$. We pay particular attention to the impact of systematics and revisit the multivariate analysis method first developed in Section 3.7 using a *Healpix* pixelization scheme.

The chapter is organized as follows. In Section 5.2, we briefly outline the *eBOSS*

LRG sample used for this work. In Section 5.3, we explain the Random forest machine learning algorithm used for estimating photometric redshifts. A study of the uniformity and homogeneity of the LRG sample as a function of redshift is outlined in Section 5.4. Section 5.5 then discusses the two-point correlation statistics which are employed to measure the correlation function. We present the results and discussion in Section 5.6 and conclude with future possibilities.

Unless stated otherwise, all magnitudes and fluxes in this chapter are corrected for extinction using the dust maps of Schlegel et al. (1998a), hereafter SFD, and are expressed in the AB system (Oke & Gunn, 1983a). The SDSS photometry has been demonstrated to have colors that are within 3% of being on an AB system (Schlafly & Finkbeiner, 2011a). We use a standard ΛCDM cosmology with H_0=100h km s^{-1} Mpc^{-1}, $h = 0.7$, $\Omega_{\rm M} = 0.3$, and $\Omega_\Lambda = 0.7$, which is broadly consistent with the recent results from *Planck* (Planck Collaboration et al., 2014).

5.2 DATA

5.2.1 Photometric data

For this analysis, we use the *SDSS-IV/eBOSS* photometric sample of ∼600,000 LRGs selected over the entire *SDSS-III/BOSS* footprint of 10,000 deg^2. *SDSS-IV/eBOSS* is an ongoing effort to obtain spectra of ∼370,000 LRGs over 7,500 deg^2 of the SDSS footprint in the redshift range $0.6 < z < 1.0$, with a median redshift of $z \sim 0.71$. The selection of this sample has been presented in detail in Chapter 3.

5.2.2 Spectroscopic data

In the first two years of observations, *eBOSS* has obtained spectra of ∼100,000 LRGs. The redshifts from these spectroscopically confirmed LRGs are used as training sets for building machine learning algorithms, which are then used for estimating photometric redshifts of the entire LRG sample. A majority of LRGs are within the redshift range $0.5 < z < 1.0$, with

very few LRGs outside this range. Hence, only the LRGs within the redshift range $0.4 < z < 1.1$ are selected for training purposes. The spectroscopic target dataset is estimated to have $\sim$10% stellar contamination. By construction, stars are excluded from the training datasets.

5.3 PHOTOMETRIC REDSHIFTS

The redshifts used in this work for measuring clustering are estimated using the Random forest machine learning algorithm. Random forest is an ensemble method that operates by building a number of decision trees and outputting the class that is the mode of the classes (in the case of classification) or mean prediction (for regression) of the individual trees (Ho, 1995). For this work, the relevant output is the regression. The Random forest regressor is a supervised learning algorithm which builds decision trees based on a set of given criteria provided by the training sample. It applies the general technique of bootstrap aggregating to decision tree learners: that is, the algorithm selects a set of random samples with replacement from the training set and fits trees to these samples (Ho, 1998). This bootstrapping procedure leads to better model performance because it decreases the variance of the model (Amit & Geman, 1997). The predictions of a single tree are highly sensitive to noise in its training set. However, the average of many uncorrelated trees is not (Friedman et al., 2000). The bootstrap sampling is a way of de-correlating the trees by showing them different training sets (Breiman, 1996). For these reasons, Random forest has been proved to have better performance on noisy data.

For this work, we use the implementation of Random forest algorithm provided in the software package Scikit-learn (Pedregosa et al., 2011). The main parameters to adjust when using this implementation of the Random forest algorithm are the maximum number of features, the number of estimators, and the depth of individual trees (Buitinck et al., 2013). The number of features is a parameter dependent on the task at hand. The number of features parameter determines the maximum number of features of the training sample to consider when looking for the best split. If this keyword parameter is set to "auto", the routine considers all the features present in the training set. For regression, the maximum

number of features in the training sample (max features = n features) is a good empirical value for this parameter. The number of estimators is the number of trees in the forest. The larger the number of estimators, the better. However, it also takes longer to compute, and the results will stop getting significantly better beyond a critical number of trees. The depth of trees is the parameter which enables individual tree learners to learn highly irregular patterns. One should be careful about this parameter as trees which run too deep tend to overfit their training sets, that is, the statistical model describes random error or noise instead of the underlying relationship in the training sets.

Here, we determine the optimum value of these parameters using the training-test split method. In this technique, a random subset of the training sample is kept aside for testing the results of training. The predictions of the forest tree learners are tested against the known (or expected) output of the test sample. For best results, it is desirable that the training sets are true representatives of the datasets on which these algorithms are ultimately applied.

In this work, we split the spectroscopic sample of ∼100,000 LRGs into two randomly-selected datasets of sizes 80% and 20%. The larger dataset, comprising of 80% of the randomly selected LRGs, is used as the training set. The smaller dataset, comprising of the remaining 20% of the randomly selected LRGs, is kept for testing purposes. At this point, we start building Random forest models on the training sample. To begin with, we use a large number for the maximum depth of trees, ∼40, and attempt to optimize the number of estimators. This optimization is achieved by building Random forest models with increasing numbers of estimators and evaluating them using the same test sample. Here, we start with five estimators and increase in steps of five till 70. To test the predictions of models, we use the standard deviation (SD) and median absolute deviation (MAD) statistics. These results are shown in Figure 5.1.

Based on these results, values larger than 60-65 yield minimal gains. Choosing a smaller number carries the risks of higher variance in the predictions. At this point, we fix the number of estimators (60) and repeat the same process to tune the maximum depth of individual trees. We vary the depth of trees in an increment of 1 and repeat the measurement of our test statistics. The results are shown in Figure 5.2. Although the trend continues to be downward, which implies that we can achieve better performance with even deeper

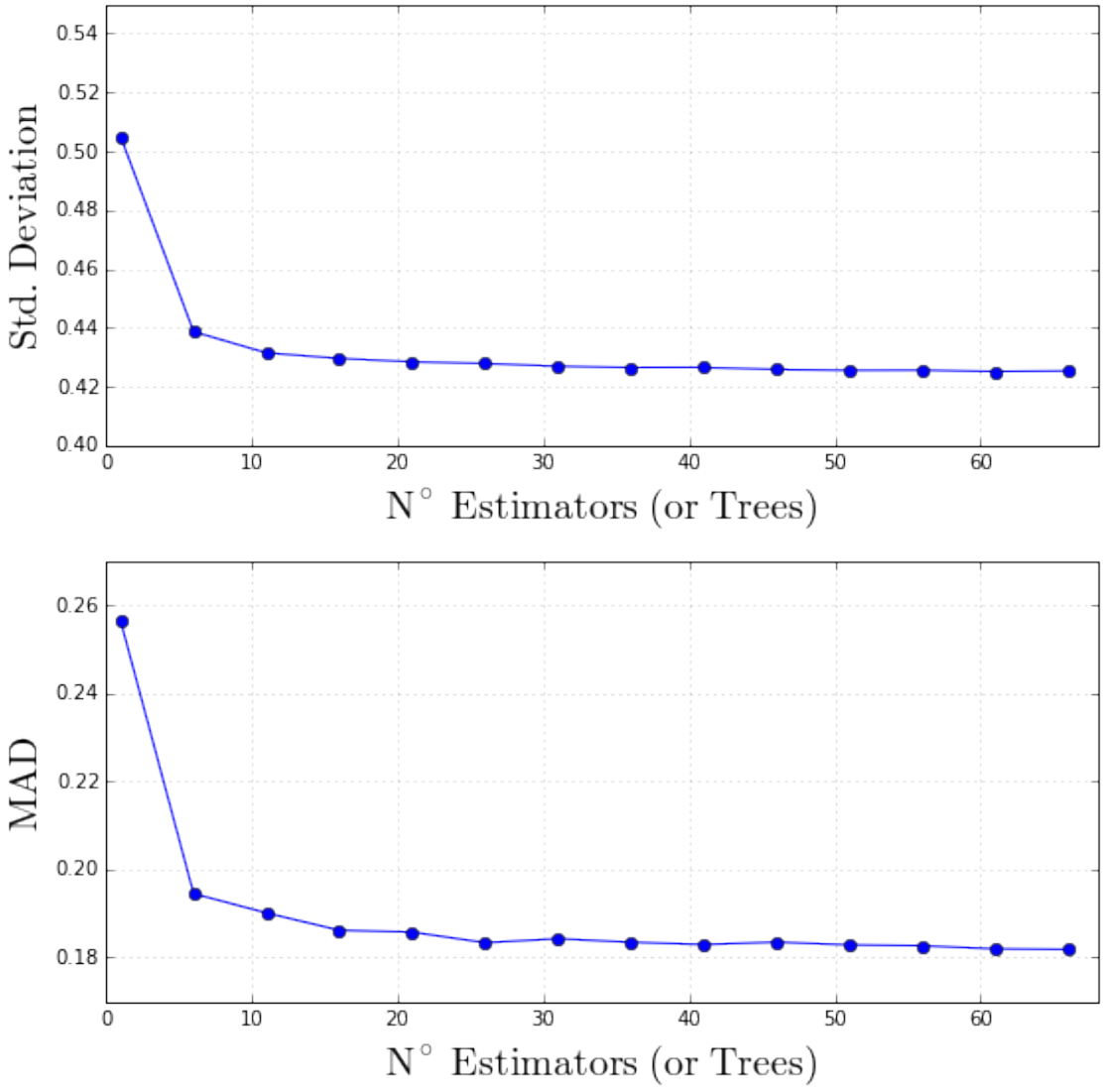


Figure 5.1: Random forest regression results as a function of the number of decision trees. The standard deviation and median absolute deviations are estimated from the predicted photo-z and spectroscopic redshifts of the test sample.

trees, we adopt 20 as the depth parameter. The motive behind this is to keep the runtime reasonable as it scales as the cubic power of the depth of trees. We also assess the fraction of outliers as a function of depth of trees in Figure 5.3. Outliers have been identified as those objects for which the error in predicted photometric redshift is greater than 0.15 (i.e., $\frac{z_{spec}-z_{phot}}{1+z_{spec}} > 0.15$).

At this point, we have obtained near-optimal values of the main parameters of the Random forest algorithm. The final algorithm assumes these values for the key parameters of the Random forest; 1) The number of estimators = 60, 2) the depth of trees = 20, and 3) the number of features = "auto." The results of photometric redshift estimation using these parameter values is plotted in Figure 5.4. We have achieved a standard deviation of $\frac{\sigma_z}{1+z} \sim 0.028$, which is an acceptable error for cosmological studies which employ photometric redshifts. For example, Padmanabhan et al. (2007) detect BAO feature in the SDSS LRG sample using photometric redshifts at $z < 0.6$ with $\frac{\sigma_z}{1+z} \sim 0.03$. The final model is ready to be applied on the full LRG sample, and we present the resulting photometric redshift distribution in Figure 5.5.

5.4 UNIFORMITY AND HOMOGENEITY OF EBOSS LRGS USING PHOTOMETRIC REDSHIFTS

In this section, we revisit the uniformity and homogeneity of *eBOSS* LRGs in detail using a *Healpix* pixelization scheme. As discussed in Section 3.2.1, we require that the target sample should be highly uniform to prevent non-cosmological signals from contaminating clustering measurements. The general methods we use for regression analysis have already been discussed in detail in Section 3.7.3. Here, we briefly review the key steps of regression analysis and incorporate the results into clustering measurements. The important difference here is that we perform this regression analysis in redshift bins of $\Delta z = 0.05$, and the resulting *Predicted Surface Density* or PSD is used for creating random catalogs against which we measure the angular clustering.

As in Sections 3.7.1 and 4.4, we assess the uniformity of the target sample by comparing

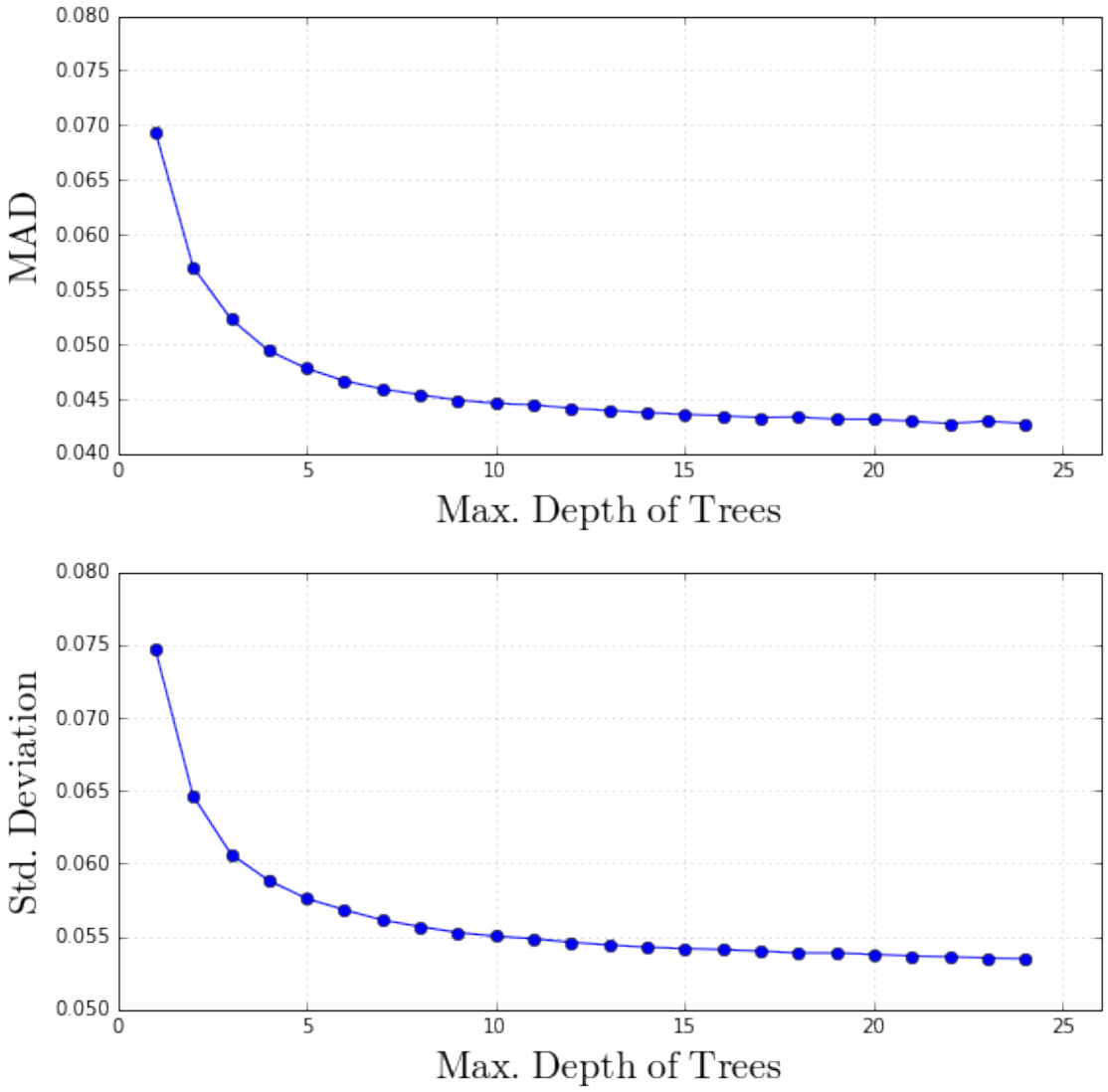


Figure 5.2: Random forest regression results as a function of the depth of individual decision trees. The standard deviation and median absolute deviations are estimated from the predicted photo-z and spectroscopic redshifts of the test sample.

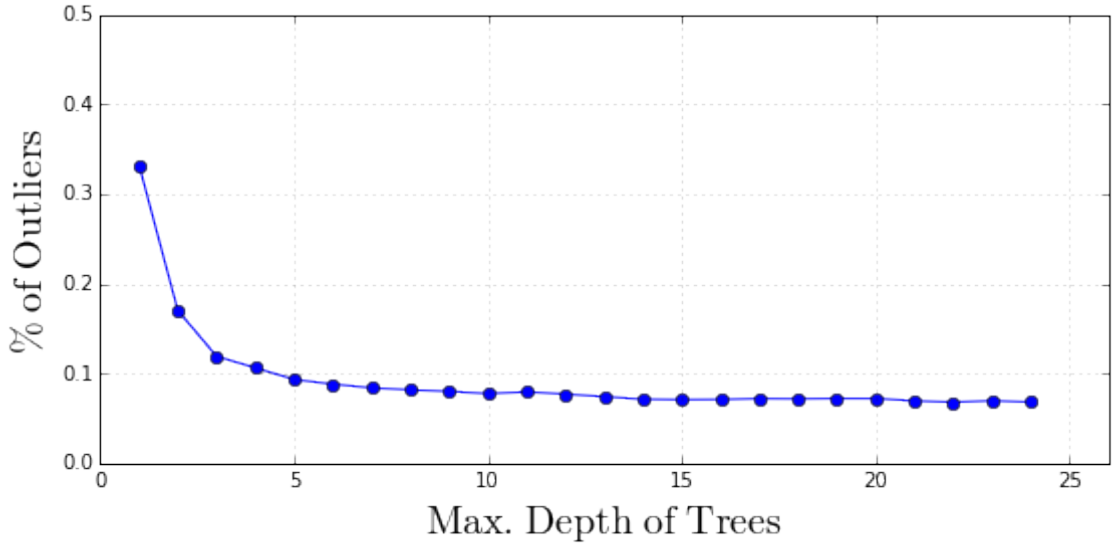
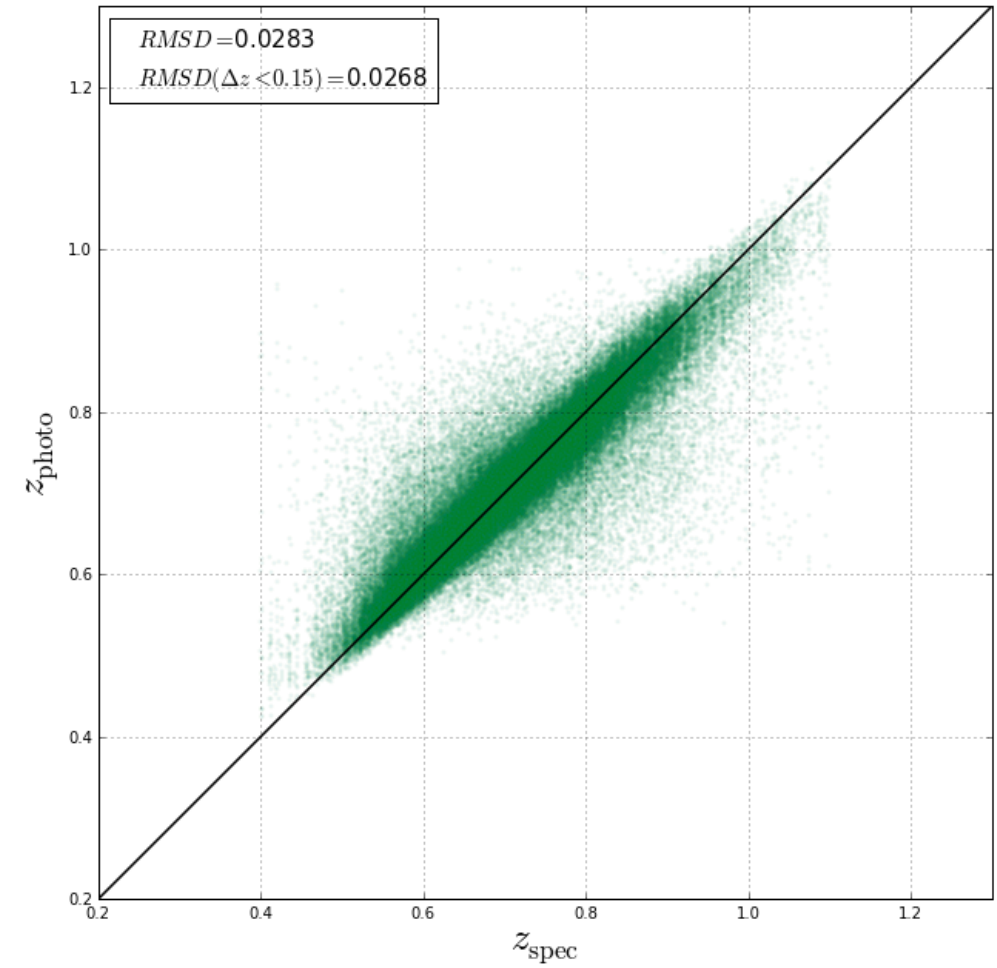


Figure 5.3: Fraction of outliers as a function of the depth of decision trees estimated from Random forest photometric redshifts and spectroscopic redshifts from a test sample. Outliers have been identified as those objects for which the error in predicted photometric redshift is greater than 0.15 (i.e., $\frac{z_{spec}-z_{phot}}{1+z_{spec}} > 0.15$)

the observed density of targets to maps of local imaging conditions and Galactic structure. We perform a multivariate linear regression analysis of the observed surface density against the set of identified potential systematics and simultaneously fit for their impact, rather than correlating against one at a time. This has the advantage of producing a model of systematic-affected density that provides accurate predictions for the combined effects of all the systematics considered.

Using the results of the regression analysis we assemble maps of the observed density and the predicted density in different redshift bins. We then identify regions within our footprint where the total span of target density fluctuation is less than 15% and consider the portion of sky with larger variations to be contaminated at an unacceptable level based on prior experience in BOSS (Ross et al., 2011a). This analysis, similar to the one performed in Section 3.7.1, results in establishing the "mask" of the *eBOSS* LRG survey.



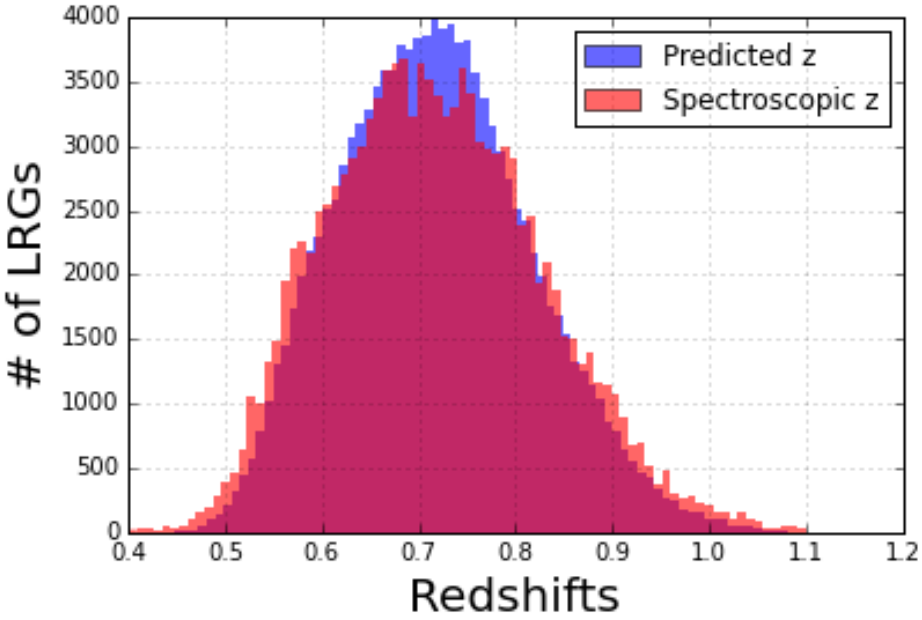


Figure 5.4: The photometric redshifts vs spectroscopic redshifts plots for the test sample using Random forest routine. The blue histogram represents the photo-z's and the red histogram represents the spectroscopic redshifts.

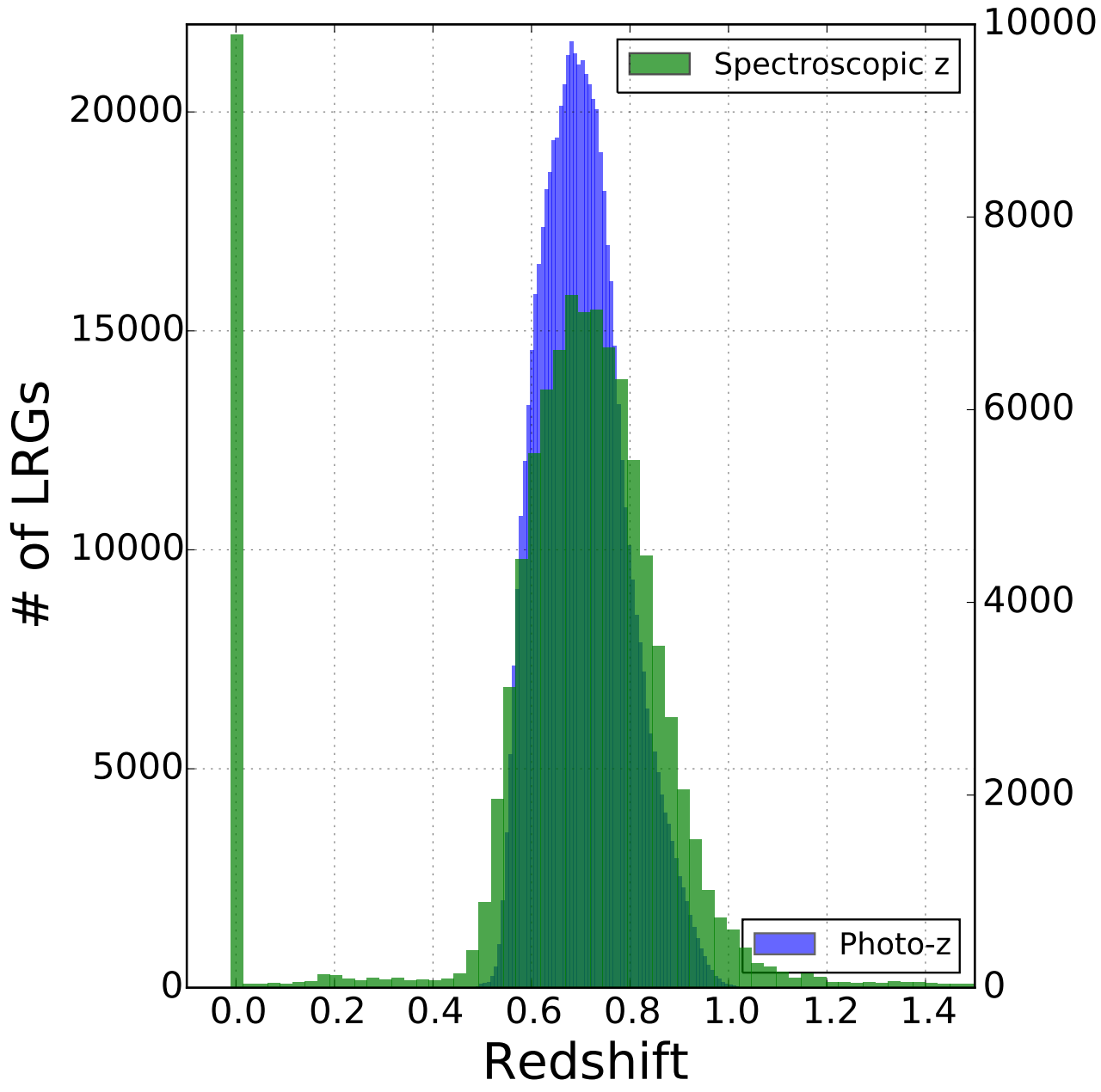


Figure 5.5: The photometric redshift histogram of the entire *eBOSS* LRG sample estimated using Random forest algorithm. The green bars represent the distribution of spectroscopic redshifts from the first two years of *eBOSS* observations. The stellar contamination, estimated to be ∼11%, is visible as the bar at $z \sim 0$.

5.4.1 Imaging systematics associated with WISE and SDSS

As described in Section 3.7.1, we have identified a broad set of seven imaging parameters that could affect *eBOSS* target selection. We reproduce them here for convenience:

1. `W1covmedian`: The median number of single-exposure frames per pixel in the WISE W1 band.
2. `moon_lev`: The fraction of frames that were contaminated with scattered moonlight in the WISE W1 band.
3. `W1median`: The median of accumulated flux per pixel in the WISE W1 band measured in units of *DN* (data number).[1].
4. **Galactic Latitude**: used as a proxy for stellar contamination.
5. **Galactic extinction**: We use r band extinction, as given by Schlegel et al. (1998a).
6. **FWHM** in the *SDSS*z-band: We use FWHM as an estimate of the 'seeing' or imaging quality for the SDSS imaging.
7. **SKYFLUX** in the *SDSS*z-band: the background sky level affects the detection of faint objects which is more difficult in the brighter regions of the sky.

We create all-sky maps of the WISE systematics using the metadata tables associated with the Atlas images and source tables provided by WISE survey team; `W1covmedian`, `W1median`, and `moon_lev` are all quantities in these tables.[2] We also create similar SDSS maps over the entire SDSS footprint. We use the seeing and the sky background in the z band since the *eBOSS* LRG selection algorithm is flux-limited in that bandpass filter. We present these maps in Figures 5.6, 5.7, 5.8, and 5.9.

5.4.2 Regression analysis using photometric redshifts

Next, we break the sky up into equal-area pixels of 0.8 deg^2 using a *Healpix* (Górski et al., 2005) pixelization scheme (NSIDE = 64) and weight all pixels equally. Using the methods described in Section 3.7.1, we calculated the observed density, SD_{obs}, and the predicted

[1] The accumulated photons in each pixel are represented by a number in units of *DN*
[2] `http://wise2.ipac.caltech.edu/docs/release/allsky/expsup\/sec2_4f.html`

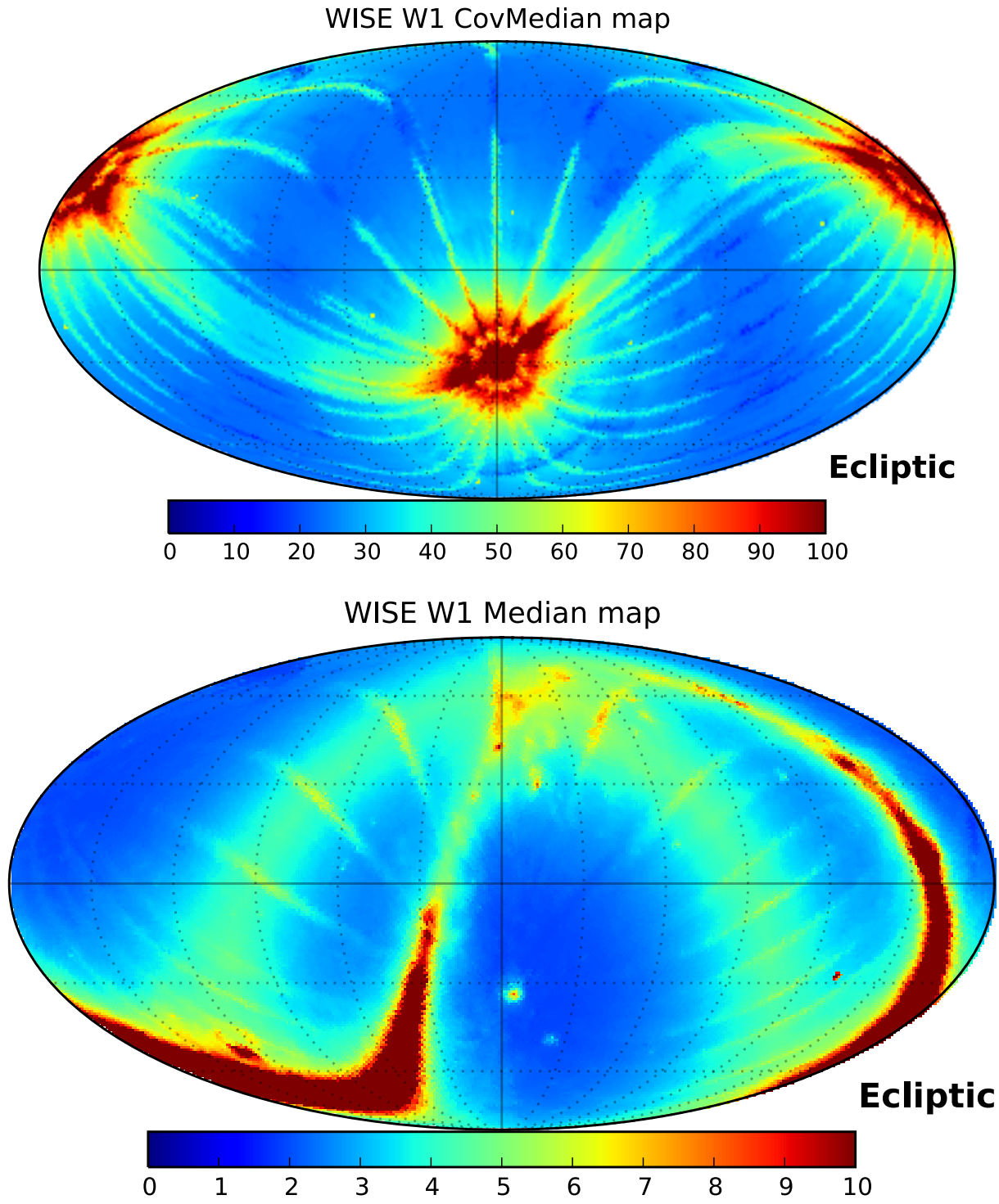


Figure 5.6: WISE all-sky systematics map for Median and Coverage Median. The maps are made using Healpix pixelization scheme (NSIDE=512). W1 covmedian is defined as the median number of single-exposure frames per pixel in the WISE W1 band. W1 median (bottom panel) is defined as the median of accumulated flux per pixel in the WISE W1 band measured in units of *DN* (data number).

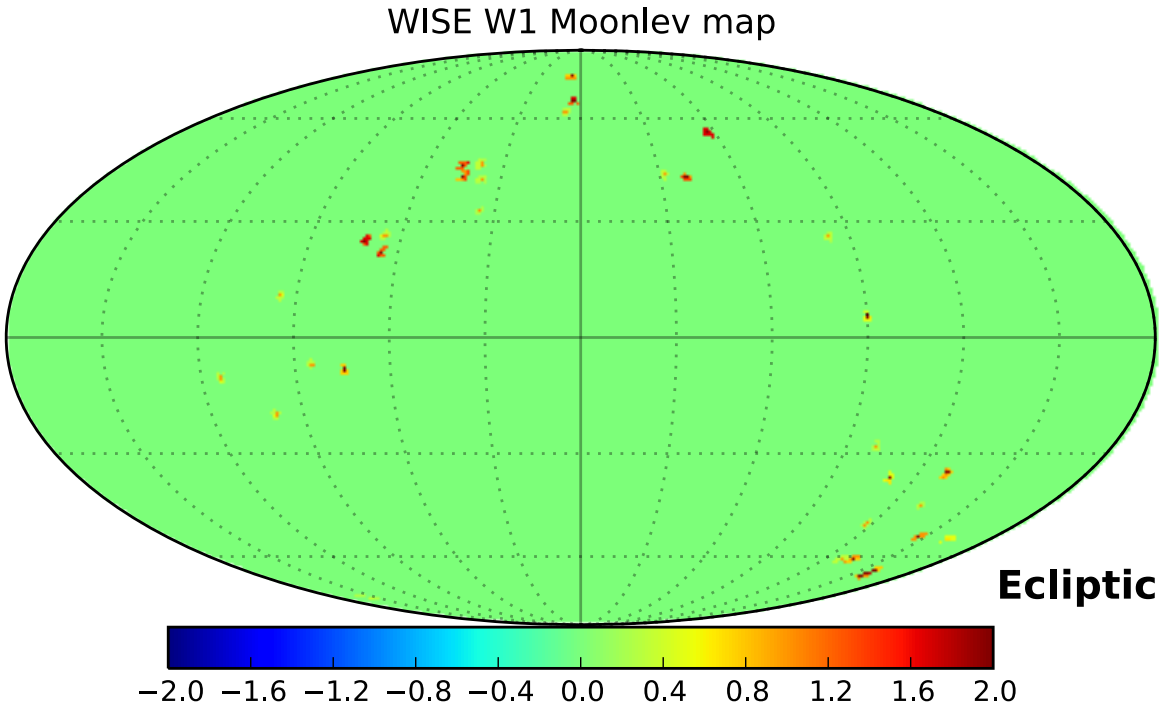


Figure 5.7: WISE all-sky systematics map for moon light contamination. The maps are made using Healpix pixelization scheme (NSIDE=512). W1 moonlev is defined as the fraction of frames that were contaminated with scattered moonlight in the WISE W1 band

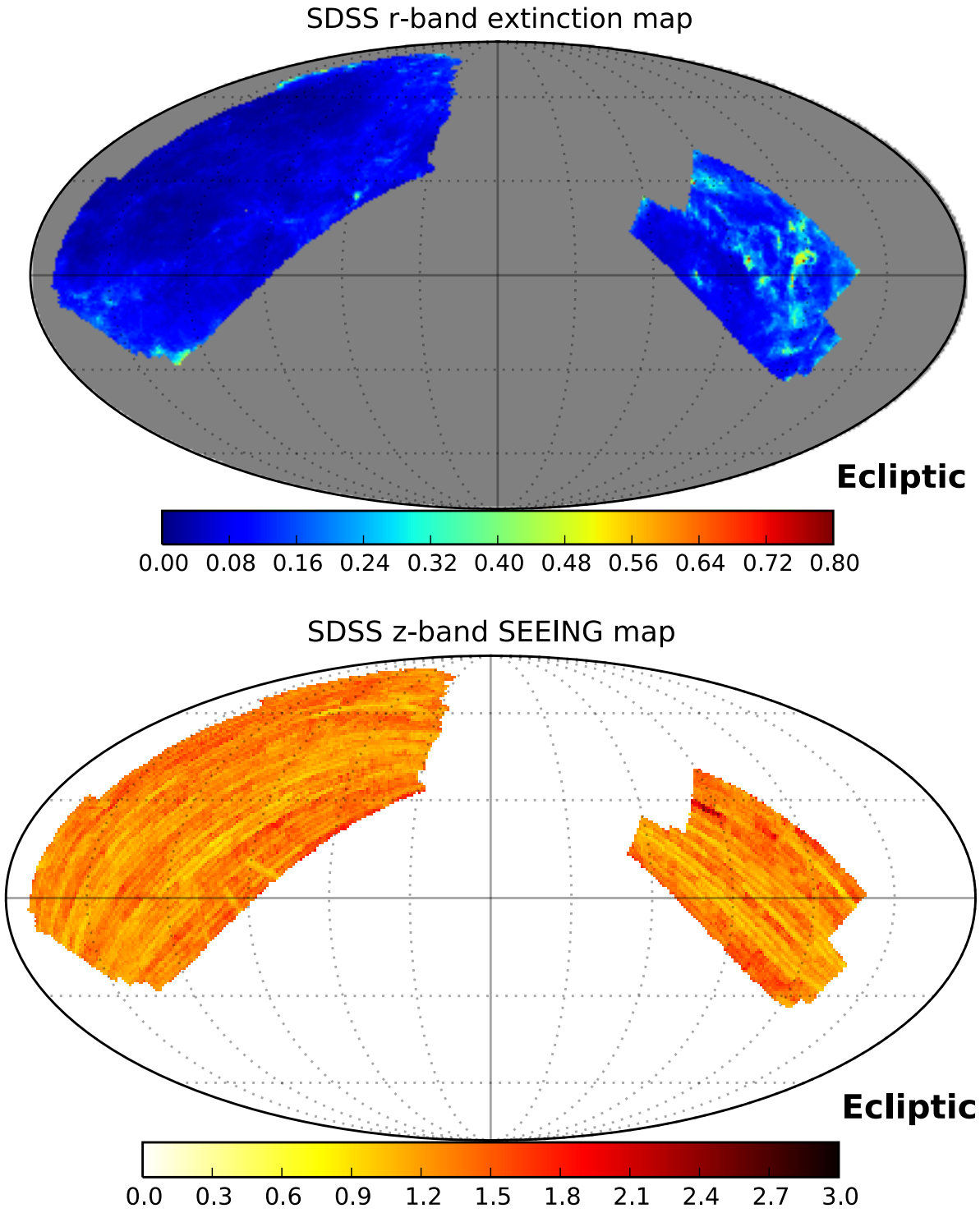


Figure 5.8: SDSS systematics map for dust extinction in r-band and observing conditions. LRG sample is flux limited in z-band. The maps are made using Healpix pixelization scheme (NSIDE=512).

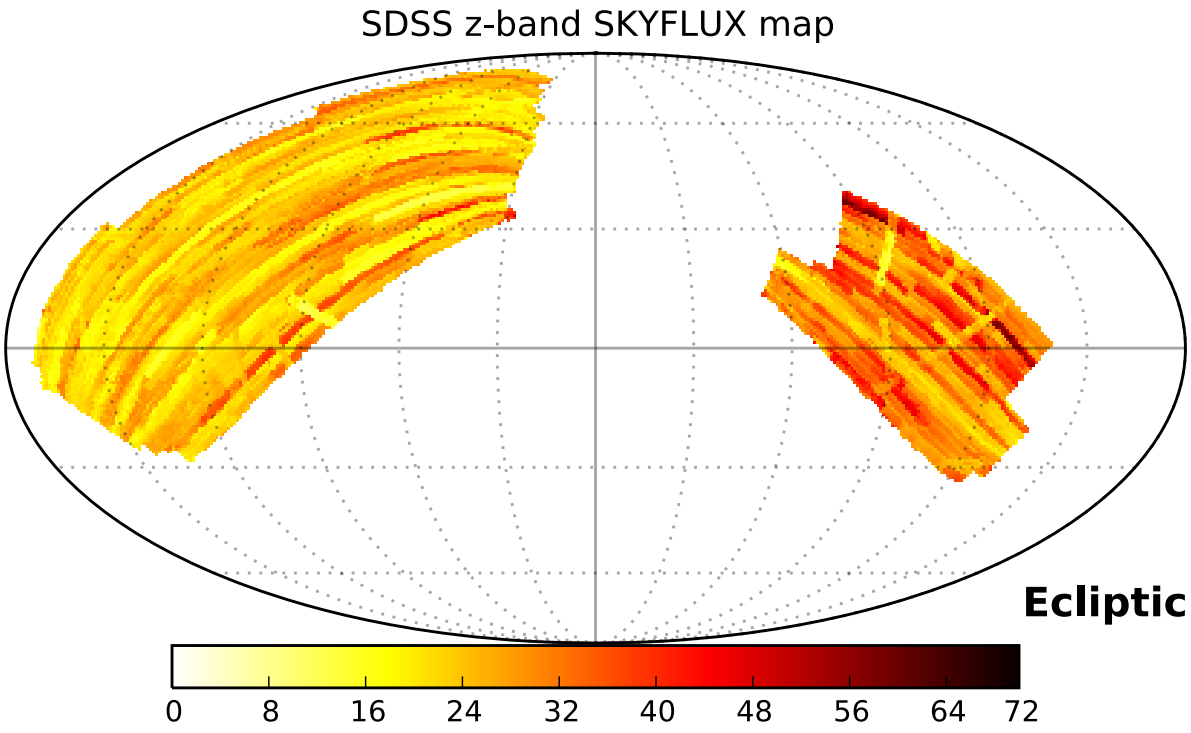


Figure 5.9: SDSS systematics map for sky background in z-band. The maps are made using Healpix pixelization scheme (NSIDE=512).

surface density, PSD. This analysis is restricted over the SDSS footprint as dictated by the SDSS imaging systematic maps and LRG footprint.

5.4.3 Analysis of regression results

The PSD is highly useful for testing the uniformity of the target sample. Multivariate linear regression provides an acceptable best-fit model which has been demonstrated in Chapter 3 by the residual analysis in Figure 3.6. This regression analysis is performed by dividing the LRG sample in redshift bins of $\Delta z = 0.05$ using the photometric estimations from Section 5.3. The *Predicted Surface Density* (PSD) from regression is shown along side the observed LRG density in Figures 5.10, 5.11, 5.12, 5.13, 5.14, and 5.15.

In addition to the regression analysis in different redshift bins, we also perform the regression over the entire sample spanning the whole redshift range $0.55 < z \sim 1.0$. This regression analysis allows us to determine what fraction of the survey footprint satisfies the requirement of less than 15% total variation in target density as described in Section 3.2.2. When the effect of systematics exceeds the 15% level, that area of the survey is excised from clustering analyses. This 15% window is not necessarily symmetric around the mean, so we fix its limits such that the footprint area satisfying the requirement is maximized. We depict the observed surface density, the predicted surface density, and the mask of the survey across the whole footprint of *SDSS* in Figures **??** and **??**. The LRG sample is estimated to be uniform and homogeneous over $\sim 92\%$ of the *BOSS* survey footprint, showing little or no dependence on imaging systematics and flux calibrations.

5.4.4 Random catalogs

The construction of random catalogs is an important step in measuring clustering. To extract precise clustering measurements, the randoms should reflect the underlying redshift distribution of the datasets (i.e., the distribution with no fluctuations from clustering). In addition, the randoms should also reflect the effects of all the imaging systematics such as seeing, sky brightness, Galactic extinction, etc., which may alter the target density in a complex manner. In a nutshell, the random sample should demonstrate the true redshift

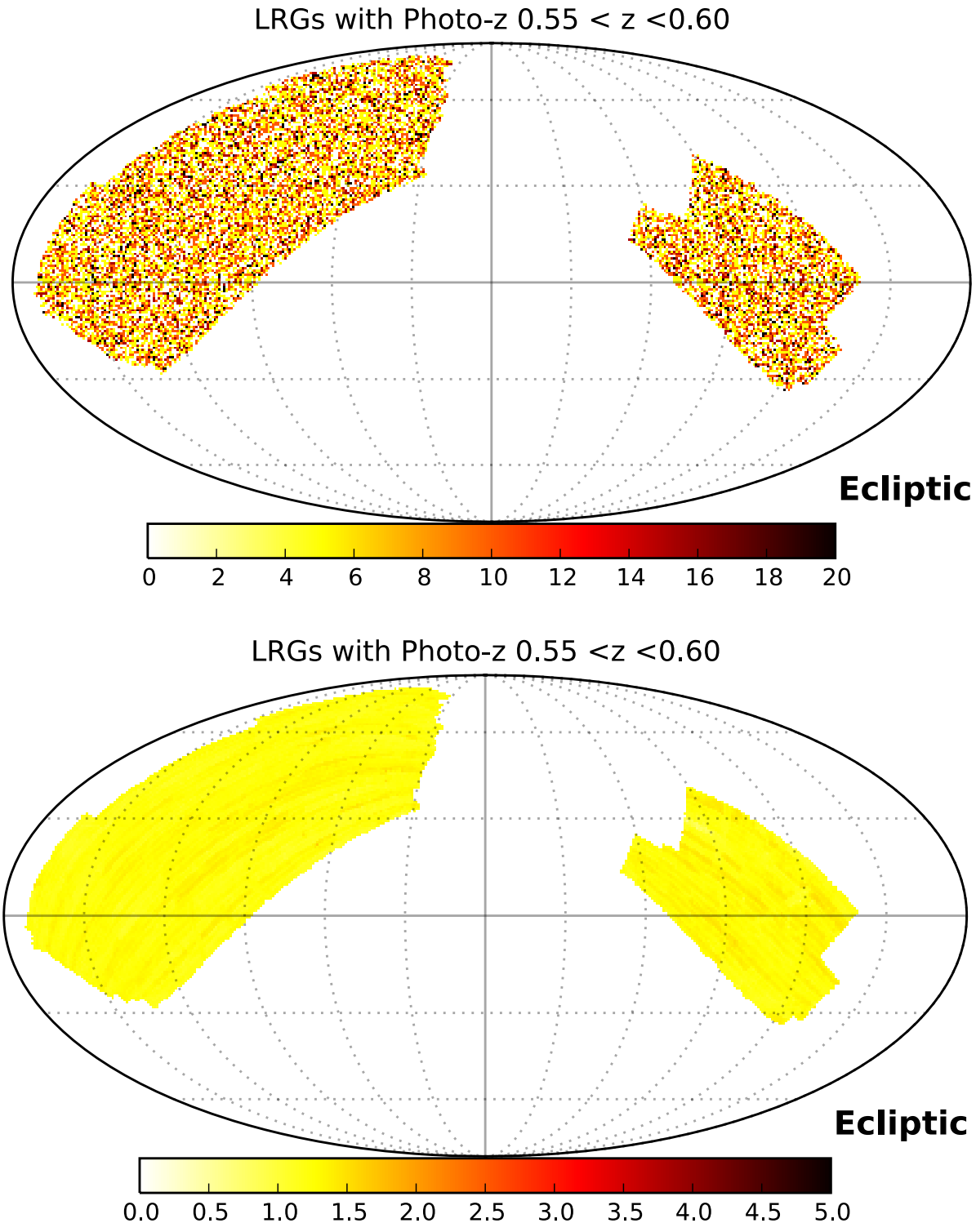


Figure 5.10: LRG predicted density and observed density at redshifts $0.55 < z < 0.60$. The maps are made using Healpix pixelization scheme (NSIDE=512). The observed density (top panel) of obtained by splitting the sample based on random forest photo-z's. The predicted density (bottom panel) is the result of multivariate regression of the observed density in this bin against the source of systematic errors.

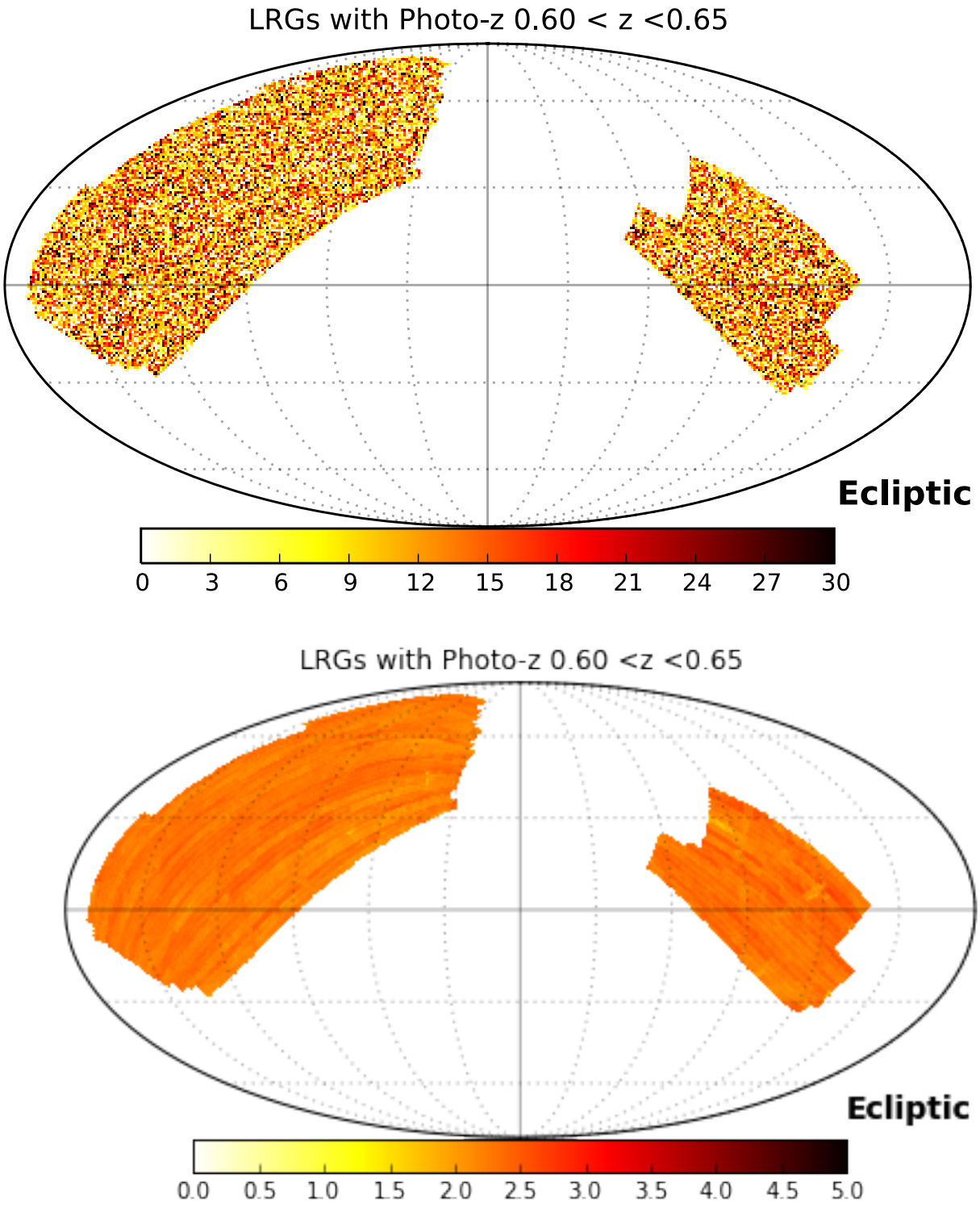


Figure 5.11: LRG predicted density and observed density at redshifts $0.60 < z < 0.65$. The maps are made using Healpix pixelization scheme (NSIDE=512). The observed density (top panel) of obtained by splitting the sample based on random forest photo-z's. The predicted density (bottom panel) is the result of multivariate regression of the observed density in this bin against the source of systematic errors.

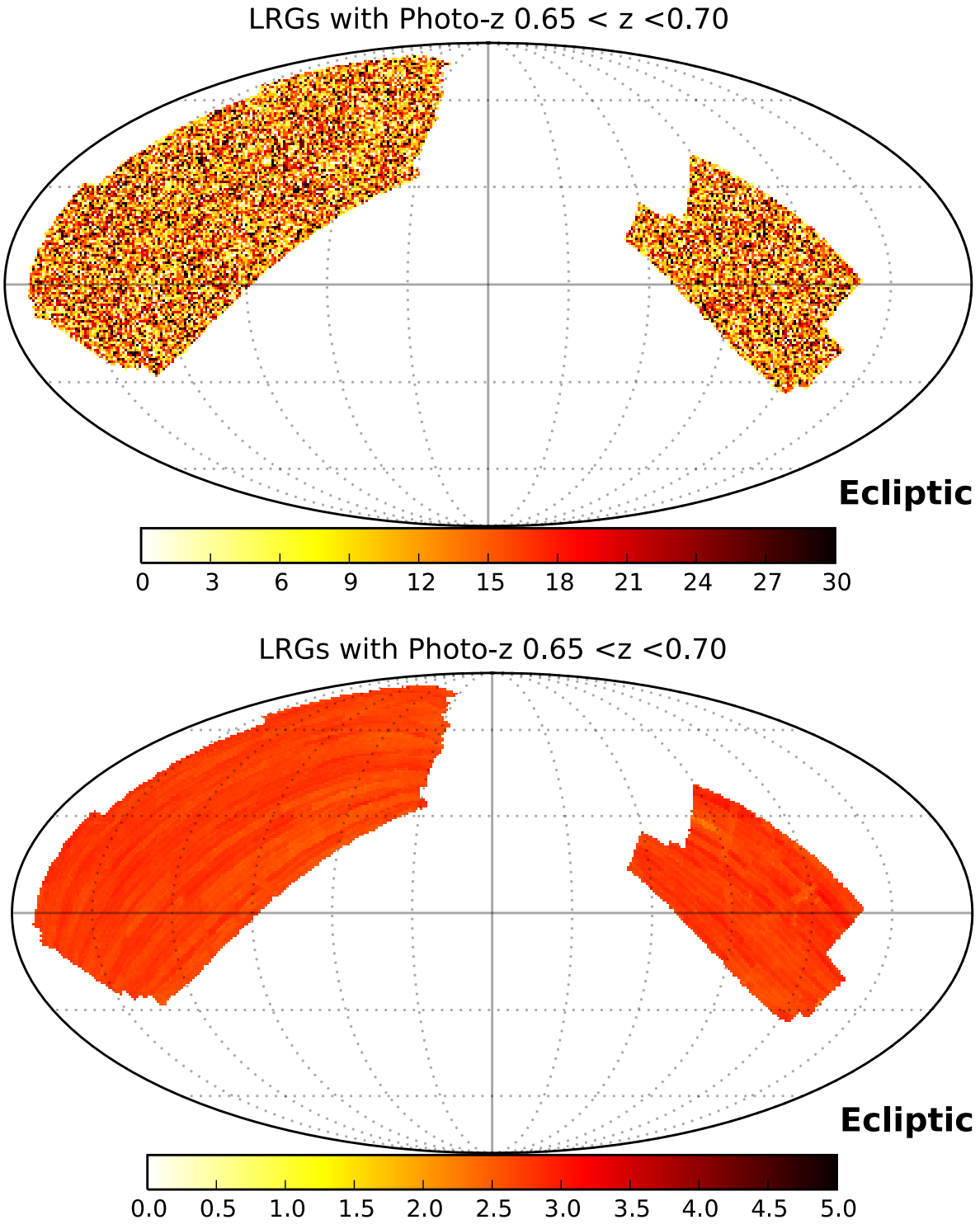


Figure 5.12: LRG predicted density and observed density at redshifts $0.65 < z < 0.70$. The maps are made using healpix pixelization scheme (NSIDE=512). The observed density (top panel) of obtained by splitting the sample based on random forest photo-z's. The predicted density (bottom panel) is the result of multivariate regression of the observed density in this bin against the source of systematic errors.

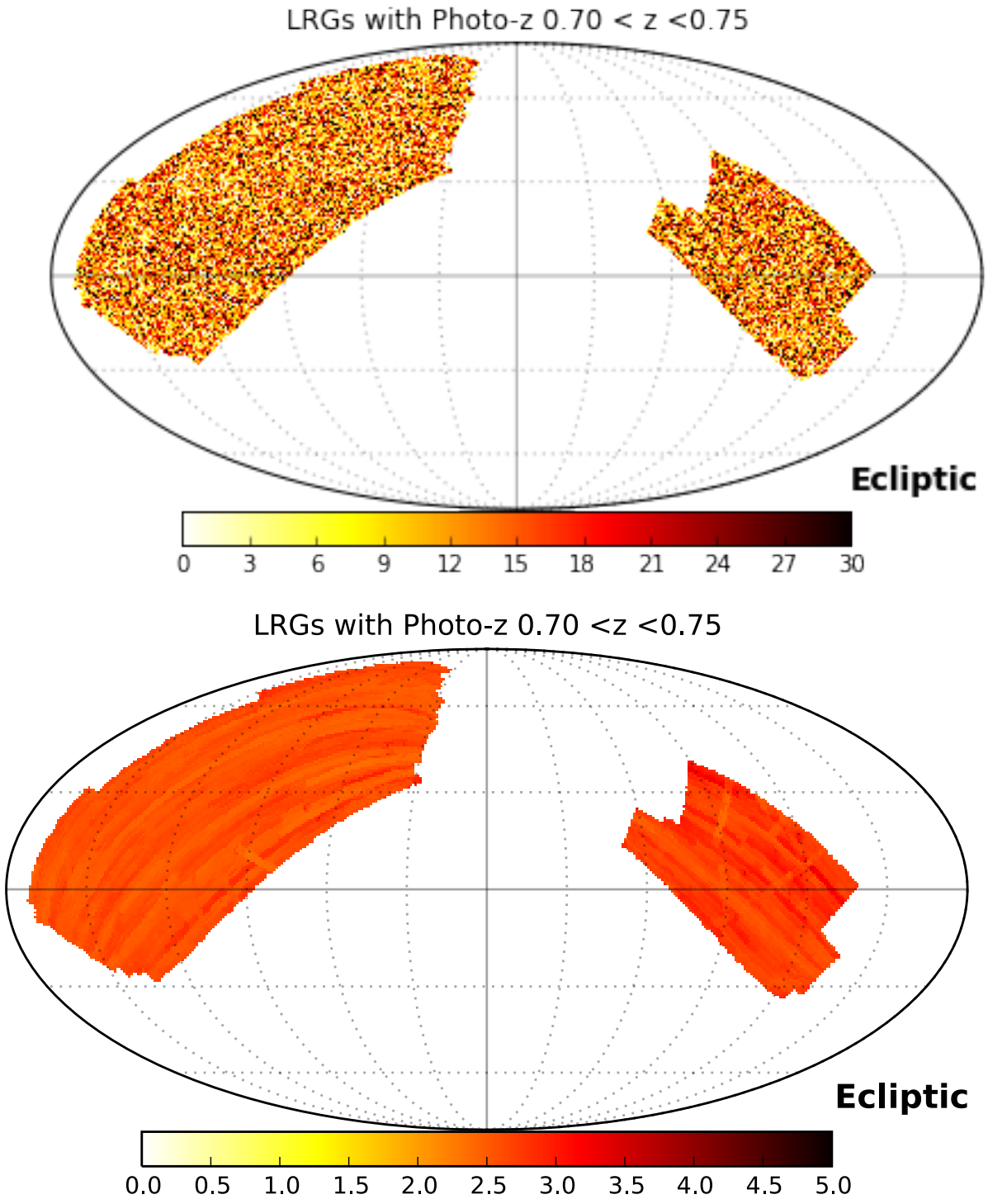


Figure 5.13: LRG predicted density and observed density at redshifts $0.70 < z < 0.75$. The maps are made using healpix pixelization scheme (NSIDE=512). The observed density (top panel) of obtained by splitting the sample based on random forest photo-z's. The predicted density (bottom panel) is the result of multivariate regression of the observed density in this bin against the source of systematic errors

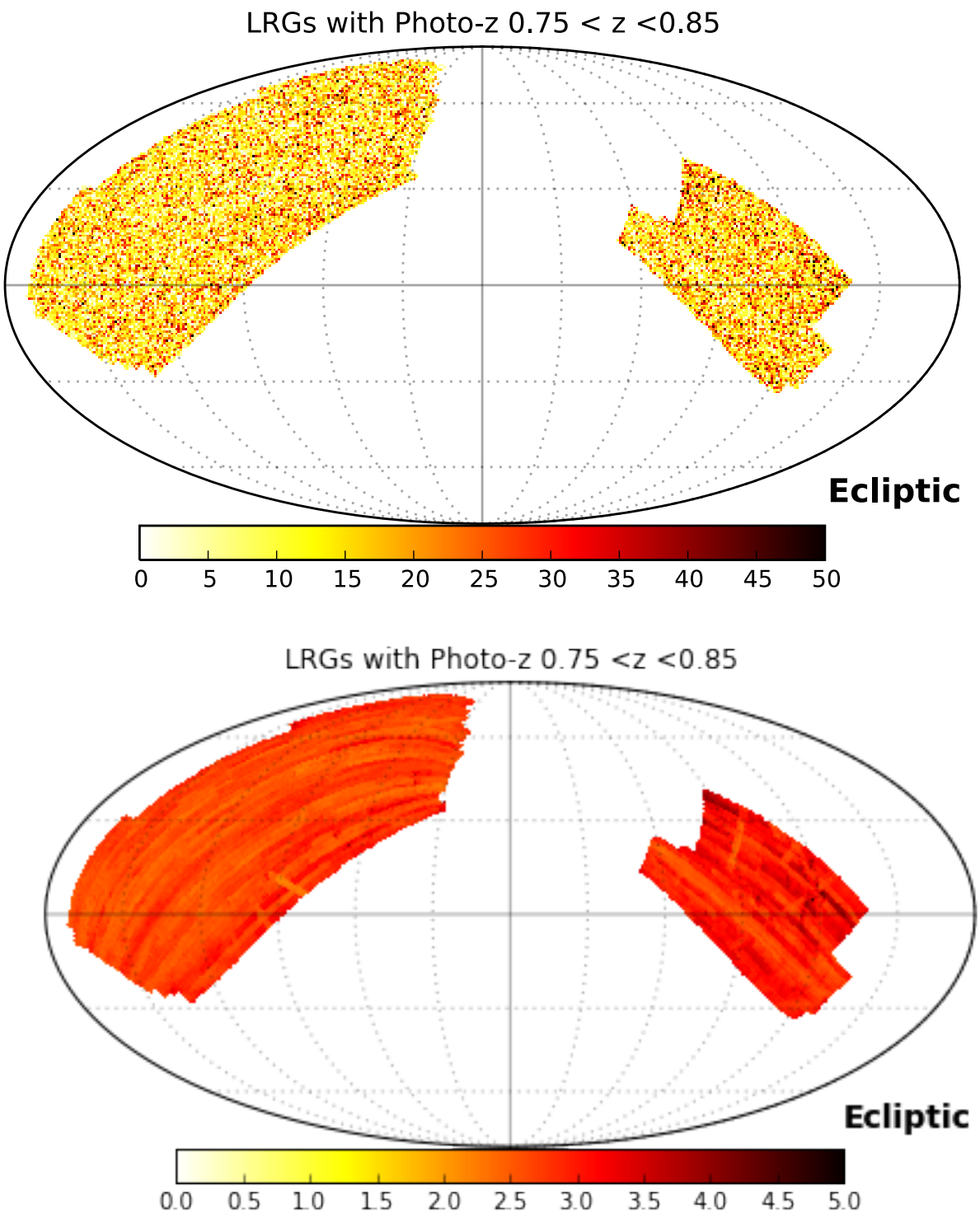


Figure 5.14: LRG predicted density and observed density at redshifts $0.75 < z < 0.85$. The maps are made using Healpix pixelization scheme (NSIDE=512). The observed density (top panel) of obtained by splitting the sample based on random forest photo-z's. The predicted density (bottom panel) is the result of multivariate regression of the observed density in this bin against the source of systematic errors

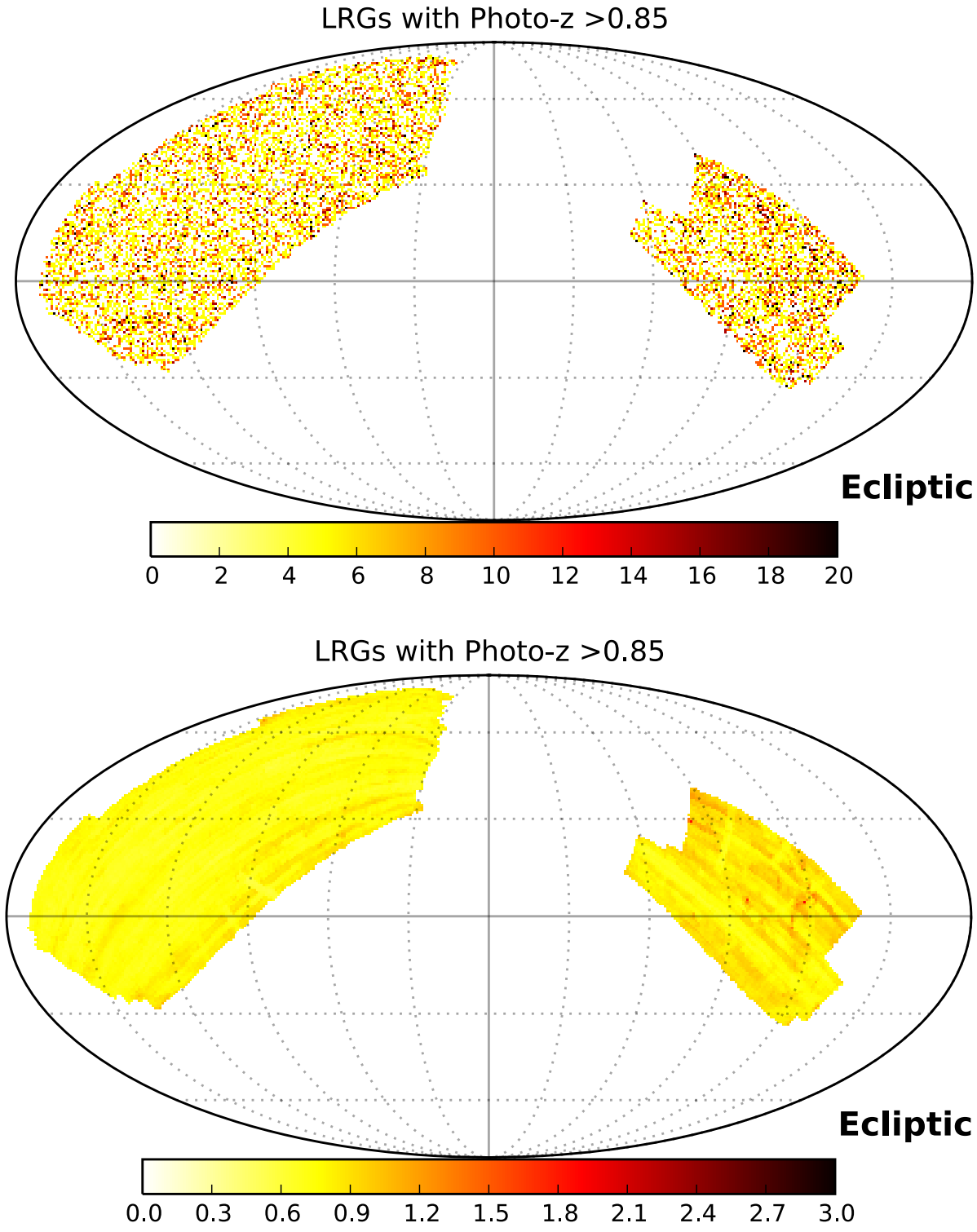


Figure 5.15: LRG predicted density and observed density at redshifts $0.85 < z \sim 1.0$. The maps are made using Healpix pixelization scheme (NSIDE=512). The observed density (top panel) of obtained by splitting the sample based on random forest photo-zs. The predicted density (bottom panel) is the result of multivariate regression of the observed density in this bin against the source of systematic errors

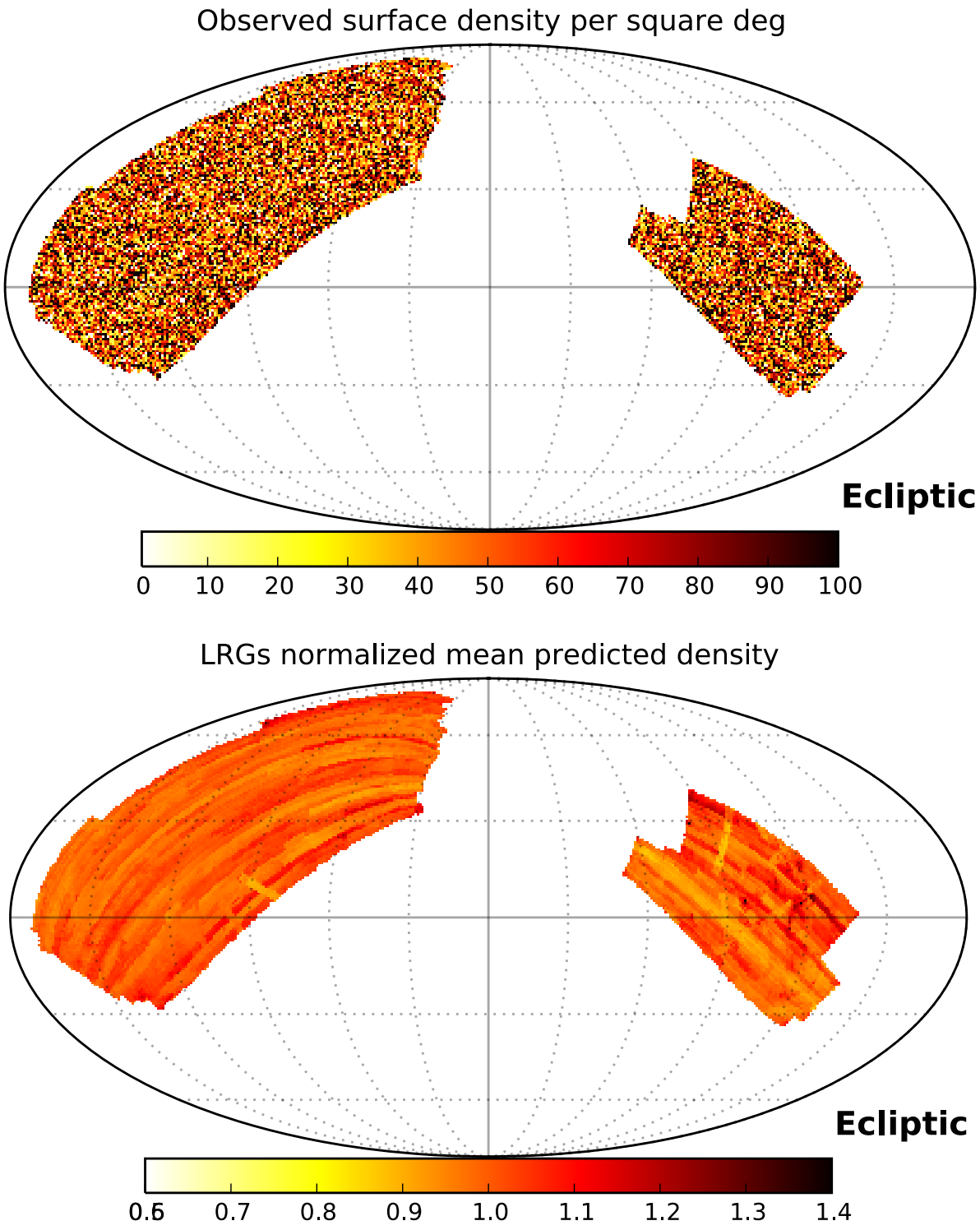


Figure 5.16: *eBOSS* LRGs observed and predicted density maps over the entire redshift range $0.5 < z < 1.0$. The maps are made using Healpix pixelization scheme (NSIDE=512). The observed density (top panel) represents the full sample. The predicted density (bottom panel) is the result of multivariate regression of the observed density against the sources of systematic errors.

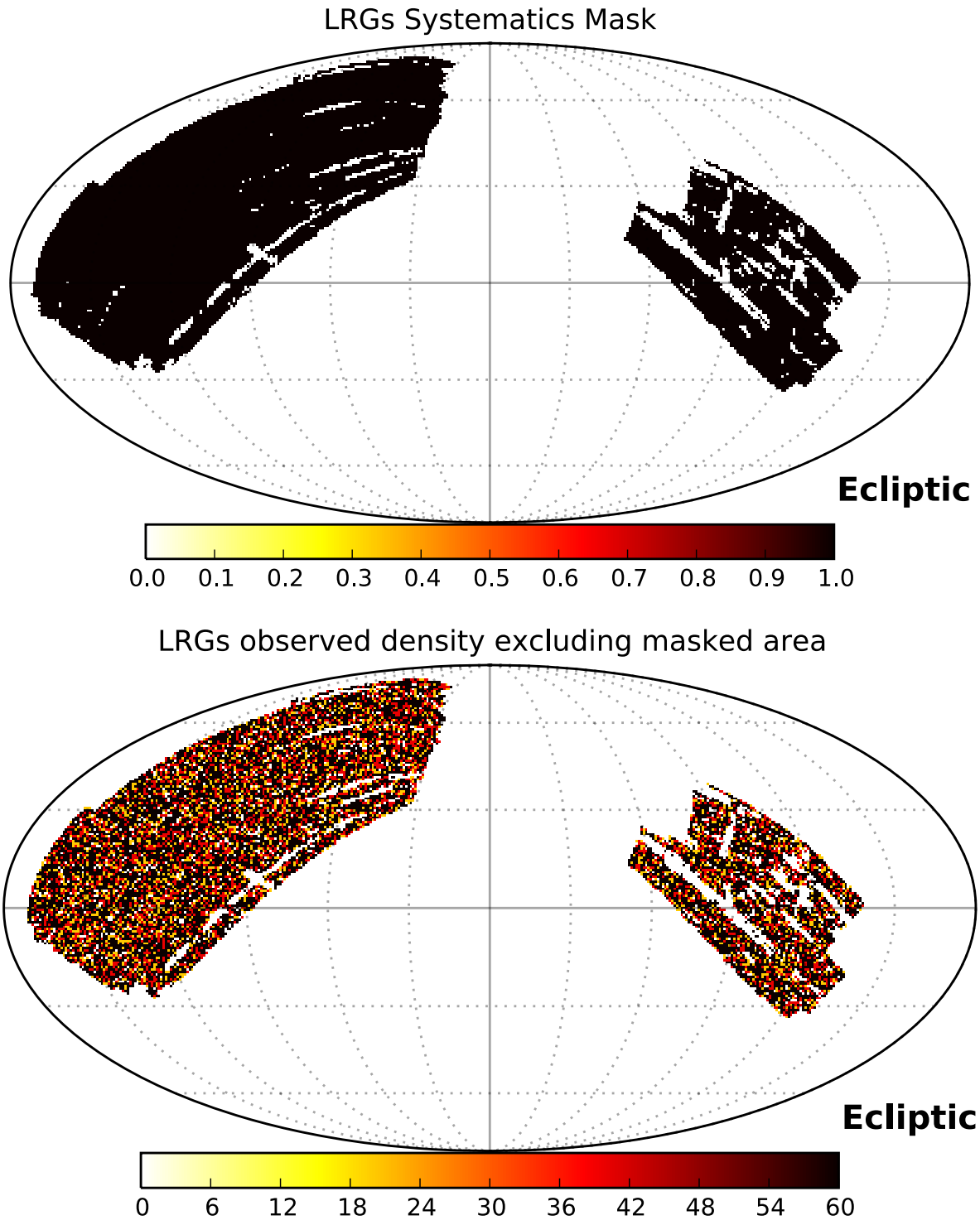


Figure 5.17: *eBOSS* LRG survey mask based on predicted density and the sample post masking. The maps are made using Healpix pixelization scheme (NSIDE=512). The survey (top panel) represents the regions showing less than 15% variation in the density around mean. The regions showing more than 15% variation in target density is excluded from LSS studies. The LRG sample post masking is shown in the bottom panel.

distribution of the datasets, and include all the effects which are not due to the clustering at relevant scales, including the mask of the survey (i.e., the detailed geometry of the area covered).

In the simplest approach, we generate random points with identical limits on sky coverage as the datasets. These points are assigned redshifts which are randomly drawn from the data. In addition, we associate a random number, $\Re$, between 0 to 1 to each point that will be used to modulate the density of the randoms. We then normalize the *PSD* maps presented in the previous section to the same scale and refer to it as the *Normalized Predicted Surface Density*, *NPSD* (i.e., the maximum value of $NPSD$ over the entire redshift range is 1). Now that the *NPSD* maps and the random factor associated with each point are on the same scale of 0 to 1, the random points with $\Re$ values greater than that of the *NPSD* in that pixel are thrown away. The remaining points provide a random sample which not only matches the actual redshift distribution of a dataset but also includes the density modulation effect of systematics. We present the resulting random samples in Figures 5.18 and 5.19.

5.5 THE 2-POINT CORRELATION STATISTICS

The two-point autocorrelation function, which is typically referred to simply as the correlation function, is the excess probability of finding a galaxy in a volume element, dV, at a separation, r, from another randomly chosen galaxy:

$$\mathrm{dP} = \mathrm{n}[1 + \xi(\mathrm{r})]\mathrm{dV}, \tag{5.1}$$

where n is the mean number density of the galaxies in question (Peebles, 1980). The excess probability means that the strength of correlation would be zero for uniformly distributed objects.

To measure the correlation function, we have to measure the counts of the pairs of galaxies as a function of their separation and normalize it by what is expected from a uniform distribution. This is typically done by constructing a random sample which is populated by random points but has the same three-dimensional distribution as the data barring variations

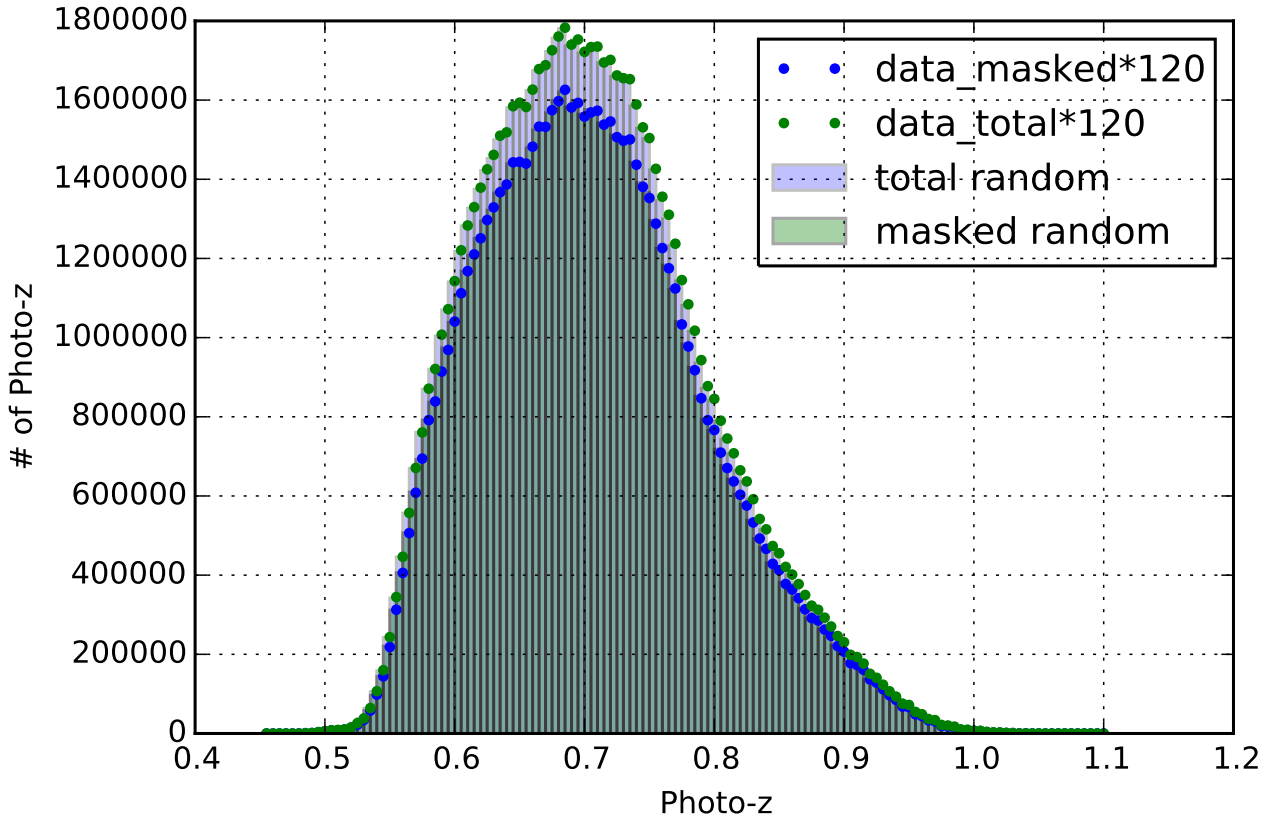


Figure 5.18: The redshift distribution of the eBOSS LRG and random sample. The data histogram is scaled by a constant factor to be more easily compared to the random histogram.

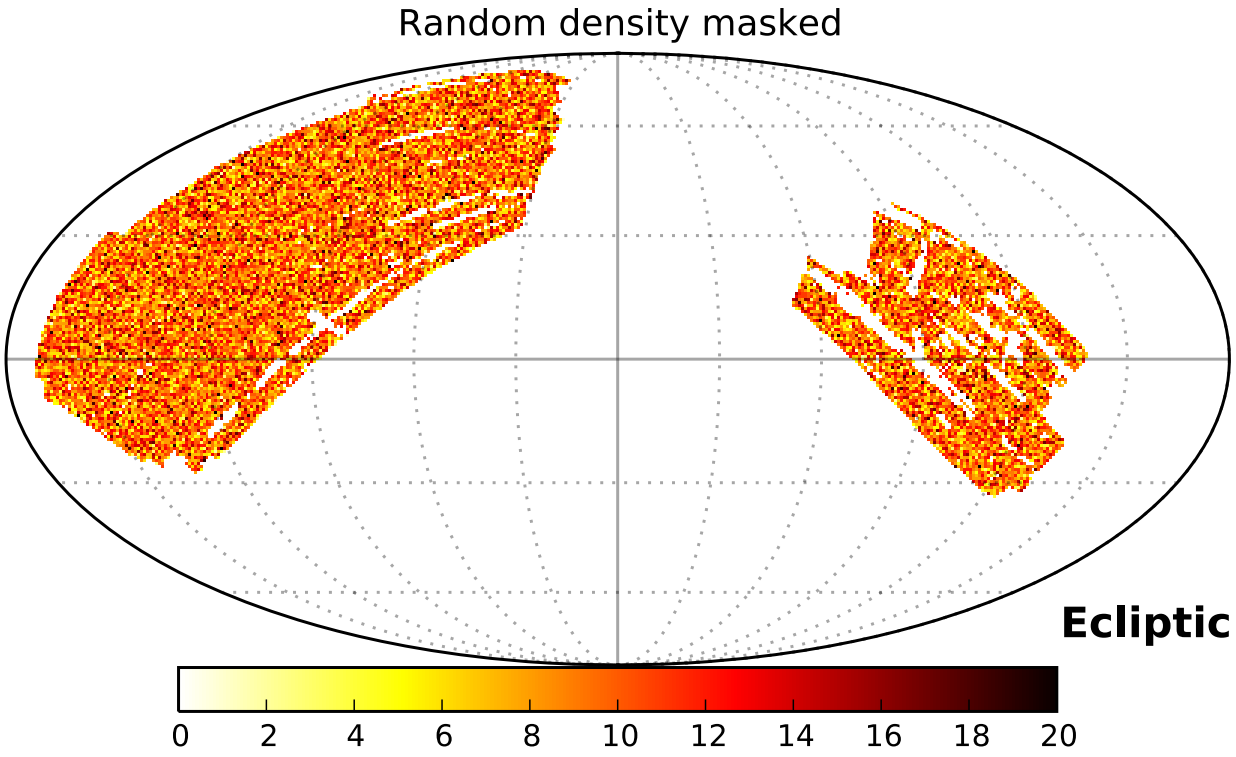


Figure 5.19: The random sample covers a footprint identical to that of the eBOSS LRG sample. Only a subset of total random sample is presented here for efficiency. The color bar represents the surface density in numbers per pixel. The pixels are 0.01 $\deg^2$ created using Healpix (NSIDE= 512).

due to the correlations we wish to measure and hence should have identical sky coverage and redshift distribution. The ratio of galaxy pair counts observed from the data relative to pair counts described using random catalogs is used for estimating $\xi(r)$.

A few estimators have been proposed and tested over time. In this work, we use the Landy & Szalay estimator (LS, Landy & Szalay, 1993) to measure the two-point autocorrelation function in LRG samples. The LS estimator has been found to be better at handling edge effects and reaches minimum variance compared to other more naive estimators (Kerscher et al., 2000):

$$\xi_{\rm LS} = \frac{({\rm DD} - 2{\rm DR} - {\rm RR})}{\rm RR}, \tag{5.2}$$

where DD are data pairs, DR are data-random pairs, and RR are random-random pairs at a given separation. If one could obtain a map of galaxies in real space (rather than redshift space), one could compute the two point autocorrelations function, $\xi(r)$ from this estimator. Once computed, $\xi(r)$ can be modeled approximately as a power law, $\xi(r) = (\frac{r}{r_0})^{\gamma}$. Here, r_0 is the characteristic scale length of galaxy clustering, defined as a length scale at which $\xi(r) = 1$. The canonical value of the scale parameter γ is $\sim$1.8 and has remain unchanged for more than 30 years (Peebles, 1980; Totsuji & Kihara, 1969).

In real galaxy surveys, we measure positions in redshift space, not real space, so we must use other statistics to constrain $\xi(r)$. This is complicated further when redshift information is not available for a given sample, as it is observationally expensive. However, it is possible to measure the spatial distribution of galaxies in two dimensions as projected on the sky. This is measured as the projected angular correlation function, w(θ). This quantity is defined, similarly to three-dimensional correlation function, as the excess probability of finding a galaxy at an angular separation of θ:

$${\rm dP} = {\rm n}[1 + {\rm w}(\theta)]{\rm d}\Omega, \tag{5.3}$$

where n is the mean number density of the galaxies per steradian and $d\Omega$ is the solid angle of a second galaxy at a separation θ from a randomly chosen galaxy. The angular correlation function can be approximated as a power law too, w(θ) = $A_w\theta^{\delta}$. Here A_w is the clustering amplitude and δ=(1-γ) is the slope of the correlation function. As the two-dimensional

clustering seen on the sky is simply the projection of the three-dimensional clustering along the line of sight, w(θ) is directly related to its three-dimensional analog $\xi(r)$. It is possible to estimate $\xi(r)$ from w(θ) using Limber's equation (Limber, 1954):

$$w(\theta) = \int_0^\infty y^4 \phi^2(y) dy \int_0^\infty \xi(\sqrt{x^2 + y^2\theta^2}) dx, \tag{5.4}$$

where y is the comoving distance and $\phi(y)$ is the radial selection function normalized such that $\int y^2\ \phi(y)\ dy = 1$. If one assumes a power-law form for the three-dimensional correlation function, $\xi(r) = (\frac{r}{r_0})^\gamma$, then the parameters, r_0 and γ can be estimated from amplitude and slope of angular clustering:

$$r_0^\gamma = \frac{A\ \Gamma(\frac{\gamma}{2})}{\Gamma(\frac{\gamma-1}{2})\ \Gamma(\frac{1}{2})}, \tag{5.5}$$

where Γ is the usual gamma function.

In this work, we measure the angular correlation, $w(\theta)$, at different redshifts using photometric redshifts and obtain the values of the parameters, A and δ, using a power-law fit. These parameters are optimized using the least square errors technique. These parameters can then be used to infer the three dimensional correlation function, $\xi(r)$ in future, using equation 5.5.

5.6 RESULTS AND DISCUSSION OF LRG CLUSTERING

We present the results of our LRG angular correlation function measurements in Figures 5.20, 5.21, and 5.22 as a function of angular separation on the sky. The angular clustering is computed using the software package CUTE (Alonso, 2012). In Figures 5.23 and 5.24, we show the clustering as a function of physical scales as projected on the sky. The solid line is the power-law fit and the parameters are estimated using least square error. The angular correlation appears to be consistent with a power law. However, we do not see a clear and obvious acoustic peak at $\sim$100h^{-1}Mpc. Theoretical estimations using Fisher matrix predict a $\sim 2\sigma$ detection of BAO for the given errors on photometric redshifts and

the number of galaxies per unit redshift interval, $\frac{dN}{dz}$. In practice, the actual detection will always be smaller than this since Fisher matrix predictions do not always take into account the systematic errors. For the *eBOSS* LRG sample, we have pushed the photometry to its limit causing larger errors on magnitude and photo-z's. We suspect that our non-detection is primarily due to systematic errors. In addition, a combination of one or more of the factors below could be responsible for diluting BAO signal:

1. **Stellar contamination**: The average stellar contamination estimated in the spectroscopic sample is $\sim 11\%$. This is expected to be higher in the photometric sample. We have attempted to use machine learning methods, but have not found a reliable way to improve star-galaxy separation for this sample. Machine learning methods explained in this work have failed to implement star-galaxy separation at desirable efficiency. This is expected to be one of the main effects diluting the BAO signal. Stellar contamination would not affect the BAO peak but weakens the BAO signal overall.
2. **Systematic effects**: Systematics effects may alter target density in a very complex way. The noisy SDSS photometry and the fact that these LRGs are at the edge of photometric detection make things more complex. The BAO peak is an extremely weak signal. It is critical to mitigate systematics effects in order to have a robust detection. Although we have implemented multivariate regression to estimate the combined effect of systematics, more work is required to completely understand these effects. As we accumulate more spectroscopic data, we will get a better understanding of systematic effects.
3. **Probability**: It is also possible that the volume we are probing has a very weak signal. In this case, the detection using photo-z becomes even harder. It is possible that we are probing a volume where any significant detection would require accurate spectroscopic redshifts. As discussed before, BAO is a feature at a scale of $100h^{-1}$Mpc. This implies that even in an enormous volume of $1h^{-1}\text{Gpc}^3$, we would have ~ 1000 realizations of the BAO peak. Due to this, the errors in the correlation function at BAO scale are large which also makes a robust detection difficult.

The values of the power-law fit parameters, A_0 and γ, are consistent with previous studies. This work can be effectively used to constrain the projected accuracy of the 3D-

correlation function for future surveys. We have successfully demonstrated the applicability of machine learning photo-z's for measuring clustering. This is critical for the future generation photometric surveys like LSST, where the spectroscopic follow-up of even a small fraction of detected objects will not be possible.

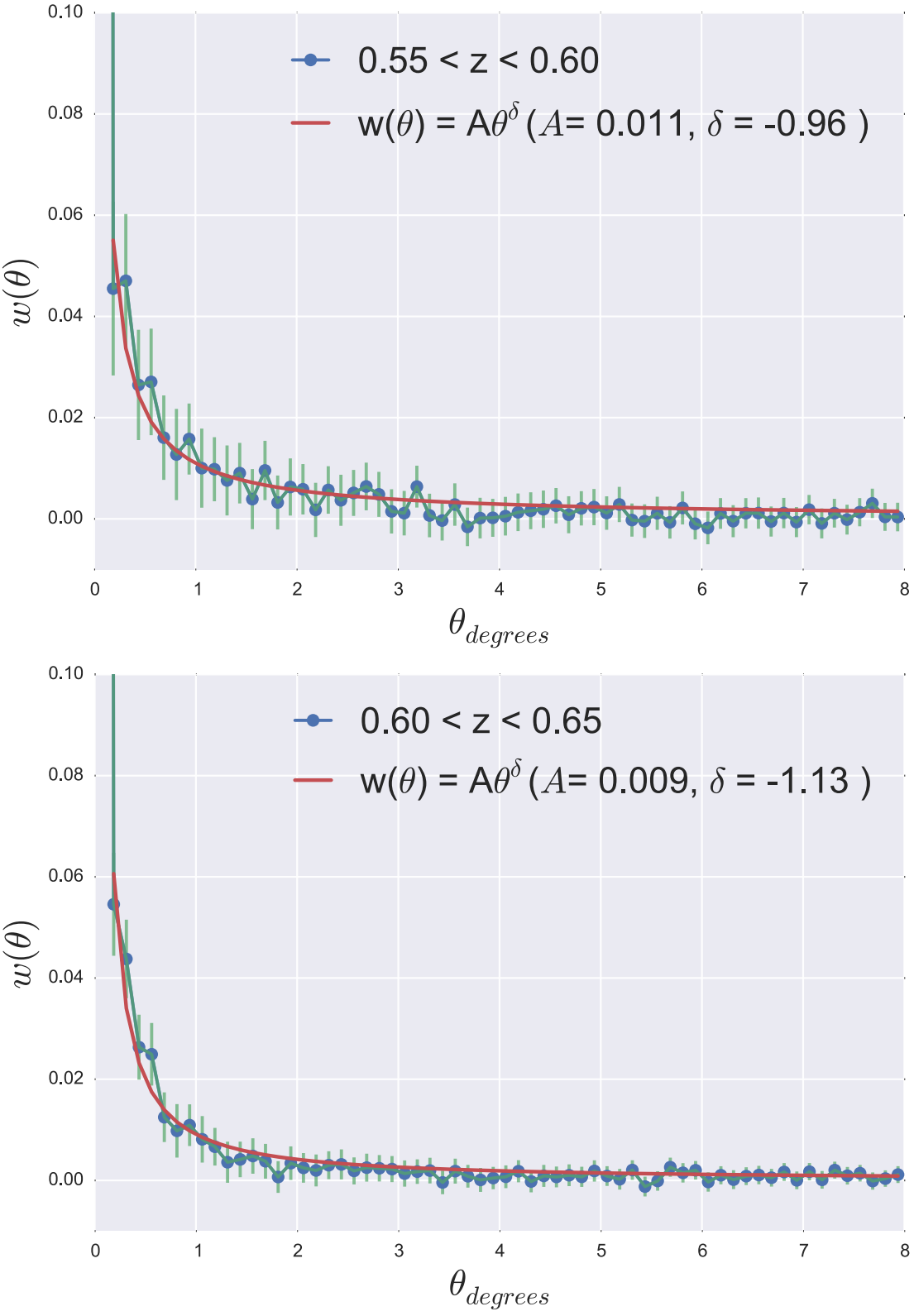


Figure 5.20: The clustering of the eBOSS LRGs as a function of angular separation at redshifts $0.55 < z < 0.60$ and $0.60 < z < 0.65$. The solid curve represents the power law fit whose parameters are optimized using the least square error method.

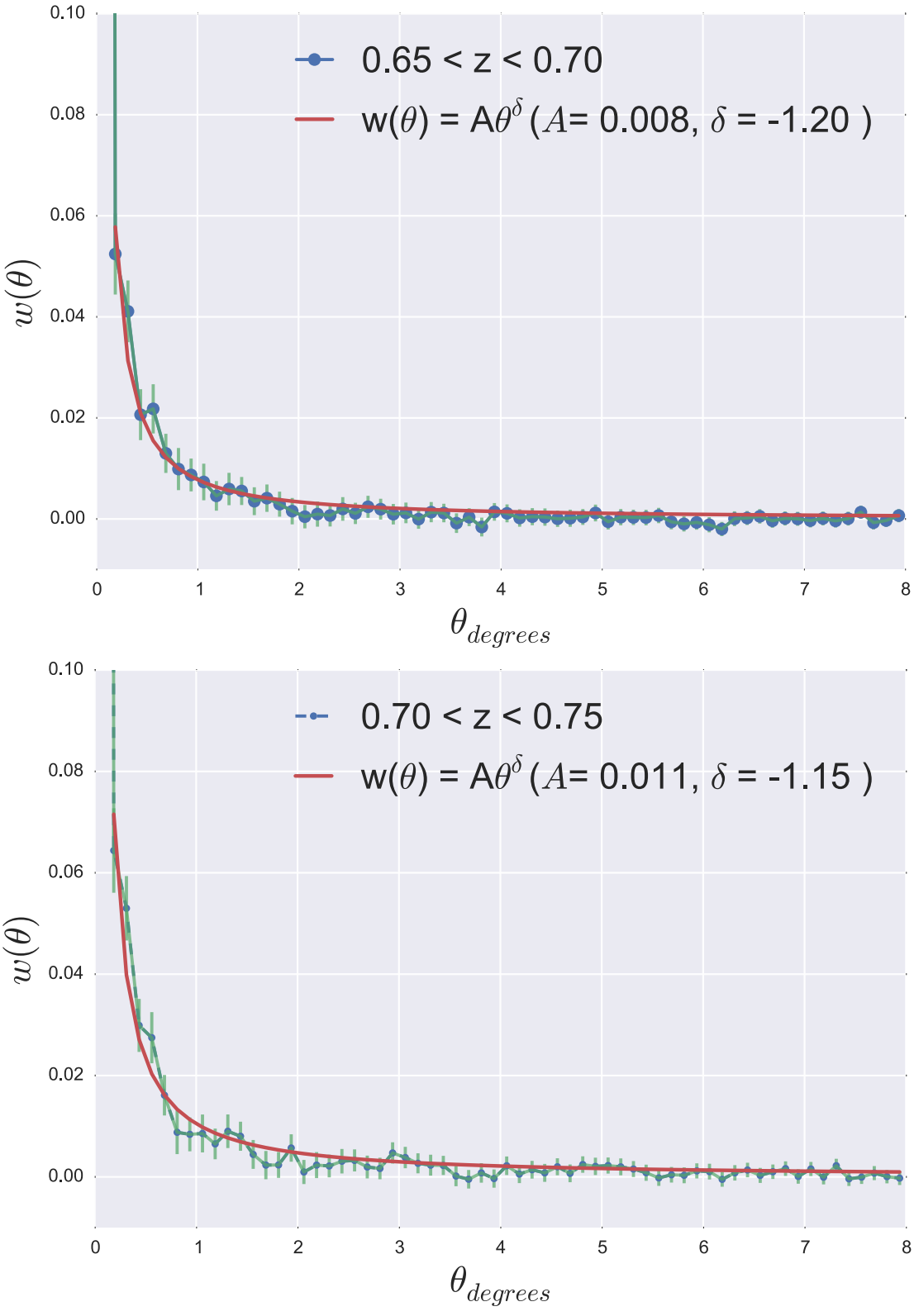


Figure 5.21: The clustering of the eBOSS LRGs as a function of angular separation at different redshifts $0.65 < z < 0.70$ and $0.70 < z < 0.75$. The solid curve represents the power law fit whose parameters are optimized using the least square error method.

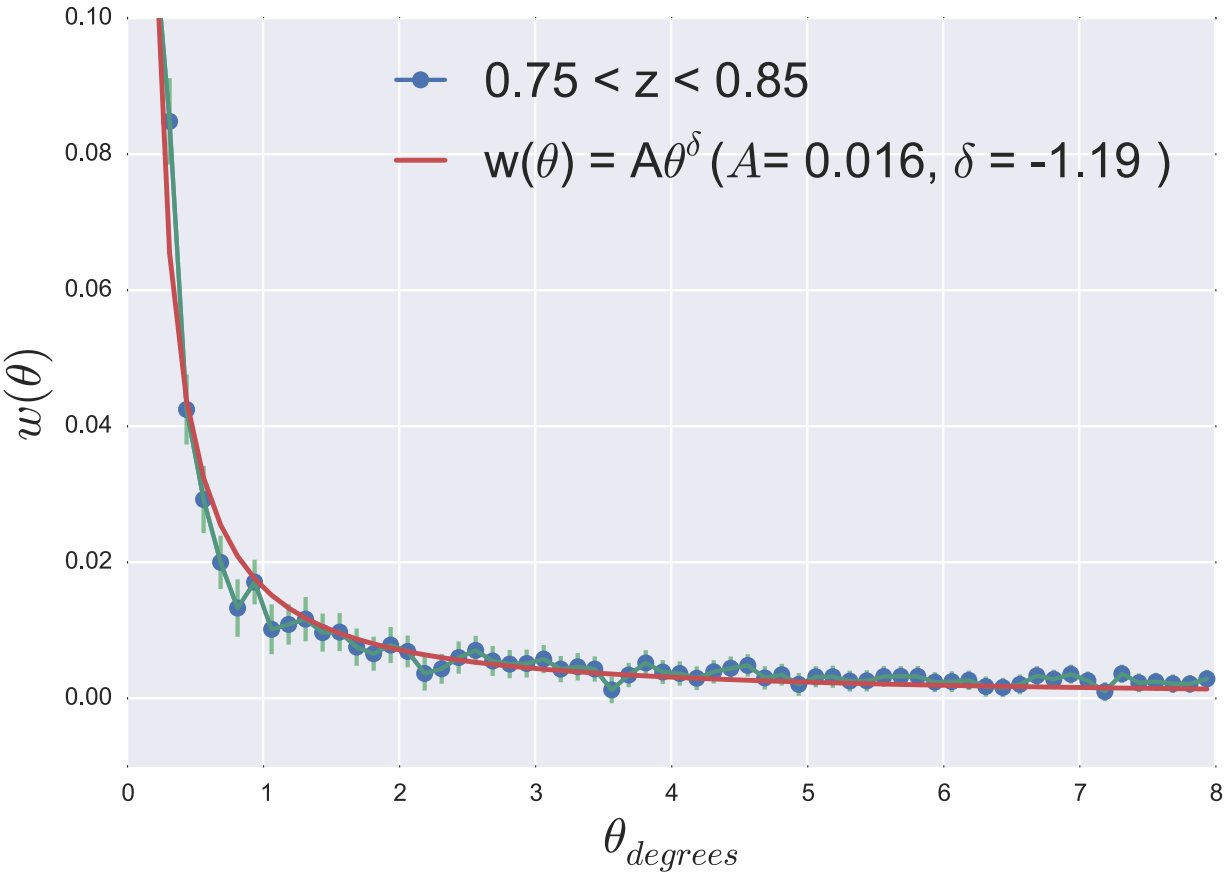


Figure 5.22: The clustering of the eBOSS LRGs as a function of angular separation at redshifts $0.75 < z < 0.85$. The solid curve represents the power law fit whose parameters are optimized using the least square error method.

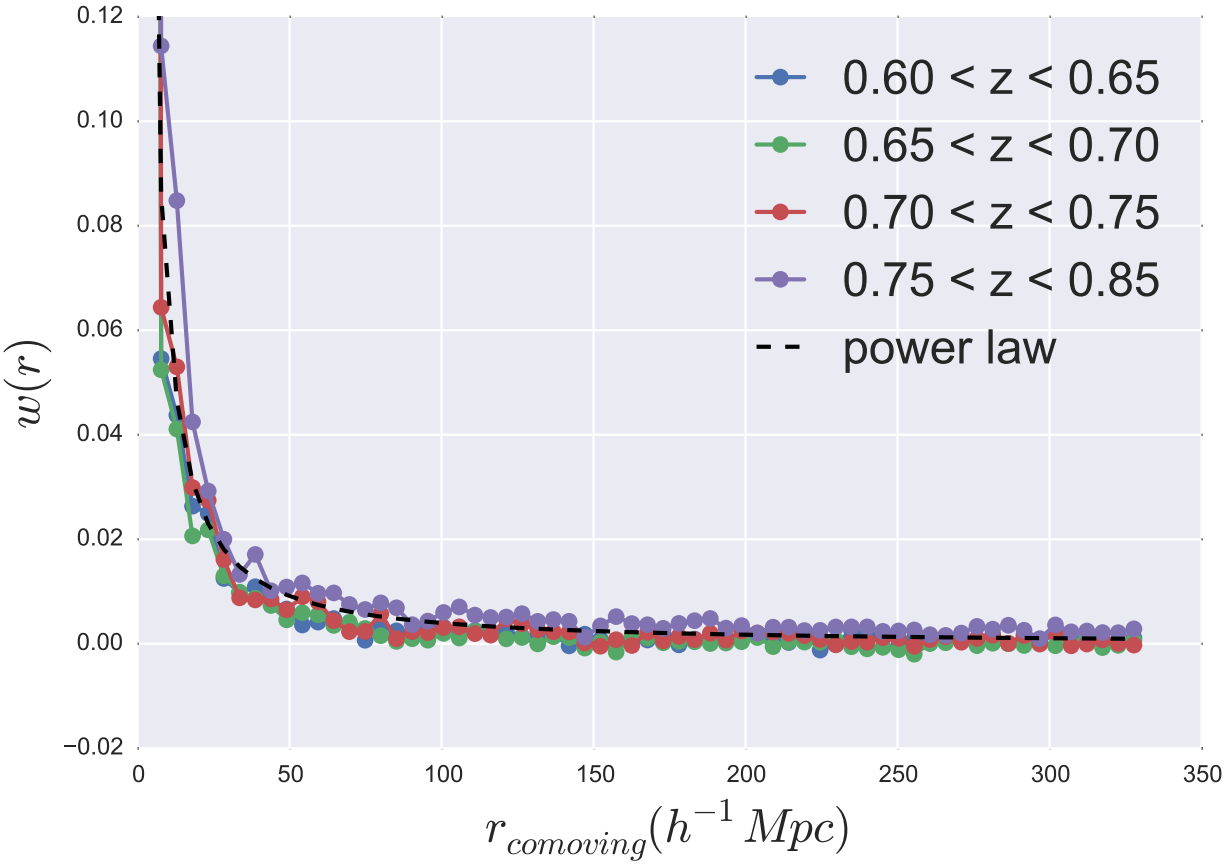


Figure 5.23: The clustering of the eBOSS LRGs as a function of their physical separation at different redshifts ranging from 0.55 to 0.85. The black dashed curve is the power-law fit to the clustering.

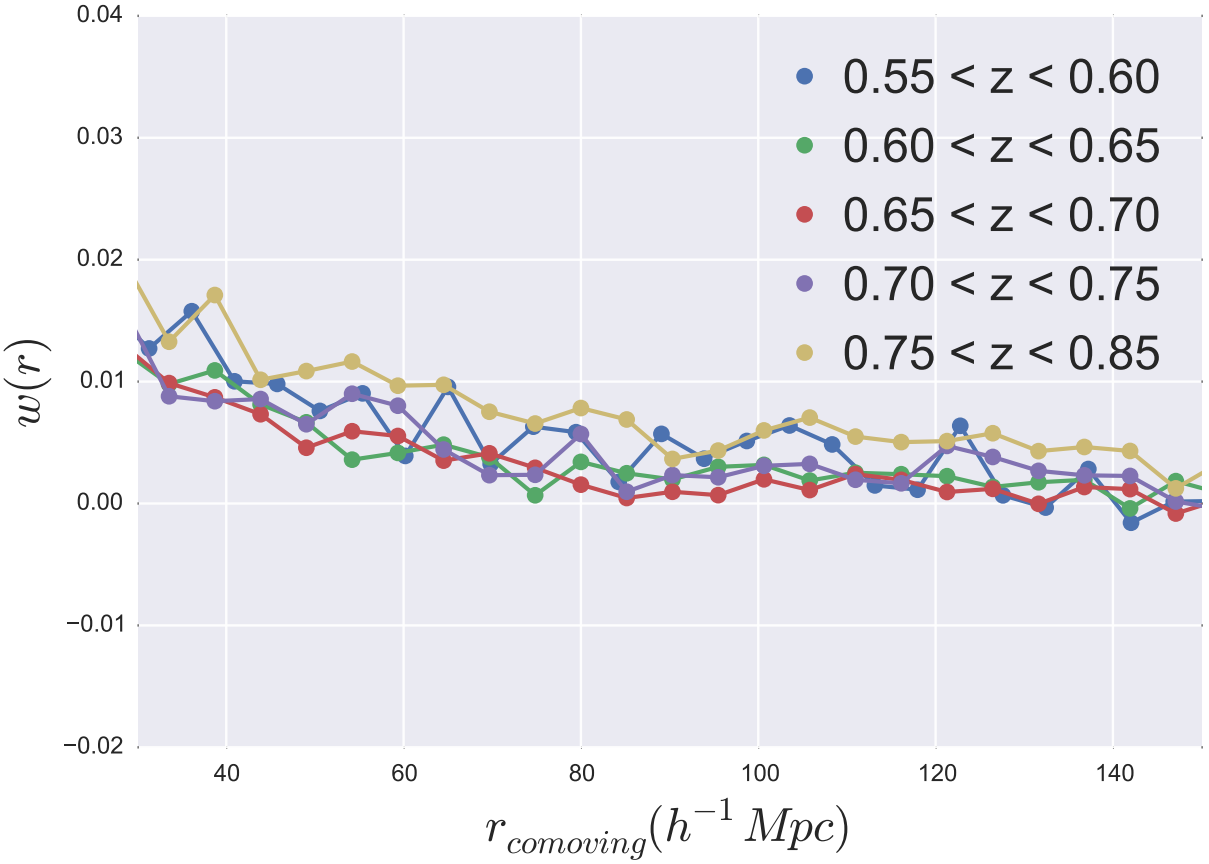


Figure 5.24: The clustering of the eBOSS LRGs as a function of their separation at different redshifts, zoomed in. We do not see an obvious BAO peak which seems to be due to a combination of factors including stellar contamination and systematic effects.

6.0 SDSS-IV/EBOSS: THE CLUSTERING OF QUASARS AND LRGS

6.1 INTRODUCTION

Quasar clustering measurements can be used to infer their active lifetimes and large-scale environments which, in turn, can be used to understand their exact fueling mechanism and the impact they have on galaxy formation and evolution (Kaiser, 1984; Mo & White, 1996). Feedback from active galactic nuclei (AGN) appears to have a profound effect on galaxy formation and evolution (Gebhardt et al., 2000). The observed correlation between central black hole mass and the velocity dispersion of stars in the bulge of a galaxy provides evidence for the possibility that AGN feedback influences their host galaxies (Springel et al., 2005). Unfortunately, quasars are found to have relatively low number density, which causes large uncertainties in clustering amplitudes estimated from the autocorrelation function especially at low redshifts where they are rare. It is possible to make more precise measurements of quasar clustering by studying the clustering of galaxies around quasars (i.e., by measuring the cross-correlation function of quasars and galaxies), as the number density of galaxies is much higher. Cross-correlation clustering measurements also provide a measure of the local environment in which quasars reside, which is relevant to understanding the physics of quasar fueling (Kauffmann & Haehnelt, 2000).

In this chapter, we present measurements of the quasar LRG cross-correlation function at redshifts $0.55 < z < 1.0$ using data from the first two years of SDSS/eBOSS observations. Although spectroscopically confirmed quasars are available over a much wider redshift range, the redshift range of the LRG sample limits this investigation to the lowest redshifts of *eBOSS* quasars. We measure the clustering strength as a function of scale in the range $0.1\ h^{-1}$Mpc $< r < 30\ h^{-1}$Mpc in redshift bins of $\Delta z = 0.05$.

The chapter is organized as follows. In Section 6.2, we briefly outline the *eBOSS* LRG and quasar sample used for this work and explain the procedure for creating the *eBOSS* large scale structure and random catalogs. In Section 6.3, we explain the cross-correlation statistics which are used for clustering measurements. In Section 6.4, we present the results of cross-correlation measurements of LRGs around quasars. We also present the autocorrelation measurements for the *eBOSS* LRGs, which is then combined with the cross-correlation measurements to give an estimate of quasar bias at different redshifts. We conclude with what this means for the halo masses in which these quasars reside and discuss ways to take this work further.

6.2 DATA

For this study, we use samples of $\sim 30,600$ quasars and $\sim 120,000$ LRGs in the redshift range $0.6 < z < 1.0$ covering $\sim 2,000$ deg^2 of the sky. These LRGs and quasars have spectroscopically confirmed redshifts from the first two years of eBOSS observations. These datasets constitute the large scale structure (LSS) catalog of *eBOSS* created for large scale clustering measurements (Tojero et al., in prep.). The distribution of LRGs and quasars is discontiguous and distributed over both the NGC and SGC portions of the SDSS footprint. The sky map and redshift distribution of both samples are shown in Figures 6.1 and 6.2. The LRGs and quasars from previous programs that used *eBOSS* target selection algorithms such as the SDSS-III/BOSS ancillary program SEQUELS are also included in this investigation. Next, we explain the methods used for creating the *eBOSS* LSS catalogs briefly.

6.2.1 Large scale structure catalogs

The LSS catalog created for *eBOSS* follows the same prescription used for *BOSS* DR12 (Reid et al., 2016). In particular, Laurent et al. (2017) found that the clustering amplitude of the quasar sample and its redshift evolution are consistent with the assumptions used in Zhao et al. (2016), and that the clustering can be modeled with the type of simulation

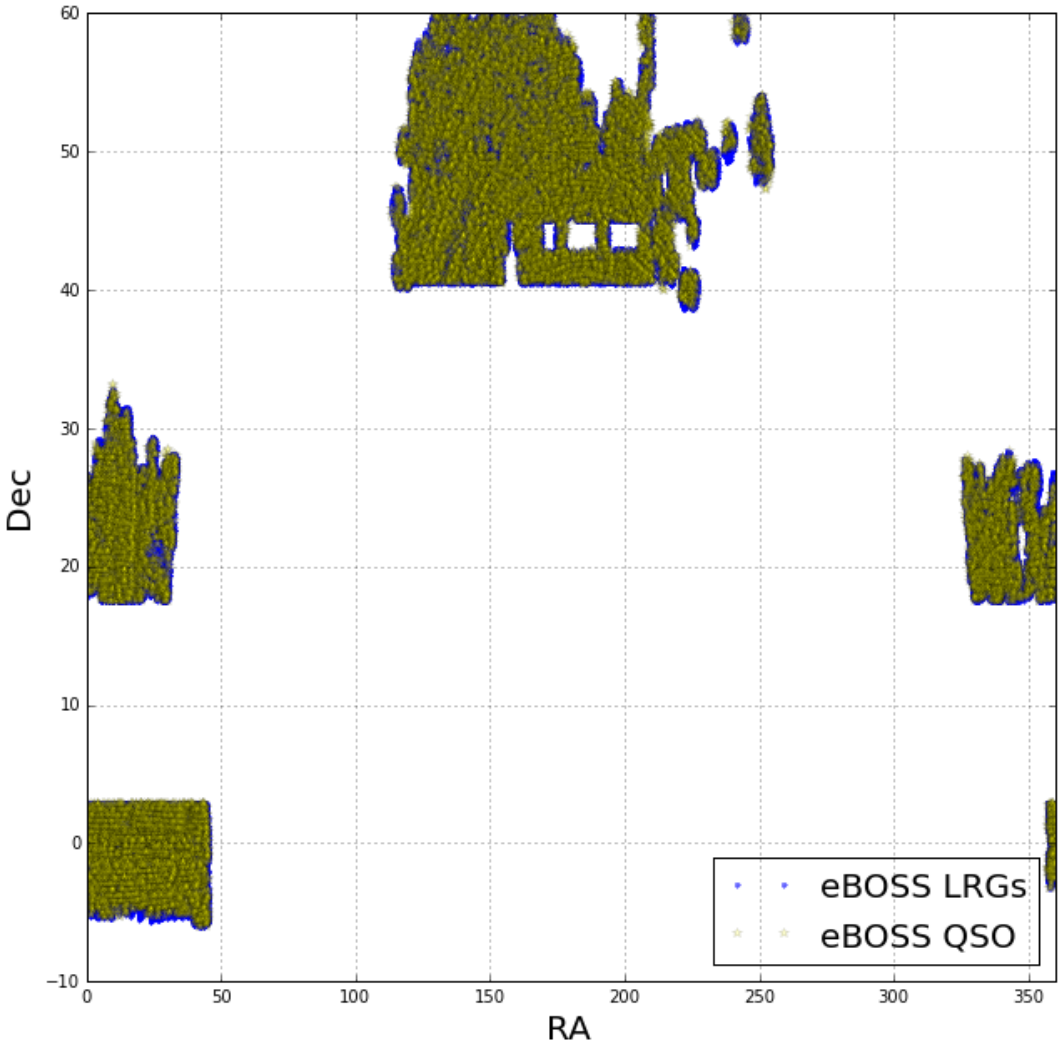


Figure 6.1: The sky coverage of the eBOSS LRG and quasar samples used in this analysis. These galaxies and quasars have spectroscopically confirmed redshifts from the first two years of eBOSS observations. These datasets cover $\sim$2,000deg^{-2} spread over both the NGC and the SGC regions of the SDSS footprint. The LRG and quasar samples completely overlap with each other on the sky since they are observed simultaneously in eBOSS.

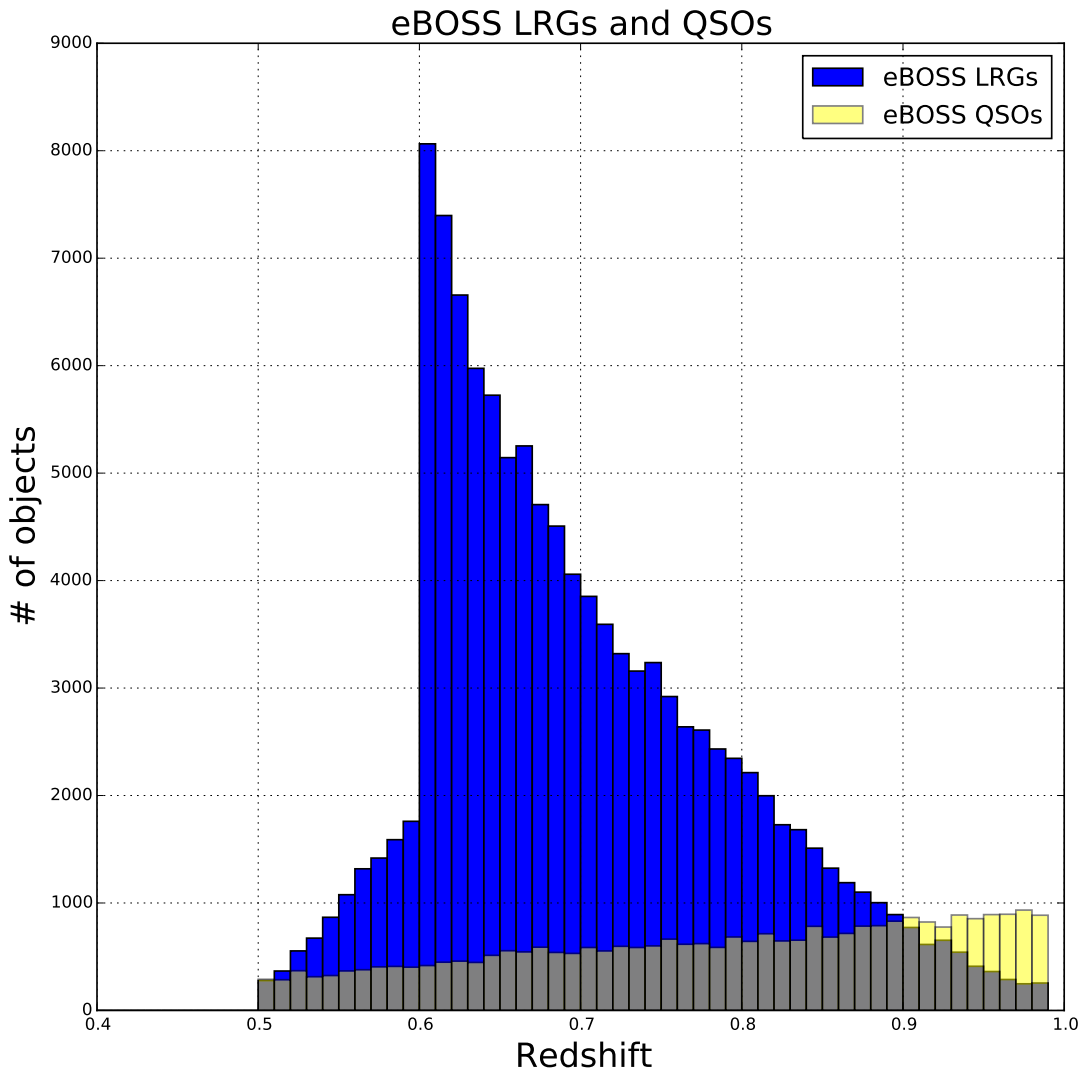


Figure 6.2: The redshift distributions of the eBOSS LRG and quasar samples. These galaxies and quasars have spectroscopically confirmed redshifts from the first two years of eBOSS observations. The quasar sample is primarily designed for higher redshifts and quasars are rare in the low-z universe resulting in a low number density at redshifts $z < 1.0$.

techniques that have been successfully applied to galaxy samples.

6.2.1.1 Footprint

Once the targets are selected using the final target selection algorithms they are fed into a tiling algorithm (Blanton et al., 2003), which assigns spectroscopic fibers to these targets within a 3 degree tile. This allocation is done in a way to maximize the number of fibers assigned to targets keeping in mind the constraints placed by the 62 arcsecond exclusion radius (Dawson et al., 2016). Targets that lie within this exclusion radius can not be assigned different spectroscopic fibers. For this reason, this radius is often referred to as the "fiber collision" radius. The tiling algorithm is sensitive to this fact, and over-dense regions of the sky receive more tiles. Quasars are assigned higher priority than LRGs in the *eBOSS* survey, and a fraction of LRGs suffer fiber collisions. Some of these collisions are resolved by overlapping tiles while others are dealt with separately. We use MANGLE software to divide the sky into unique sectors and within each sector, we compute the following:

- $\mathtt{N}_{star}$: the number of spectroscopically confirmed stars
- $\mathtt{N}_{cp}$: The number of targets that did not receive a spectroscopic fiber due to fiber collisions
- $\mathtt{N}_{Missed}$: The number of targets that will be observed in the future
- $\mathtt{N}_{Legacy}$: The number of targets that already have good spectroscopic redshifts (excluded from tiling)
- $\mathtt{N}_{zFailed}$: The number of targets that did not yield good spectroscopic redshifts
- $\mathtt{N}_{BadClass}$: The number of targets with spectroscopic classification that does not match its target class; for, the quasar sample, these are exclusively galaxies.
- $\mathtt{N}_{Good}$: The number of targets that yield good spectroscopic redshifts.

We then define a targeting completeness parameter per sector and per target class as

$$\mathrm{C}_{eBOSS} = \frac{N_{Good} + N_{zFailed} + N_{BadClass} + N_{cp} + N_{star}}{N_{Good} + N_{zFailed} + N_{BadClass} + N_{cp} + N_{star} + N_{Missed}} \tag{6.1}$$

All sectors with $C_{eBOSS} > 0.5$ are retained in the LSS catalog; the average completeness of the remaining sectors is $\sim$93% in both galactic caps. In addition, we also define a redshift completeness parameter in each sector as:

$$\mathrm{C}_z = \frac{N_{Good}}{N_{Good} + N_{zFailed}}, \tag{6.2}$$

which evaluates target redshift efficiency averaged over each sector. Sectors with $C_z > 0.5$ are retained in the LSS catalogs.

6.2.1.2 Veto Mask

We apply the same veto mask as was applied on *BOSS* DR12 regions to exclude sectors in problematic areas (Reid et al., 2016). These include:

- `Bright stars mask`: Regions around bright stars based on the Tycho catalog (Høg et al., 2000).
- `Bright object mask`: Regions around bright galaxies and stars not included in the Tycho catalog (Rykoff et al., 2014).
- `Bad photometry mask`: Regions strongly affected by dust extinction and bad seeing.
- `Center posts`: Center posts anchor the spectrographic plates and prevent any fibers from being placed there.

6.2.1.3 Spectroscopic completion and systematic weights

Spectroscopic completeness is affected by a combination of factors mentioned in Section 6.2.1.1. For example, targets missed due to fiber collisions are preferentially in denser regions of the sky. They do not happen randomly on the sky and have a higher bias that should be taken into account. These fiber collisions are corrected by transferring the weight of the lost target to the nearest neighbor of the same target class with a valid redshift and spectroscopic classification. The spectroscopic redshift completeness weights are identified in the LSS catalogs as w_c.

Similar to spectroscopic weights, we also assign a systematics weight. This is estimated based on the relation between the target number density and the sources of potential systematics as explained in the Section 3.7.1. The systematic weights are identified in the LSS catalogs as w_{sys}. In the end, a total weight, $w_{tot} = w_c \star w_{sys}$, is assigned to each target which is used in clustering measurements.

6.2.2 Random catalogs

Random catalogs have been constructed in a way that matches the angular and redshift distribution of the data, but with approximately 40 times the number density. These catalogs are necessary to compute the correlation functions of the quasars and LRGs. In the simplest approach, we began by generating a set of points randomly distributed over the eBOSS footprint where the angular number density in each sector is subsampled to match the value of C_{eBOSS} in that sector. This is implemented using the MANGLE software. Next, we apply the same *veto* masks, that are applied on the LRG and quasar datasets, on these random points. Each random point is then assigned a redshift which is drawn randomly from the datasets. Finally, each random redshift drawn from the LRG and quasar datasets is weighted by the total LRG and quasar weight, which includes spectroscopic completeness weight (w_c) and systematic weight (w_{sys}), $w_{tot} = w_c \star\ w_{sys}$. Random samples generated in this way represent the true distribution of the LRG and quasar sample in the absence of any cosmological clustering.

6.3 CROSS-CORRELATION STATISTICS

Cross-correlation functions are defined similarly to the two-point autocorrelation function. The two-point autocorrelation function (typically referred to as the correlation function) is the excess probability of finding a galaxy in a volume element, dV, at a separation, r, from another randomly chosen galaxy:

$$dP = n[1 + \xi(r)]dV, \tag{6.3}$$

where n is the mean number density of the galaxies in question (Peebles, 1980). The definition implies that the strength of correlation would be zero for uniformly distributed objects. In a way, the two-point correlation functions provide a measure of lumpiness of the distribution of a sample. Analogously to the auto-correlation function, the two-point cross-correlation

function is the excess probability of finding an object from a given sample in a volume element, dV, at a separation, r, from a random object in another sample. In this work, we measure the cross-correlation of the eBOSS LRGs with eBOSS quasars:

$$dP(LRG|Q) = n_{LRG}[1 + \xi_{Q,LRG}(r)]dV, \tag{6.4}$$

which is the excess probability of finding an LRG in a volume element dV at a separation r from a quasar (Q), where n_{LRG} is the number density of LRGs.

Here we use the Landy & Szalay estimator (LS, Landy & Szalay, 1993) to measure the cross-correlation function between the quasar and LRG samples. The LS estimator has been found to be better at handling edge effects and minimizes variance compared to more naive estimators (Kerscher et al., 2000). It is defined by:

$$\xi_{LS} = \frac{(DD - 2DR - RR)}{RR}, \tag{6.5}$$

where DD are counts of data pairs, DR are data-random pairs, and RR are random-random pairs at a given separation. This form of the LS estimator is suited for two-point autocorrelations, which incorporates only one dataset. For cross-correlations, we have two different datasets with separate randoms corresponding to each of them. Hence the LS estimator needs to be generalized in order to compute cross correlations. Specifically, we measure the observed number of galaxies around each quasar as a function of distance as well as the observed number of galaxy and quasar pairs for random distributions. This information is then used to calculate the generalized LS estimator which is computed as:

$$\xi_{LS}(r) = \frac{(GQ - GR_Q - QR_G - R_GR_Q)}{R_GRQ}, \tag{6.6}$$

where GQ are LRG-quasar pairs, R_Q and R_G are the randoms corresponding to the quasar and the LRG samples, respectively, R_GR_Q are random-random pair counts from the randoms corresponding to the LRG and quasar samples, and GR_Q are the LRG-quasar random pair counts, while QR_G are the quasar-LRG random pair counts.

We then compute the cross-correlation function, $\xi_{Q,LRG}(r)$, in real space using comoving distances in redshift bins of $\Delta z = 0.05$. $\xi(r)$ can then be modeled as a power law, $\xi(r) = (\frac{r}{r_0})^\gamma$. A power-law fit to the computed cross-correlation function can be used to recover the values of r_0 and γ.

6.4 RESULTS AND DISCUSSION OF QUASAR-LRG CLUSTERING

We show the results of LRG-quasar cross-correlations in Figures 6.4, 6.5, 6.6, and 6.7. The blue line is the observed cross-correlation function between the eBOSS quasar sample and LRGs. The solid red line is a power-law fit to the observed cross-correlation function. The error bars are estimated from jackknife resampling of the LRGs and quasar samples.

For Jackknife resampling, we start with dividing the sky into regions of equal area (pixels). This is a straightforward exercise using the *Healpix* pixelization scheme (Górski et al., 2005). The LRGs and quasars are assigned to these pixels based on their position on the sky. The current spectroscopic datasets spans almost 88 of these equal area pixels. Next, we combine four contiguous nearby pixels to yield 22 non-overlapping regions of equal area on the sky. Once the datasets are divided into smaller independent subsets, we measure the clustering in each of these regions. The error bars are the standard errors $\frac{\sigma}{\sqrt{(n-1)}}$ on $\xi(r)$, where σ is the standard deviation.

We present the values for the parameters, r_0 and γ, in Figures 6.8. We obtain a mean value for $r_0 \sim 4.75 \pm 0.15 h^{-1}$Mpc for the quasar-LRG cross-correlation with $\gamma \sim$1.5. The best fits are derived using the least square error method over the jackknife samples. We present the redshift dependence of these parameters in Figure 6.9. Although, it seems that r_0 evolves with the redshift, the evidence is not strong enough. The slope of clustering, γ, is consistently around 1.5. These values are within reasonable deviations with the results reported in literature, e.g., Porciani et al. (2004) report $r_0 \sim 4.7$ h^{-1}Mpc and a slope of $\gamma \sim$1.8 for the auto-correlations of quasars.

Using the overlapping spectroscopic LRG sample, we compute the auto-correlations for LRGs in each redshift bin. These auto-correlations are modeled as power laws and we estimate the parameters r_0 and γ in a similar fashion to the cross-correlations. We show the results of LRG auto-correlations in Figures 6.10, 6.11, 6.12, and 6.13. We present the values for the parameters, r_0 and γ, in Figures 6.14. We also present the redshift dependence of the power-law parameters in Figure 6.15. The values of the parameters, $r_0 \sim 6.8 \pm 0.20 h^{-1}$Mpc and $\gamma \sim$1.8, are also consistent with values in the literature. This enables us to compute the relative bias of quasars with respect to LRGs, $b_{q,lrg} = \frac{\xi_{q,lrg}(r)}{\xi_{lrg}(r)}$, assuming a linear bias.

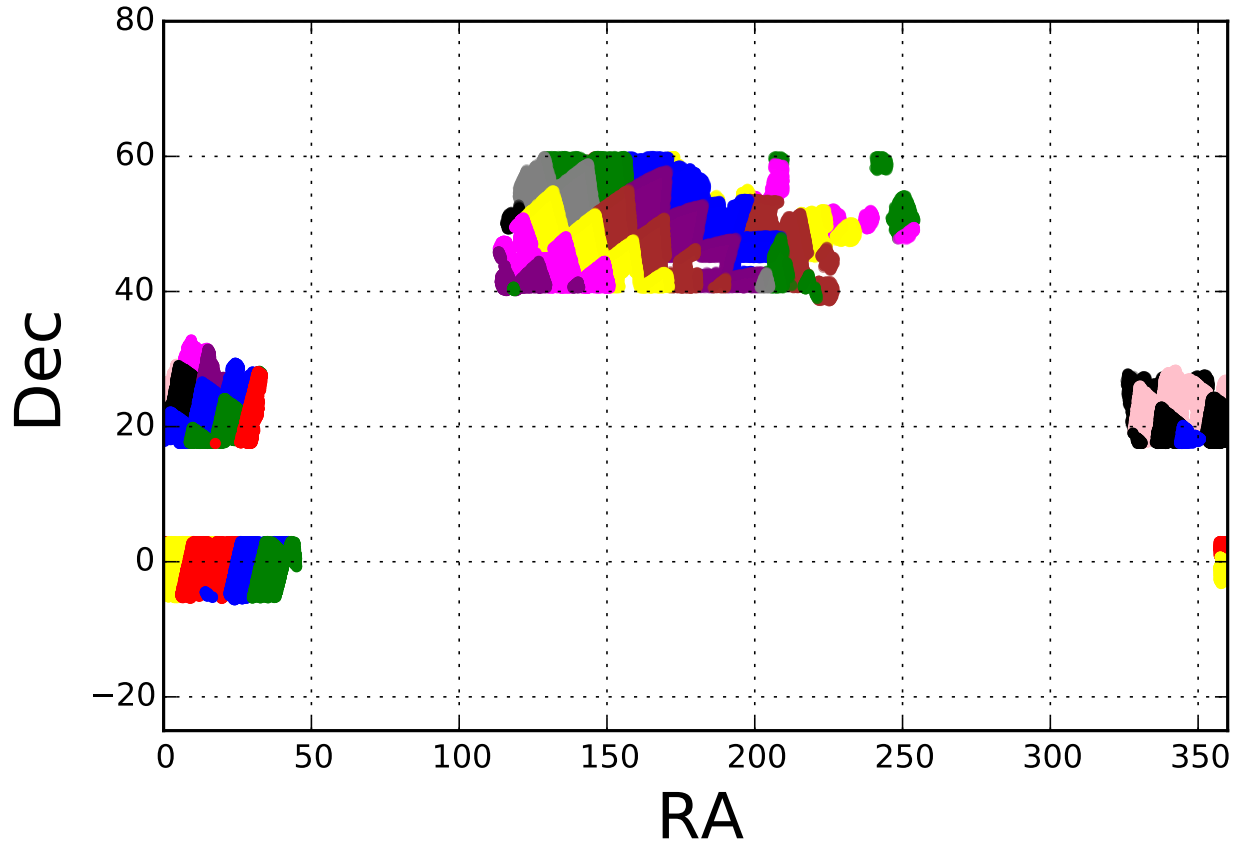
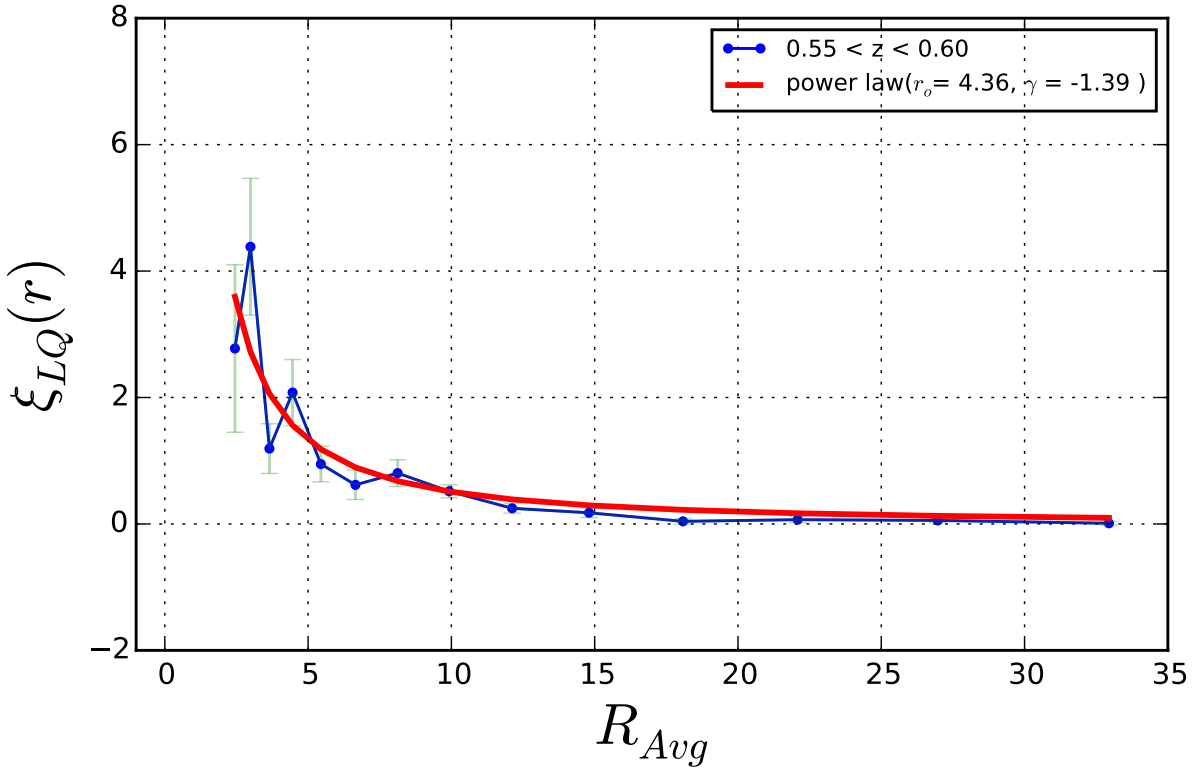


Figure 6.3: The *jackknife* resampling the eBOSS LRG and QSO samples. The datasets are split into 22 non-overlapping equal-area regions on the sky using *Healpix*. Clustering is measured in each region separately to estimate the uncertainty and robustness of r_0 and γ.



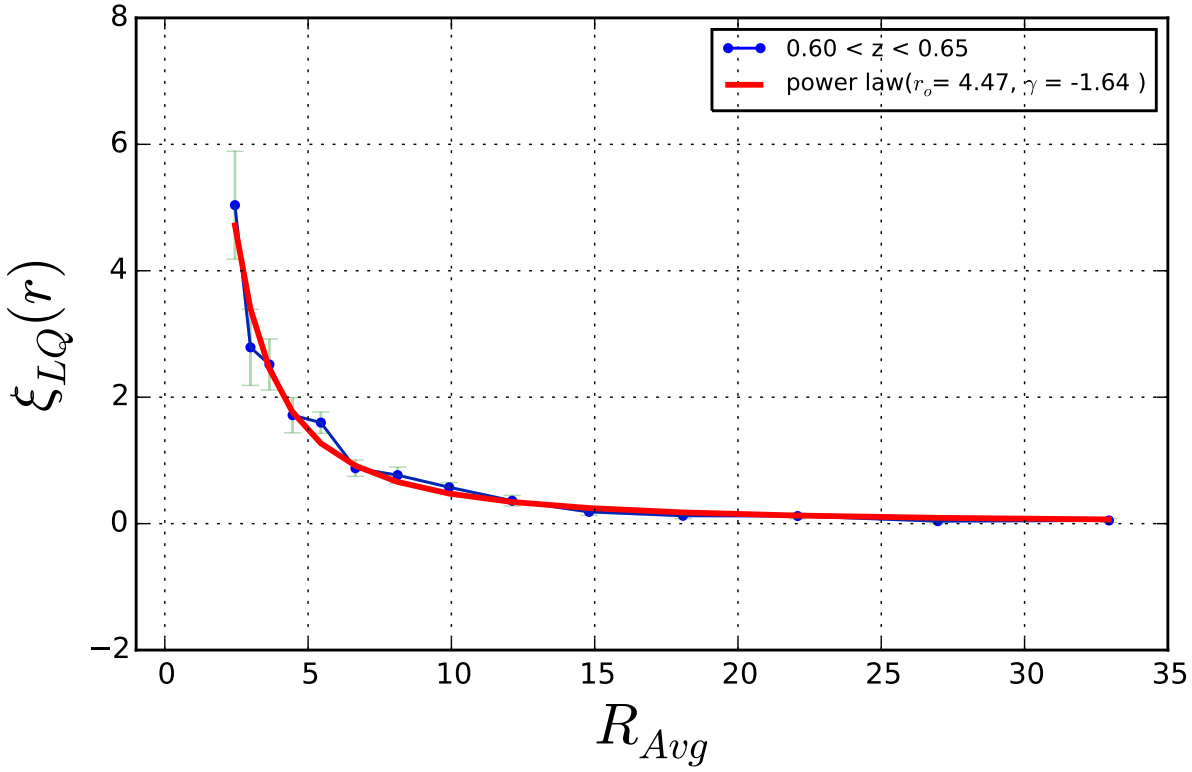
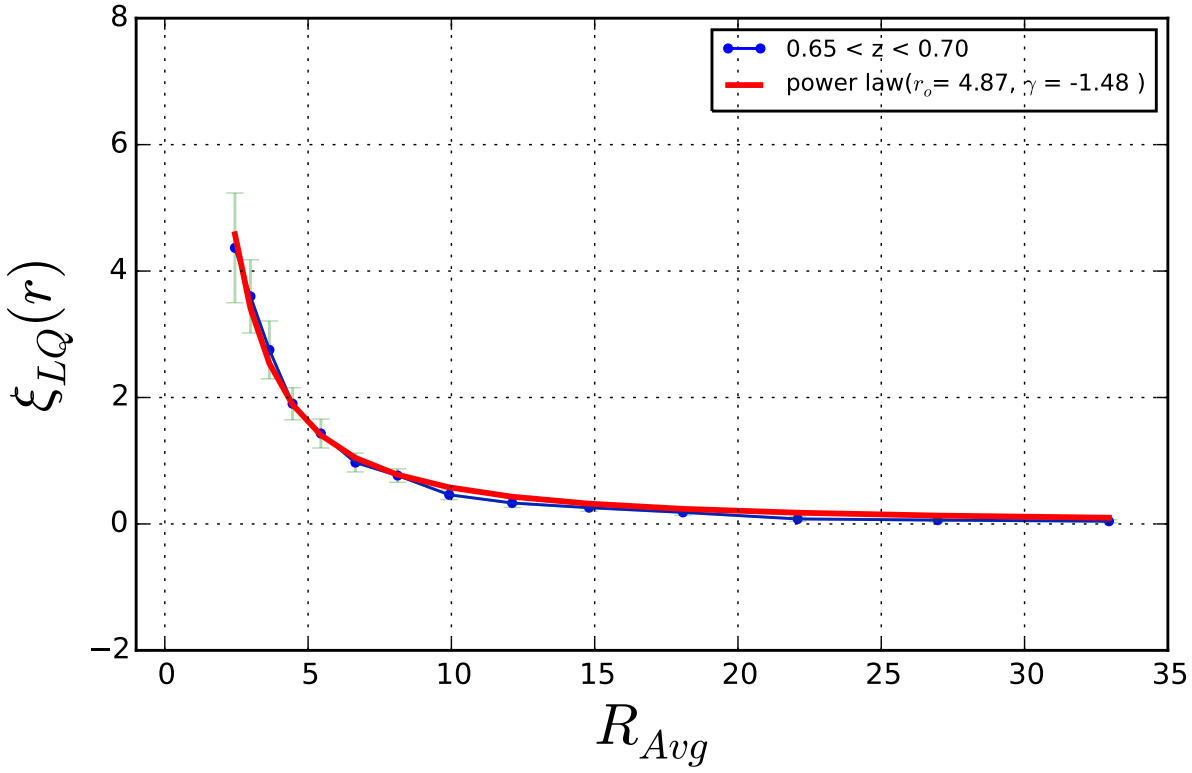


Figure 6.4: The clustering of the eBOSS LRG and QSO samples at redshifts $0.55 < z < 0.60$ and $0.60 < z < 0.65$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.



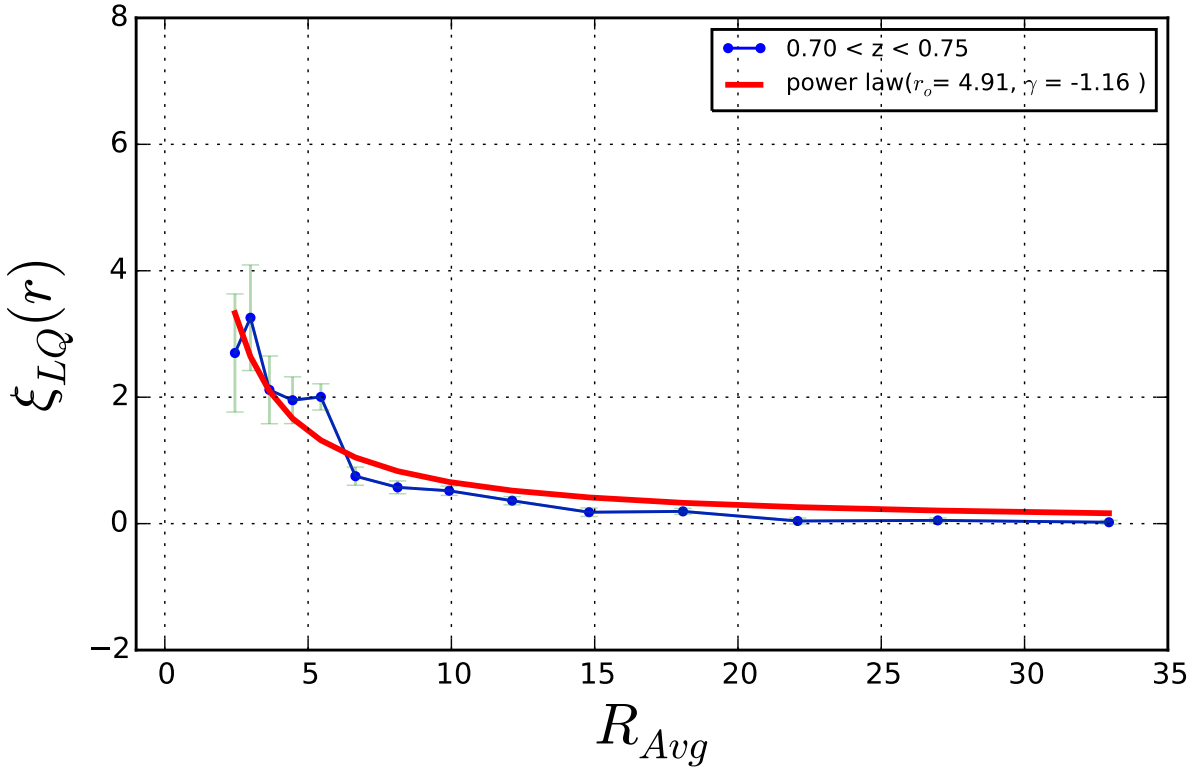
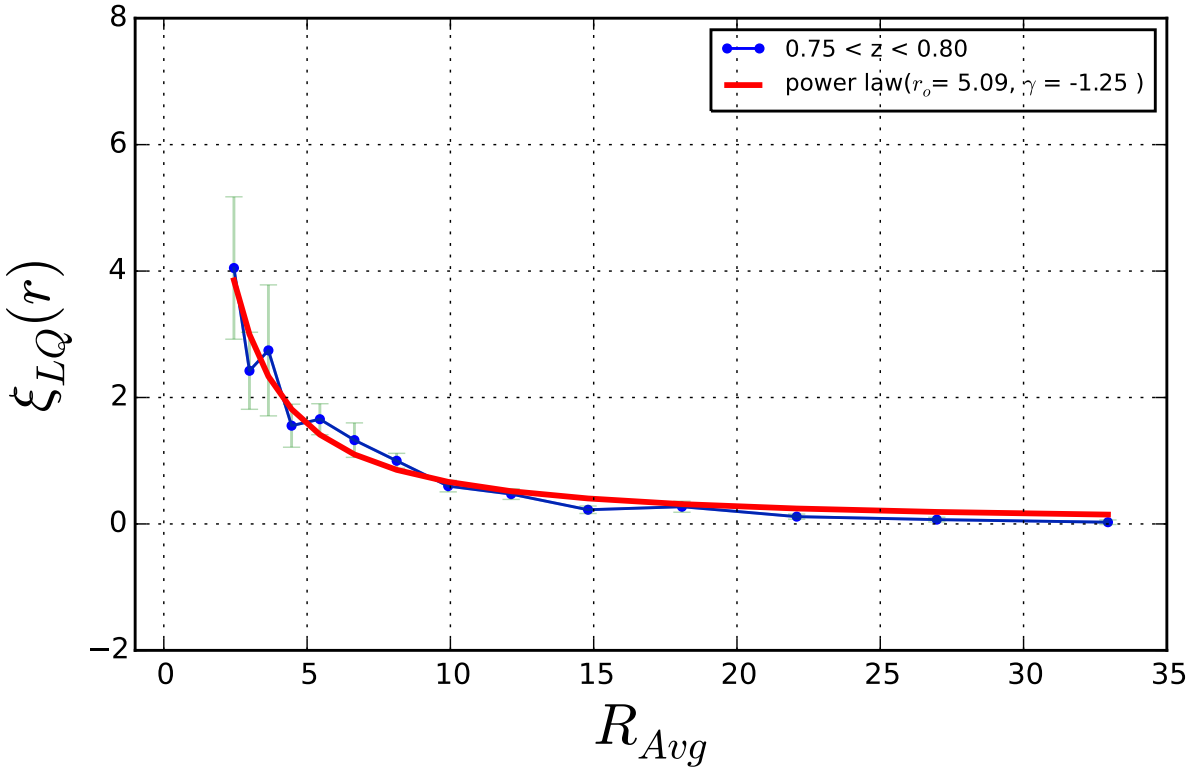


Figure 6.5: The clustering of the eBOSS LRG and QSO samples at redshifts $0.65 < z < 0.70$ and $0.70 < z < 0.75$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.



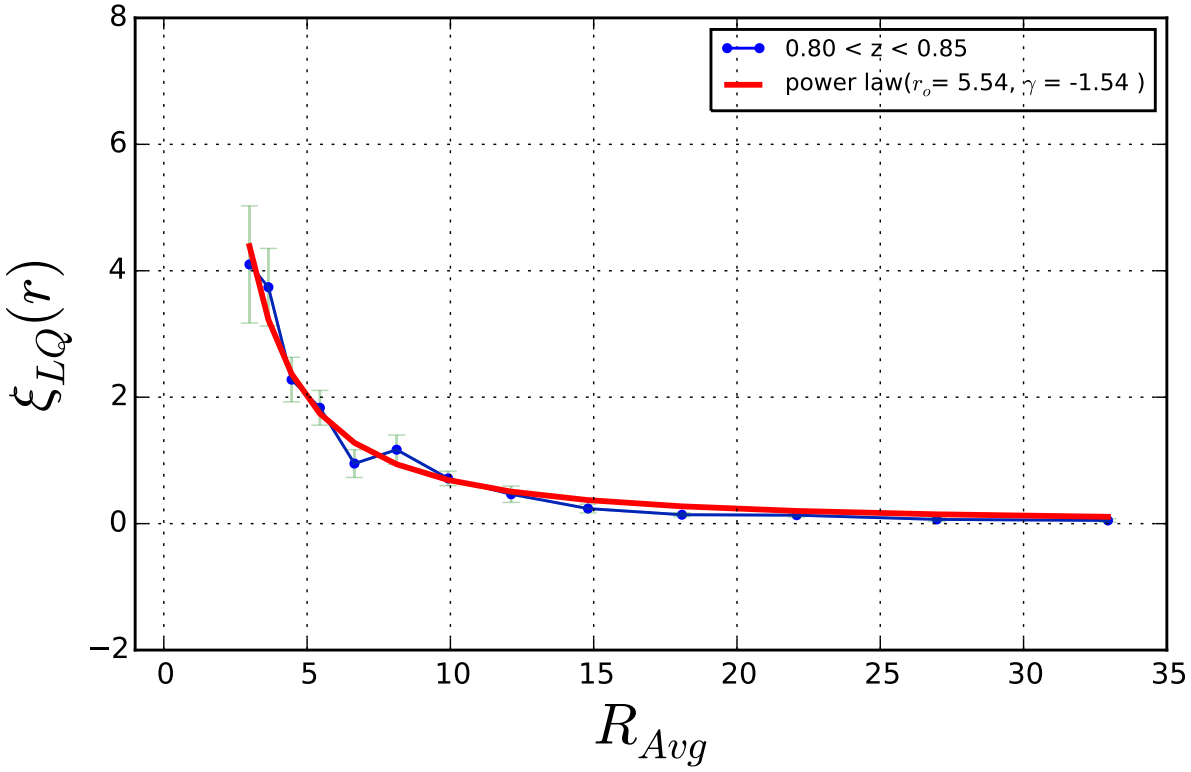
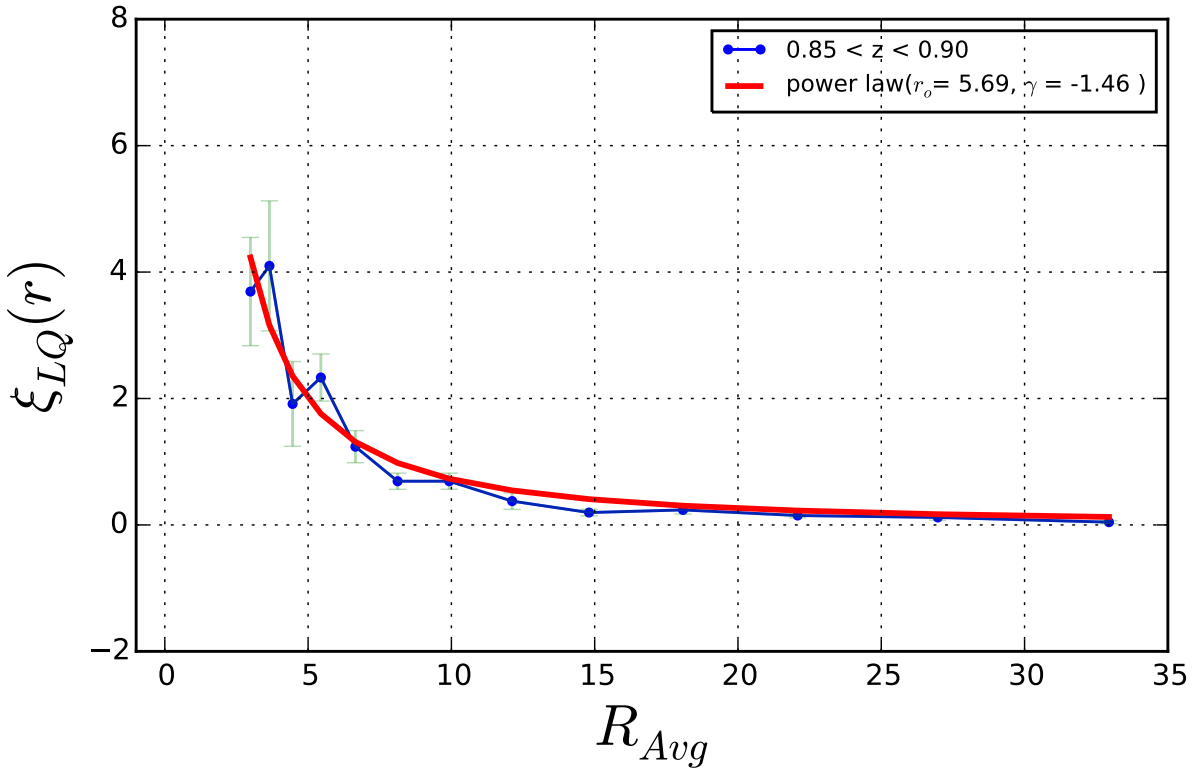


Figure 6.6: The clustering of the eBOSS LRG and QSO samples at redshifts $0.75 < z < 0.80$ and $0.80 < z < 0.85$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.



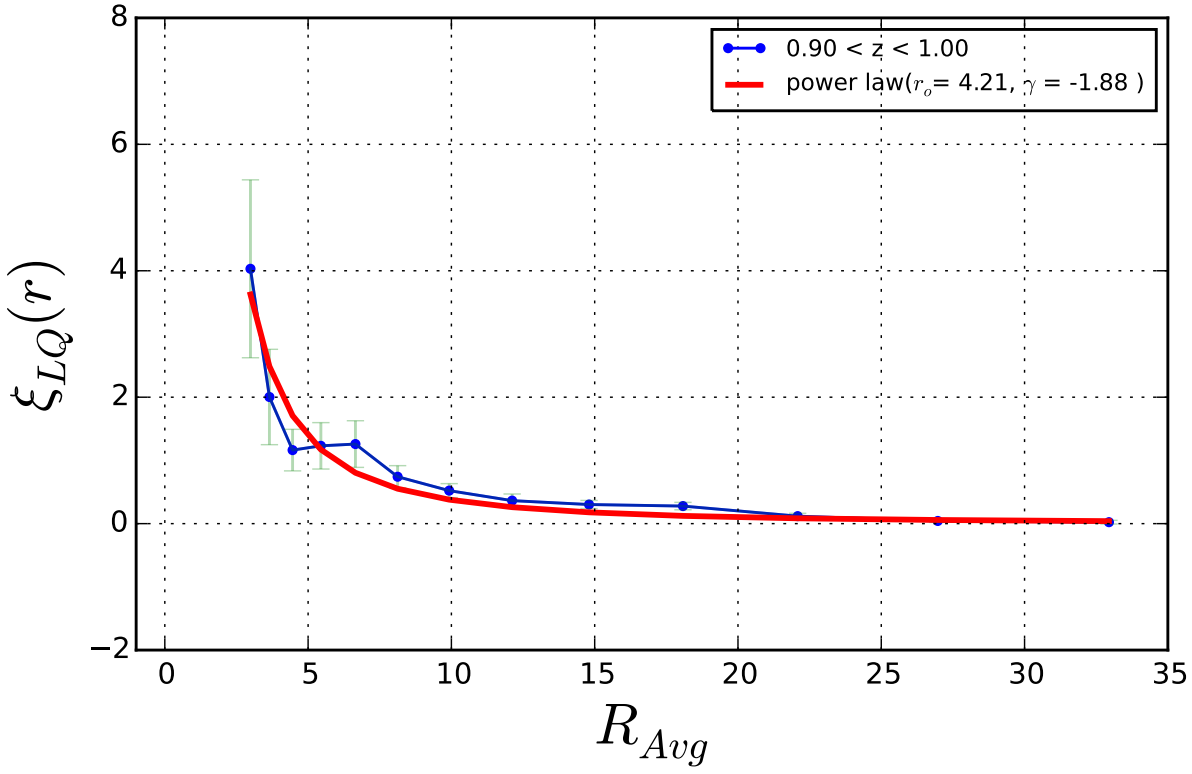


Figure 6.7: The clustering of the eBOSS LRG and QSO samples at redshifts $0.85 < z < 0.90$ and $0.90 < z < 1.0$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.

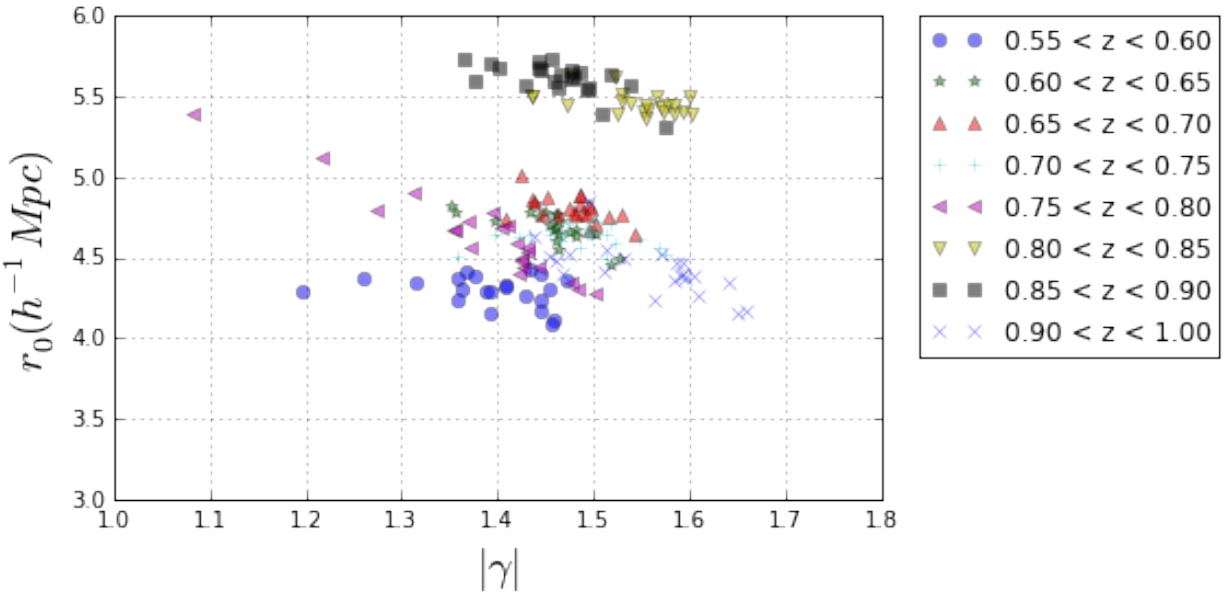


Figure 6.8: The parameters r_0 vs. γ, for the cross-correlation function of the eBOSS LRG and QSO samples. These parameters are optimized using least square errors and *jackknife* resampling.

We present the relative bias of quasars at $r_{avg} \sim 5h^{-1}$Mpc and $r_{avg} \sim 10h^{-1}$Mpc in different redshift bins in Figure 6.16. Our results imply that quasars live in halos which are less massive than the dark matter halos in which LRGs reside. This seems consistent with the fact that quasar cluster similar to the normal blue galaxies.

LRGs and quasars are observed simultaneously in eBOSS. Since quasars are accorded higher priority within eBOSS, LRGs that are very close to quasars suffer from fiber collisions and cannot be simultaneously observed. Fiber collisions result in under-sampling of the regions of the sky with a higher density of targets. This under-sampling leads to underestimating the correlation function on small scales. More work is needed to understand the effects of fiber collisions on galaxy-quasar clustering. For the reasons of fiber collisions and the fact that a power-law does not necessarily represent the clustering at very small scales, the regions within $r < 2.5$ are avoided in our modeling.

Once the impact of fiber collisions is accounted for, the behavior of the cross-correlation function at small scales can be explained better. These results will be more robust when we have a large spectroscopic sample for analysis. This will be possible in the future.

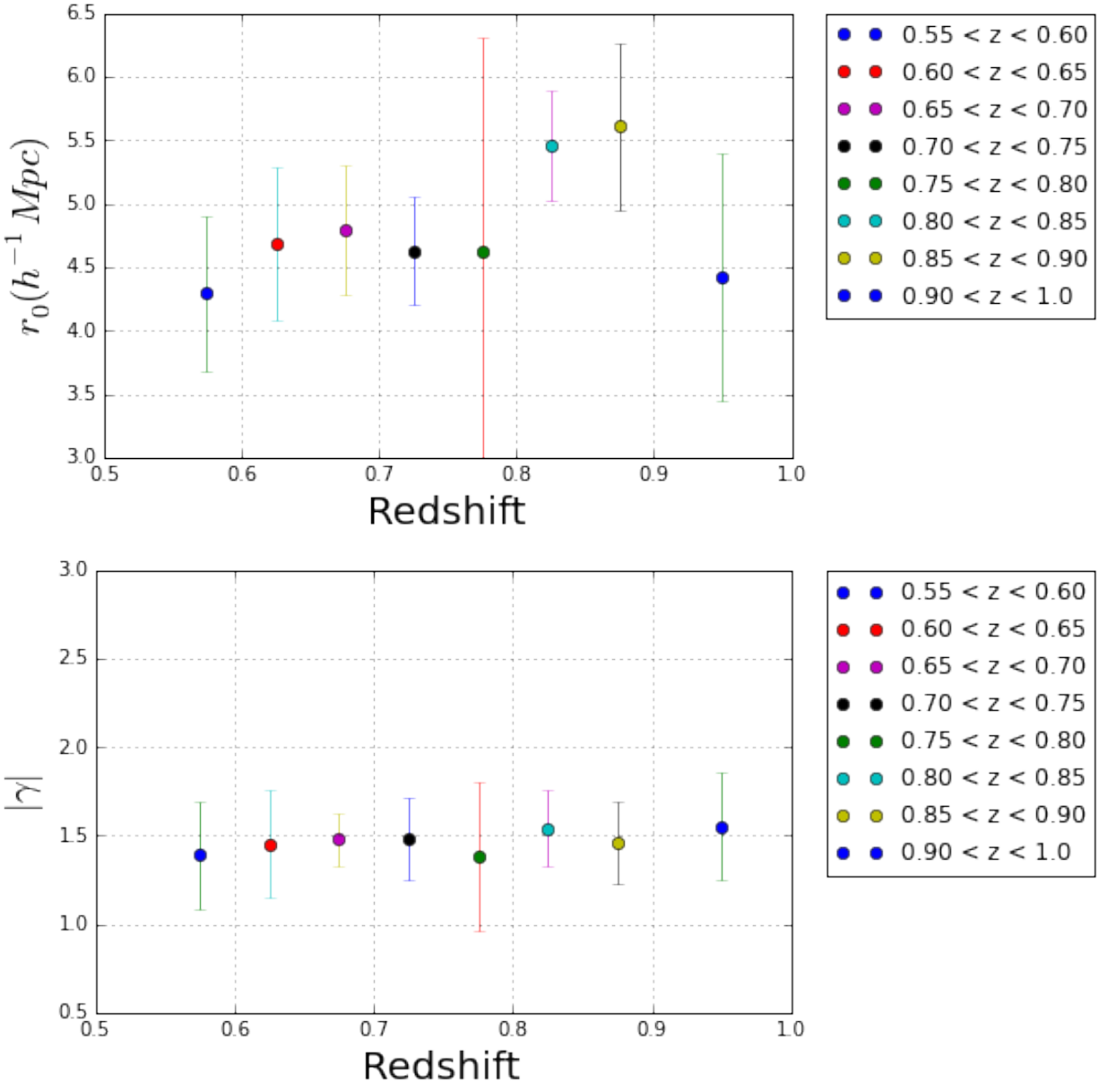
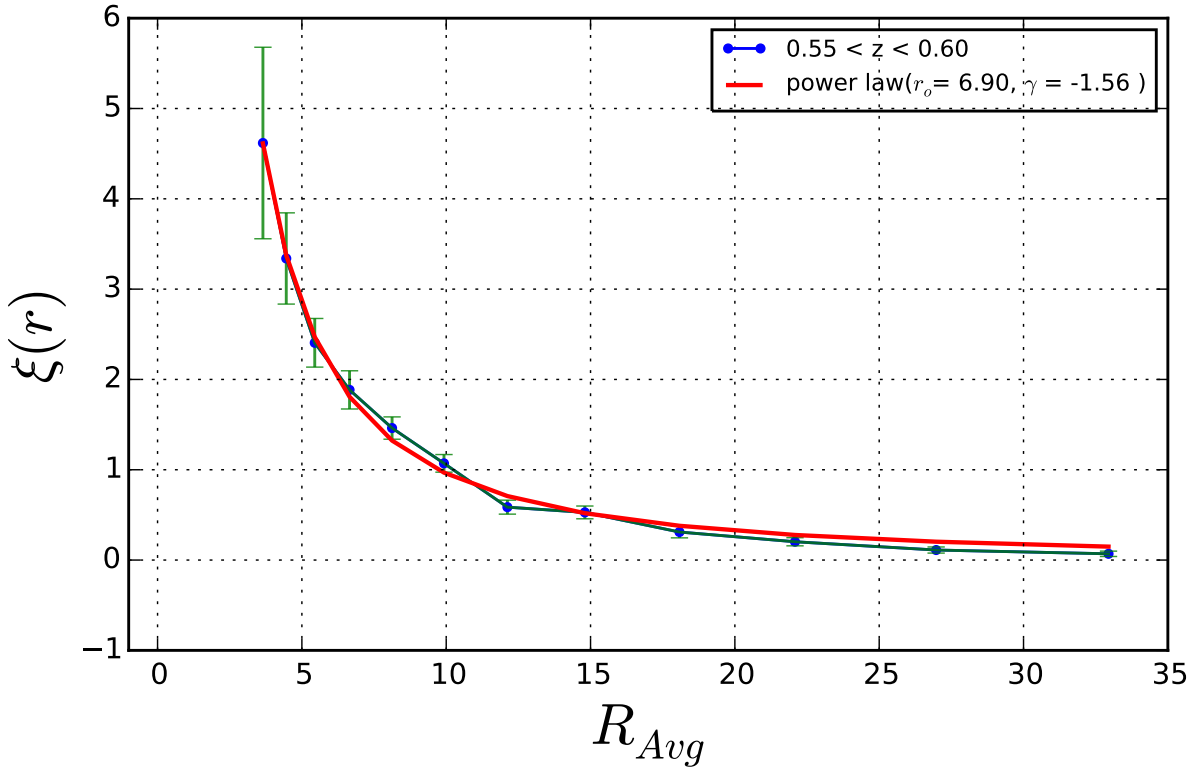


Figure 6.9: The parameters, r_0 and γ, as a function of redshift for the cross-correlation function of the eBOSS LRG and QSO samples. These parameters are optimized using least square errors and *jackknife* resampling.



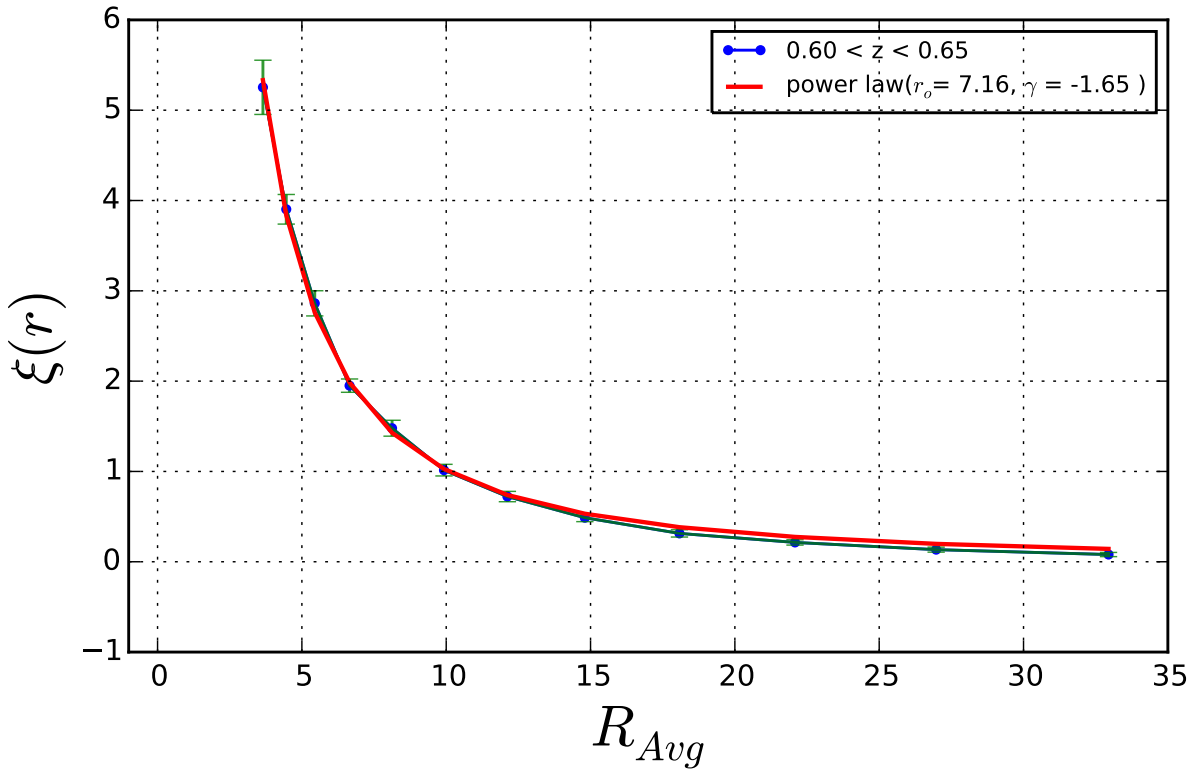
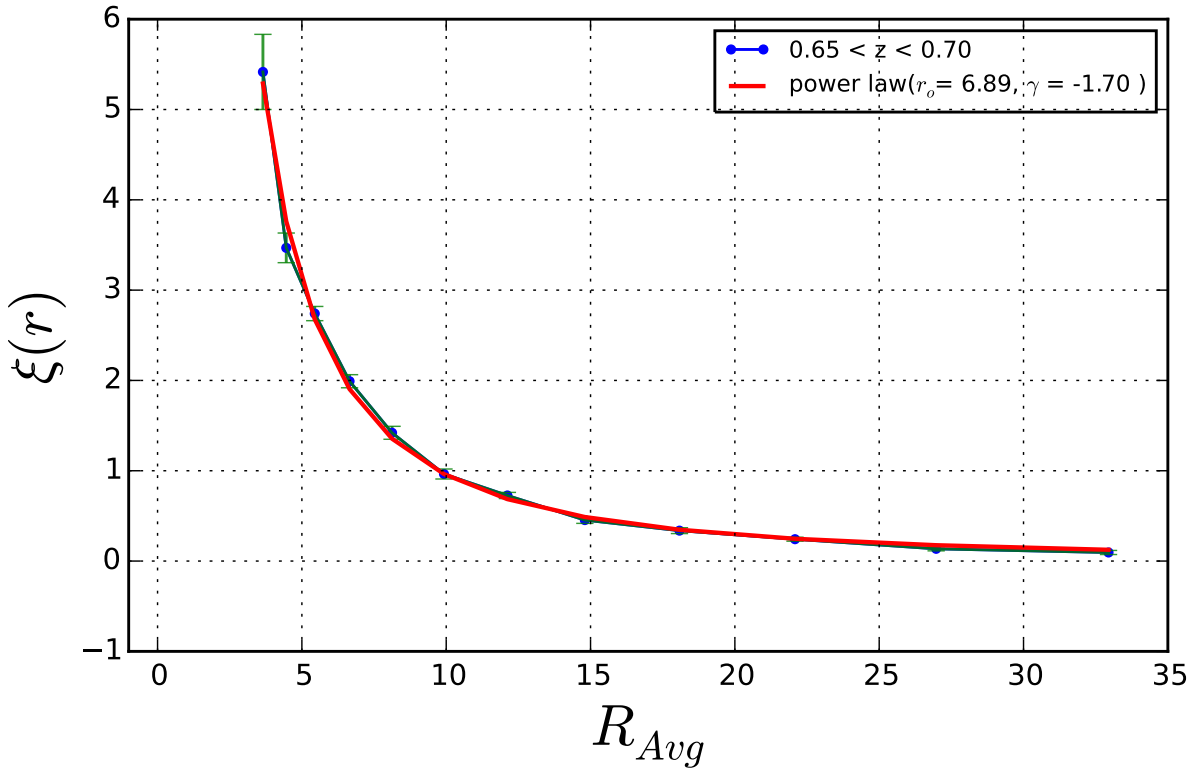


Figure 6.10: The clustering of the eBOSS LRG sample at redshifts $0.55 < z < 0.60$ and $0.60 < z < 0.65$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.



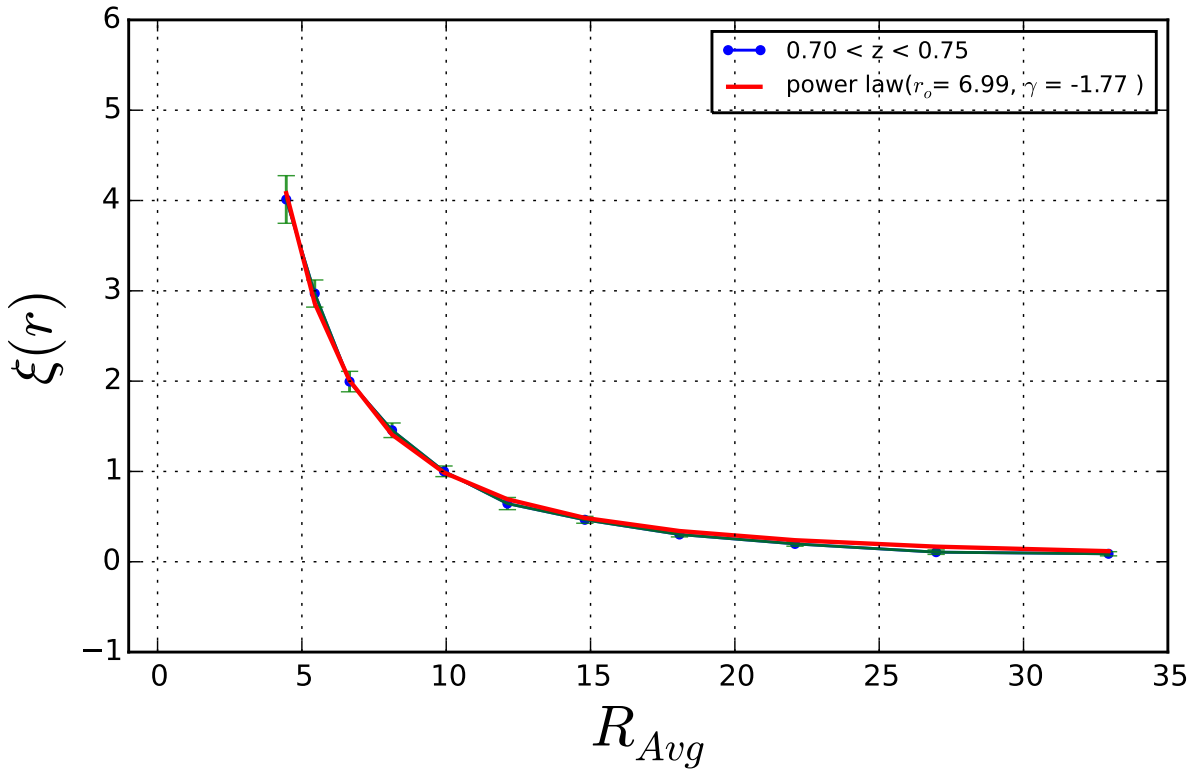
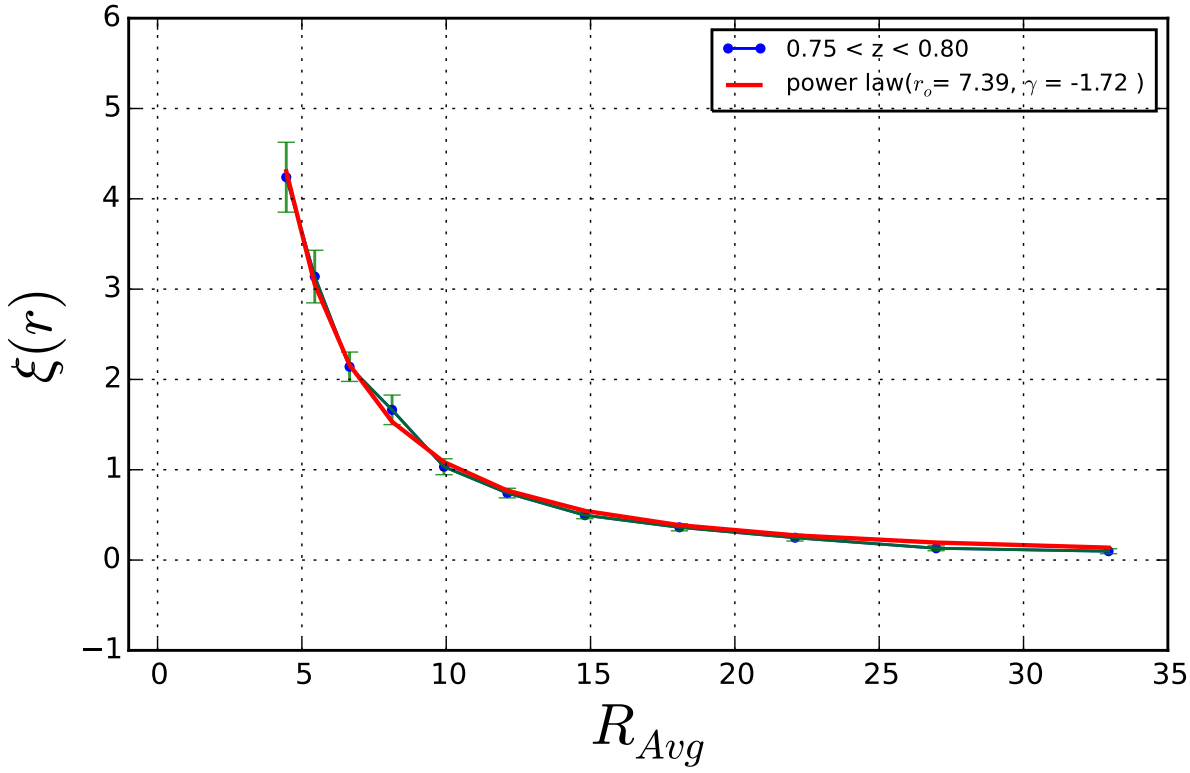


Figure 6.11: The clustering of the eBOSS LRG sample at redshifts $0.65 < z < 0.70$ and $0.70 < z < 0.75$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.



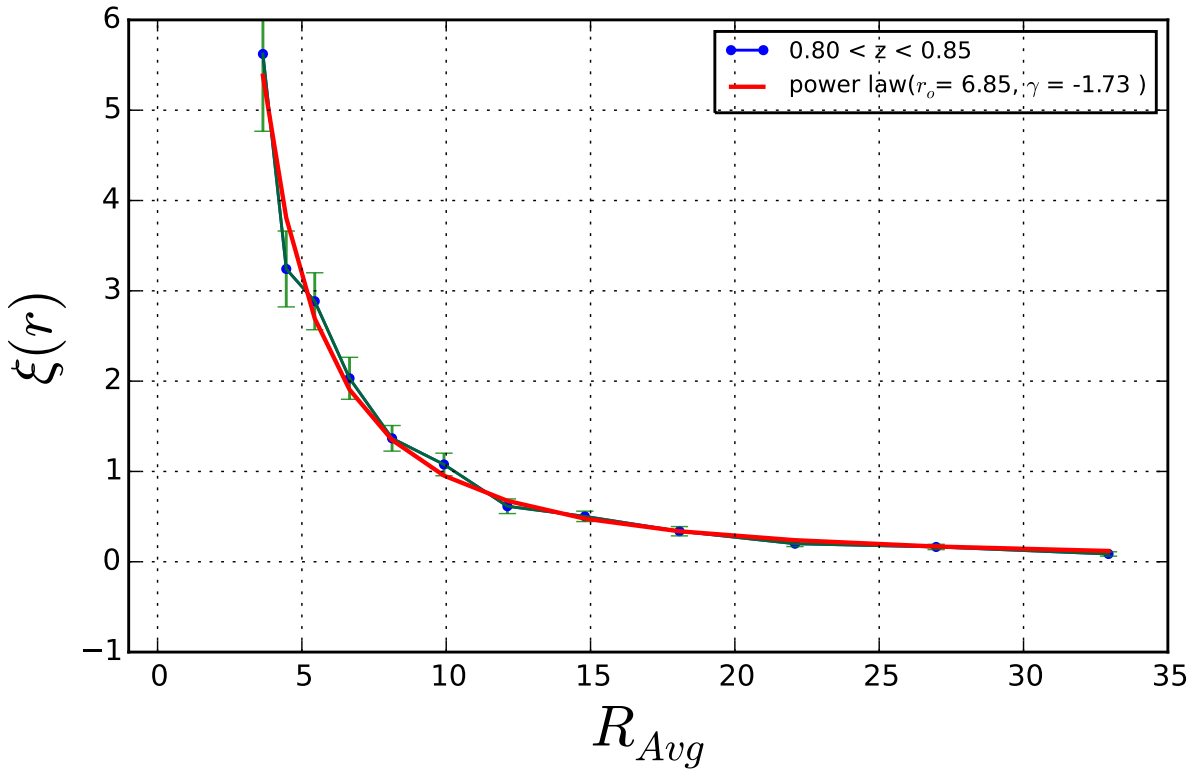
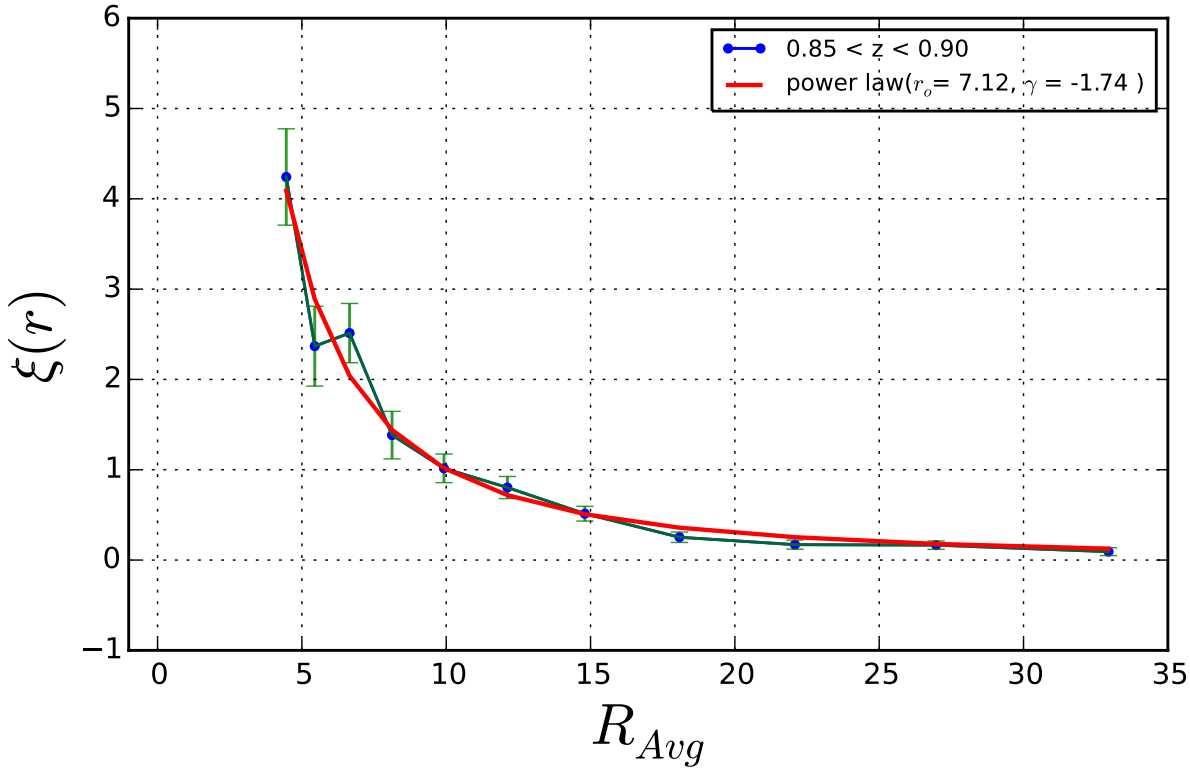


Figure 6.12: The clustering of the eBOSS LRG sample at redshifts $0.75 < z < 0.80$ and $0.80 < z < 0.85$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.



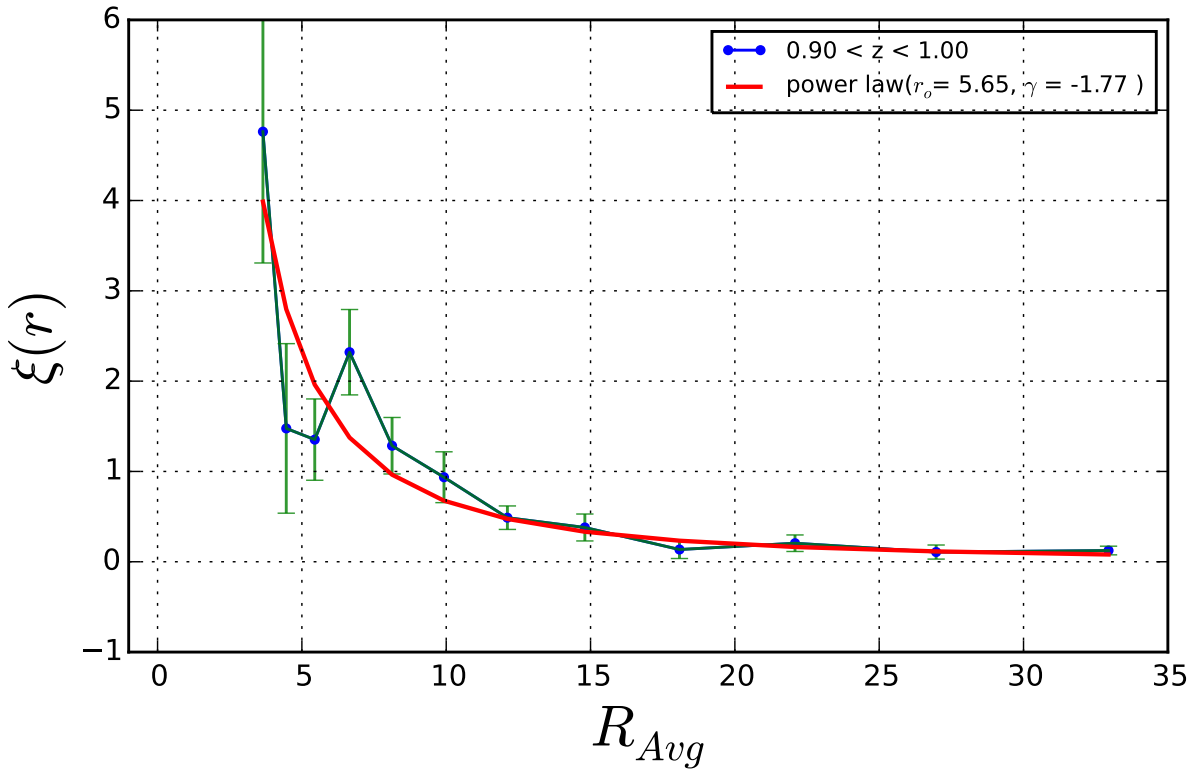


Figure 6.13: The clustering of the eBOSS LRG sample at redshifts $0.85 < z < 0.90$ and $0.90 < z < 1.0$. The error bars are the standard errors on the clustering amplitudes derived from *jackknife* resampling.

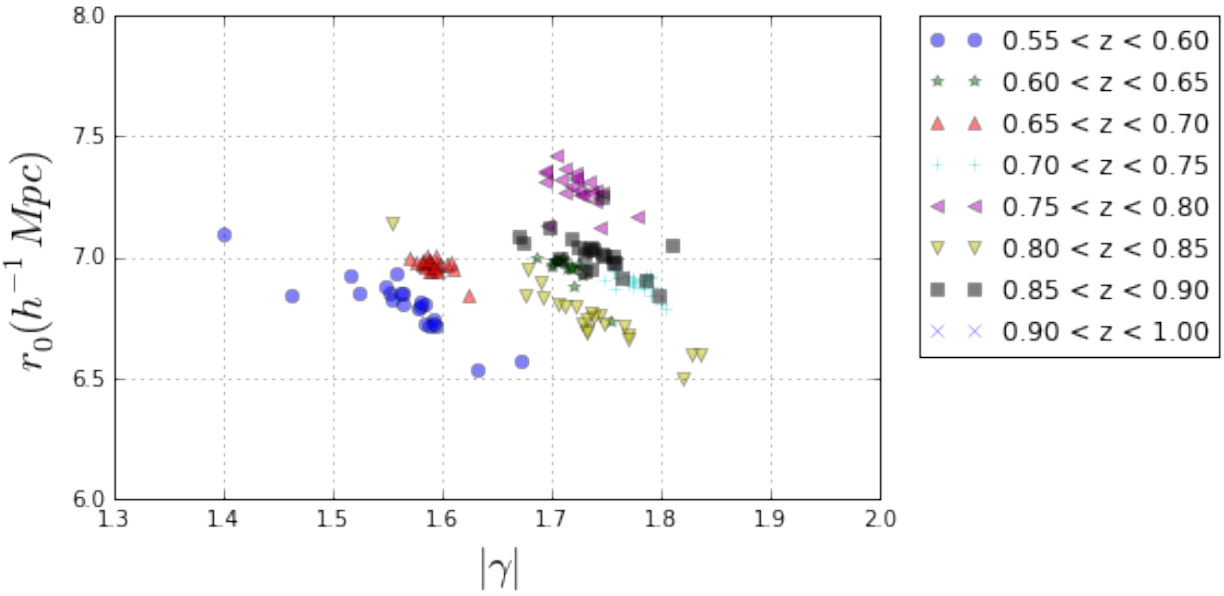


Figure 6.14: The parameters r_0 vs γ for the auto-correlation function of the eBOSS LRG sample. These parameters are optimized using least square errors and Jackknife resampling.

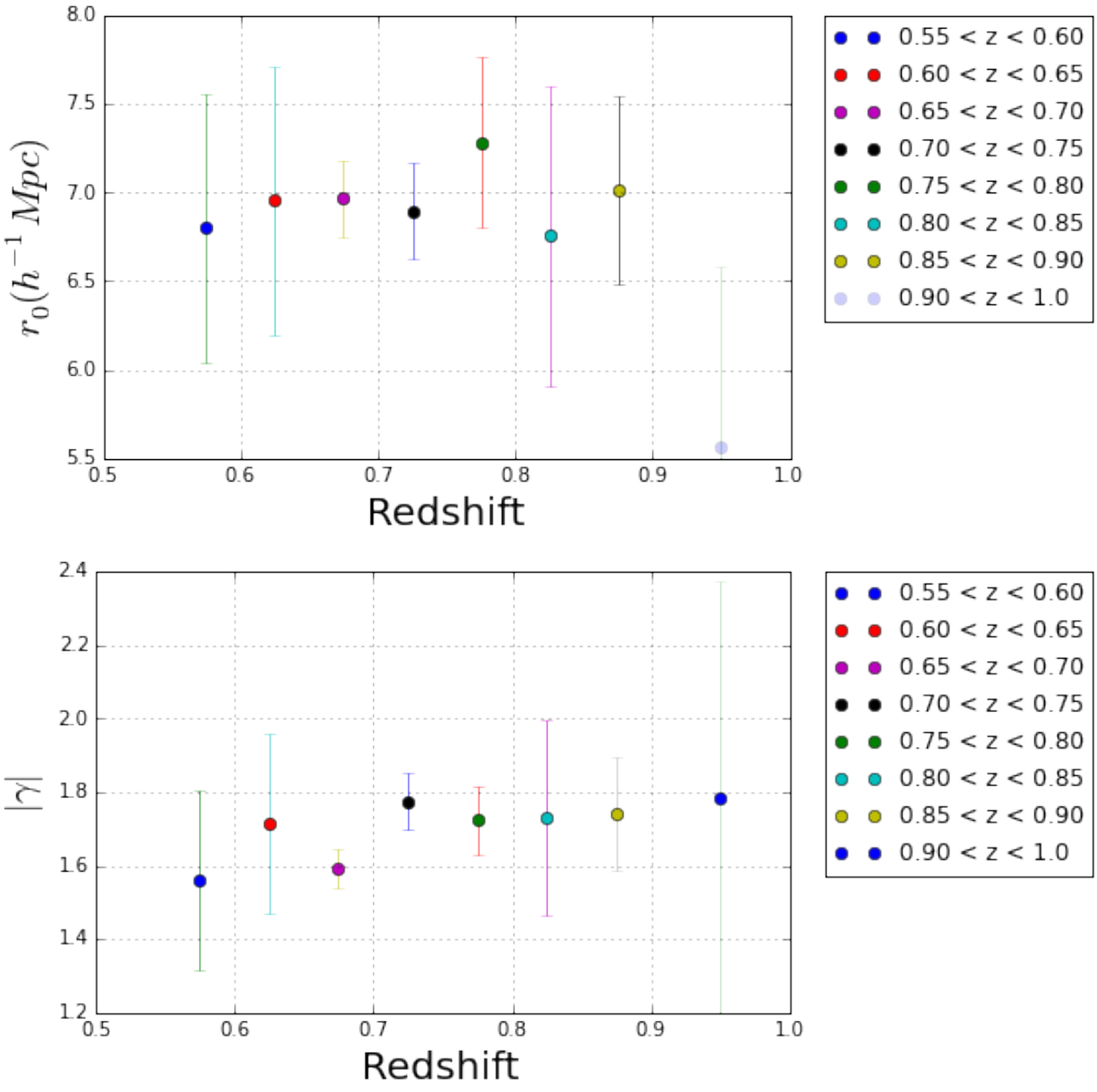


Figure 6.15: The parameters, r_0 and γ, as a function of redshift for the auto-correlation function of the eBOSS LRG sample. These parameters are optimized using least square errors and Jackknife resampling.

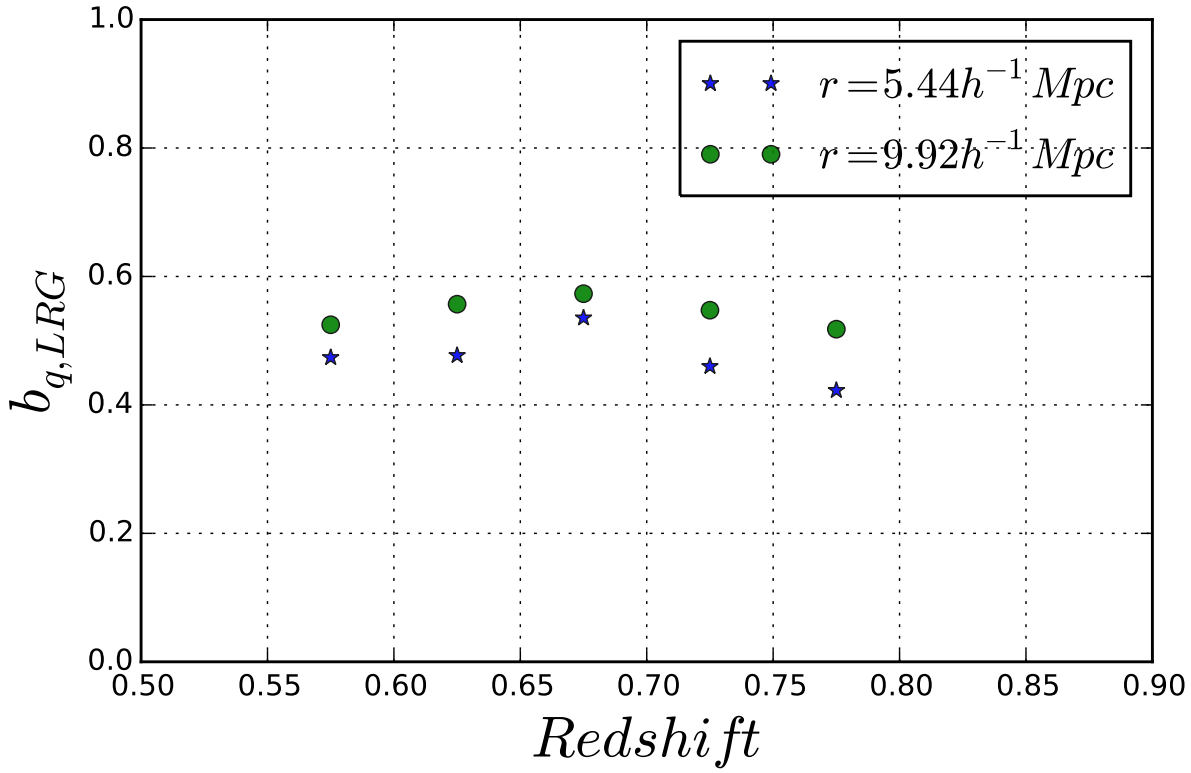


Figure 6.16: The relative bias, $b_{q,lrg}$, of quasars with respect to LRGs as a function of redshift for the eBOSS quasar sample. This parameter is estimated from the auto-correlation and cross-correlation of LRGs and quasars.

7.0 DISSERTATION CONCLUSION

In this dissertation, we have focused on applying new and innovative techniques to assemble the next generation of galaxy and quasar catalogs for spectroscopic surveys such as SDSS-III/IV and DESI, which aim to precisely measure the BAO scales. These galaxy catalogs have undergone (or are undergoing) successful spectroscopic follow-up observations. We have employed modern statistical tools to understand the systematic errors associated with the SDSS and WISE imaging data which were used to assemble these samples. These systematic errors can limit our ability to measure clustering on large scales. Overcoming these challenges is crucial for extending our knowledge of the universe through these samples.

The methods developed in this thesis have been adopted by other international teams, such as the 2-degree Field Lensing Survey (2dFLenS), to assemble and analyze different galaxy samples. We have developed a machine learning algorithm for estimating redshifts of galaxies using their brightness and color. This is an important step for data-intensive astronomy as the next generation of photometric surveys such as LSST, WFIRST, and TMT will produce an unprecedented volume of data. Although photometric redshifts are less accurate than spectroscopic redshifts, they are invaluable in cases where obtaining spectroscopy is either extremely difficult or impractical.

The SDSS-IV/eBOSS quasar sample provides another unique tracer for studying the large-scale structure of the universe, and has been used to obtain the first successful detection of a BAO feature at $z \sim 1.5$ (Ata et al., 2017). Going forward, quasars can be effectively used as direct tracers of matter density rather than using tracers of intervening neutral hydrogen alone (i.e., the Lyman-α forest). Readers who are interested in the technical details of the SDSS-IV/eBOSS survey validation results produced in this dissertation should refer to Appendix B. In the following section, I will briefly summarize each of the preceding chapters

and then conclude with ideas on how to further this work.

7.1 SUMMARY OF PREVIOUS CHAPTERS

In Chapter 2, we established a new method for selecting high redshift ($z > 0.6$) LRGs by combining optical and infrared photometry. We emphasized the key role LRGs play in tracing the large-scale structure of the universe. We then explained in detail the construction of datasets which are used for LRG selection and various key aspects of these data sets. We used a combination of optical photometry from CFHTLS and HST, infrared photometry from the WISE satellite, and spectroscopic or photometric redshifts from the DEEP2 Galaxy Redshift Survey or COSMOS to establish this selection algorithm. We presented the shortcomings of previously used methods for selecting LRGs at redshifts $z > 0.6$. These methods depended on optical photometry alone for selecting LRGs. Identifying LRGs with shallow optical photometry becomes prohibitively difficult at higher redshifts as the 4000 Å break passes into the near-infrared, and LRG colors overlap strongly with the colors of M stars. We demonstrated that a simple cut in optical-infrared color-color space provides an efficient method for differentiating LRGs from other types of objects. We presented further optimization of this algorithm by fine-tuning the free parameters of the selection while analyzing the results using ROC and FOM statistics. The methods explained in this chapter have been adapted by other investigators for assembling multiple LRG catalogs for the SDSS-III/BOSS, SDSS-IV/eBOSS, DESI, and 2dFLenS surveys, the goal of which is to measure the BAO feature in LRG samples at higher redshifts.

In Chapter 3, we presented an application of the methods developed in Chapter 2 to assemble LRG targets for SDSS-IV/eBOSS. In a six-year program which started in 2014, eBOSS will perform spectroscopic observations of more than 375,000 new LRGs with a median redshift $z \sim 0.71$ over an area of 7,500 deg^2, making it the largest galaxy survey to date by volume. The eBOSS LRG sample is designed to achieve measurements of the angular diameter distance, $d_A(z)$, to an accuracy of 1.2% and measurements of $H(z)$ to 2.1% when combined with the $z < 0.6$ sample of BOSS galaxies. We began by explaining the success-

ful historical background of previous LRG surveys which inspired astronomers to push the detection of LRGs to higher redshifts. We explained, in brief, the SDSS and WISE imaging surveys which are used to select LRG targets. The detailed explanation of the algorithm employed for the eBOSS LRG selection followed, with some key features of the sample. We presented the results from the SDSS-III/BOSS SEQUELS ancillary program, which was a key survey validation effort for eBOSS.

Exploring systematics that can affect the inferred clustering of targets is often considered only when survey data is used for science analyses. In this thesis, we investigated these issues while simultaneously exploring target selection methods, enabling more informed decisions regarding survey strategy. For instance, foreknowledge of which areas of the survey may pose problems for controlling clustering measurements allows the survey footprint to be modified. We assessed the uniformity of the target sample by comparing the observed density of targets to maps of local imaging conditions and Galactic structure. We applied a multivariate linear regression analysis of surface density against a broad set of tracers of potential systematics; the intention is similar to, e.g., Scranton et al. (2002); Ross et al. (2011a); Ho et al. (2012); Leistedt et al. (2013); Giannantonio et al. (2014), but unlike these investigations, we simultaneously fit for the impact of a wide variety of systematics rather than studying correlation against one systematic at a time. Our approach has the advantage of producing a model of systematics-affected density that will provide accurate predictions for the combined effects of all the systematics considered, even if the input systematic maps are covariant with each other (as, for instance, stellar density and dust extinction must inevitably be). The full potential of this LRG sample can only be realized a few years down the line when most, if not all, of the spectroscopic observations are complete.

In Chapter 4, we presented a detailed analysis of core quasar targets from SDSS-IV/eBOSS. The target selection algorithm used to assemble the quasar sample is beyond the scope of this thesis. Instead, we focused on the key features and statistical analysis of the quasar sample, a work which we led. eBOSS will perform spectroscopic observations of more than 500,000 new quasars in the redshift range $0.9 < z < 2.2$, and will provide the first quasar based BAO distance measurements with the expected precision of 2.8% and 4.2% on $d_A(z)$ and $H(z)$, respectively. eBOSS is the first instance of quasars targeted as direct

tracers of the underlying dark matter distribution on BAO scales, requiring that challenges presented by the target selection requirements of clean and uniform selection be overcome. Similar to the analysis of the eBOSS LRG sample in Chapter 3, we presented a detailed investigation of the uniformity of the target sample through a multivariate linear regression analysis of surface density against a broad set of tracers of potential systematics. The quasar sample shows different statistical dependences on potential systematics than LRGs, which was expected since quasars look very similar to stars in optical imaging. The resulting model of systematic-affected density that will enable accurate predictions for the combined effects of all the systematics considered is analyzed separately in the NGC and SGC regions for reasons explained in the chapter. Quasar and LRG observations are carried out simultaneously in eBOSS, and the full potential of this quasar sample will be realized in a few years when most of the spectroscopic observations are complete. It is worth noting that the quasar sample has already yielded a successful detection of the BAO signal at $z \sim 1.5$ using the first two years of eBOSS data, resulting in a determination of the angular diameter distance, $d_A(z)$, to an accuracy of 4.4% (Ata et al., 2017).

In Chapter 5, we discussed the efforts towards measuring BAO in the eBOSS LRG sample using photometric redshifts. The eBOSS LRG sample consists of ∼600,000 galaxies ranging from redshift $z = 0.5$ to 1.0 over 10,000 deg^2 of the sky, with a median redshift of 0.71. The photometric redshifts are estimated using the Random forest machine learning algorithm, resulting in an RMS error of $\sigma_z/(1+z) \sim 0.028$. We achieved this precision through building robust machine-learning models using a training sample of ∼100,000 LRGs from the first two years of eBOSS observations. We presented the measurement of the angular clustering of this sample, in 8 different redshift bins ranging from $z = 0.5$ to 0.95. We also explained, in detail, the construction of random samples capturing the actual redshift distribution of the LRG sample and the effects of systematic variations in target density. Although there was no clear detection of BAO, we demonstrated the ability to make precise clustering measurements with photometric surveys. This lack of robust detection of BAO could be due to a combination of factors, including but not limited to bad luck, systematic errors, redshift precision, or dilution by stellar contamination. This work has the potential for better measurements with a better understanding of systematic errors, star-galaxy sep-

aration, and more accurate photometric redshifts (as will be provided by BASS/MzLs and DeCALS surveys over the DESI survey footprint; Zhou et. al., in prep.).

Finally, in Chapter 6, we presented initial results from an ongoing effort to measure the clustering of eBOSS quasars around eBOSS LRGs using cross-correlation statistics. Quasar clustering measurements can be used to infer their lifetimes and large-scale environments, which helps to explore questions such as the physical mechanism fueling quasars and their impact on galaxy formation and evolution. For this study, we used samples of $\sim 30,600$ quasars and $\sim 120,000$ LRGs in the redshift range $0.6 < z < 1.0$. These LRGs and quasars have spectroscopically been confirmed over the first two years of eBOSS observations. We modeled the observed clustering using a power law and estimated the robustness of parameters using Jackknife resampling. We presented the initial measurements in redshift bins of $\Delta z = 0.05$. LRGs are lower priority targets than quasars in eBOSS, and hence the cross-correlation is strongly affected by fiber collisions. More work is needed to understand the effects of this on the clustering measurements at small scales. This work is still in progress.

7.2 FUTURE WORK

There are many ways that the work described in this dissertation may be expanded upon, particularly relating to dark energy experiments. The LRG selection methods detailed in Chapter 2 have already been adopted by several international teams to assemble and analyze different galaxy samples. A generalization of our LRG selection method has been implemented to create a massive galaxy catalog for investigating Sunyaev-Zeldovich effects in the CMB. These methods have the potential to be adapted further for targeting even higher redshift ($z > 1.0$) LRGs using redder wavelength channels.

Understanding systematic effects is among the most important aspects of wide field spectroscopic surveys. Spectroscopic probes will only be able to meet the projected cosmological constraints if the final clustering measurements are limited by statistical errors alone. This requirement demands stringent control of systematic uncertainties that can modulate the data on large scales, such as the impact of stellar contamination and dust extinction

on target selection efficiency, variations in seeing that alter target selection and redshift success, and so on. As we push our measurements to higher redshifts, systematic effects present even greater challenges. The multivariate regression technique discussed in Chapters 3 and 4 provides a robust method for understanding and handling all the relevant systematics collectively. Studies of systematics at the target selection phase of surveys also provide a significant advantage as they enable well-informed decisions regarding survey strategies. These include, but are not limited to, possibly avoiding regions of the sky with excessive dust extinction, stellar contamination, etc.

Another important extension of this work would be to investigate the possible improvements in the machine learning approach for estimating photometric redshifts. The next generation of dark energy experiments like LSST, WFIRST, *Euclid*, etc., will provide an unprecedented volume of data for which spectroscopic follow-up is neither feasible not financially practical. Machine learning presents a cheaper, fast and efficient alternative. In Chapter 5, we have demonstrated that photometric redshifts can be used to study large scale structure with imaging surveys. They are useful for spectroscopic surveys like DESI and *EUCLID*, as well, to provide early survey results. An adaptation of our methods would be to use realistic galaxy templates as a training sample to train models which should provide better estimates of photometric redshifts for objects lacking analogs in training sets. The next generation of photometric surveys will provide deeper and less noisy images compared to imaging surveys like the SDSS, which will improve photo-z estimations. The field of machine learning has grown exponentially over the last two decades and will continue to do so. The possibilities of its application in astronomy are endless, and the methods employed here are only one simple example. The undergoing and planned Dark Energy BAO experiments have the ability to enhance our understanding of the universe in an unprecedented way.

APPENDIX A

A.0.1 LRG_IDROP

Objects with LRG-like colors which are too faint for detection in the i band but still have a robust detection in the z-band can be targeted via a different color-cut. The r band photometry for these objects becomes quite noisy and hence it is not used in selection. Instead, we can use a similar selection in a different optical-infrared color-color space:

$$i_{Model} > 21.8, \tag{A.1}$$

$$z_{Model} \leq 19.5, \tag{A.2}$$

$$i - z > 0.7, \tag{A.3}$$

$$i - W1 > 2.143 \times (i - z) - 2.0. \tag{A.4}$$

Equations A.3 and A.4 represent an analogous color selection to equations 3.10 and 3.11, but using the i and z bands instead of r and i. Equation A.2 ensures that the objects are well-detected in the z-band despite having a noisy detection (if any) in bluer bands. This selection contributes a few targets, $\sim$ 200 over the entire footprint, which are expected to be at higher redshifts than the standard *eBOSS* LRG sample.

APPENDIX B

B.0.1 RESULTS FROM A LARGE PILOT SURVEY, SEQUELS

As mentioned in Section 3.4, the basic ideas underlying the *eBOSS* selection algorithm can be implemented in a variety of optical-infrared color spaces. To determine the optimum selection algorithm between two candidate methods, we selected ∼70,000 LRGs over an area of ∼700 deg^2 with 120.0°$< \alpha <$ 210.0°and 45.0°$< \delta <$ 60.0°. These LRGs were selected by algorithms utilizing two different optical-IR color spaces, and were used to test our selection efficiency and redshift success. The parameters of the selection algorithms were tuned such that one obtains a target density of ∼60 deg^{-2} from each one. In the following sub-sections, we explain the two selection algorithms with their commonalities and major differences.

B.0.2 COMMON CUTS FOR *SEQUELS* LRG SAMPLES

First, we require that the RESOLVE_STATUS bit corresponding to SURVEY PRIMARY is nonzero in order to remove duplicate objects. We also require the photometric flag have the CALIB_STATUS bit set for all of the r, i, and z bands used for photometric color determinations. In addition, the following flux limits are applied over the entire sample:

$$z_{Fiber2} \leq 21.7, \text{and} \tag{B.1}$$

$$i_{Model} \geq 19.9, \tag{B.2}$$

B.0.2.1 `r/i/z/WISE LRG selection` In the first selection, we identify LRGs using *r-W1*, *r-i* and *i-z* color. This selection algorithm is very similar to the selection described in Section 3.5, differing only due to changes in flux limits to improve completeness. In addition to the common cuts described above, we apply the following selection criteria:

$$z_{Model} \leq 19.95, \tag{B.3}$$

$$r - i > 0.98, \tag{B.4}$$

$$r - W1 > 2.0 \times (r - i), \text{and} \tag{B.5}$$

$$i - z > 0.625, \tag{B.6}$$

where all variables have the same meanings as in section 3.5.2. These equations and their relevance have been explained previously in Section 3.5.

B.0.2.2 `i/z/WISE LRGs` The second selection is implemented exclusively in *i-W1* and *i-z* optical-IR color-color space, eliminating any use of the *r* band. This selection algorithm is similar to the one explained in Section A, differing primarily in its flux limits, which have been tuned to produce the same target density as the *r/i/z/WISE* selection. In addition to the common cuts, we apply the following selection criteria:

$$z_{Model} \leq 19.5, \tag{B.7}$$

$$i - z > 0.7, \text{and} \tag{B.8}$$

$$i - W1 > 2.143 \times (i - z) - 2.0 \tag{B.9}$$

The equations and their relevance are the same as explained previously in LRG_IDROP (Section A).

B.0.3 DETAILS OF THE *SEQUELS* SURVEY

SEQUELS was conceived as a precursor of *eBOSS* enabling us to test the reliability and efficiency of our selection algorithms while simultaneously producing data that could be combined with the full *eBOSS* dataset to constrained cosmology. It provided a sufficiently large dataset to enable robust tests of selection algorithms. It was also critical in testing and demonstrating our ability to meet *eBOSS* requirements via these selection algorithms. We applied both of the selection algorithms explained in the section Section B.0.2 in parallel over the entire *SDSS* footprint. The final *SEQUELS* LRG sample consisted of the objects selected by either or both of the selection algorithms explained above.

B.0.3.1 Targeting bits In order to identify LRGs selected via different algorithms, we assign them different values of the `eBOSS_TARGET0` tag. For LRGs selected in *i/z/WISE* color space, `eBOSS_TARGET0` is set bit-wise to 1.[1] For LRGs selected via *r/i/z/WISE* selection, `eBOSS_TARGET0` is set bit-wise to 2. LRGs which pass both of the selection criteria have both bits set.

B.0.3.2 Overall characteristics of *SEQUELS* LRGs The two classes of LRGs, i.e., *r/i/z/WISE* selected and *i/z/WISE* selected, were analyzed separately. We found that $\gtrsim$ 87% of spectra yielded secure redshift measurements. Redshift measurements are checked via visual inspection of the spectra. The remaining 13% were found to have small differences between the depths of the lowest chi-squared minima, and hence were judged not to be reliable; this generally occurred due to low signal-to-noise ratio in the spectra. 8% of the total targets were both classified securely and found to be stars. These two factors (13% of targets having no definitive redshift measurement and another 8% being stars) make it impossible to reach the required efficiency at targeting $0.6 < z < 1.0$ LRGs of 80%. We meet the requirement set on the *eBOSS* median redshift using the *r/i/z/WISE* algorithm, but not the *i/z/WISE* algorithm. Among the objects which failed to yield a secure redshift measurement, most were noise-dominated. We tabulate the key results in Table B1.

[1] bit 0, 1, 2 are used to indicate $2^0 = 1$, $2^1 = 2$, and $2^2 = 4$, respectively.

Table B1. Summary of *r/i/z/WISE* and *i/z/WISE* in comparison to key *eBOSS* requirements.

Requirement	***r/i/z/WISE***	***i/z/WISE***	Summary
# of targets: $> 375,000$	450,000 ($\sim$ 60 targets deg^{-2})	450,000 ($\sim$ 60 targets deg^{-2})	Easily achievable
Median Redshift: >0.71	0.716	0.697	*i/z/WISE* failing marginally
Fraction at $0.6 \lesssim z \lesssim 1.0$: $> 80\%$	$\sim$ 71%	$\sim$ 64%	Both samples fail to meet

Note. — The *r/i/z/WISE* selection meets the basic median redshift requirement which is necessary to achieve our science goals. However, both algorithms fail to meet the redshift efficiency requirement. *r/i/z/WISE* selects more high-redshift LRGs and hence was chosen as the preferred selection algorithm for *eBOSS*.

In Figure B1 we present the redshift distributions, *N(z)*, of *r/i/z/WISE* and *i/z/WISE* LRGs. We find that the *i/z/WISE* selection algorithm selects a significantly higher fraction of fraction of galaxies at $0.4 < z < 0.5$ compared to the *r/i/z/WISE* selection. This causes the median redshift and targeting efficiency to fall below our requirements, as seen also in Table B1. Overall, *r/i/z/WISE* was found to be more suitable for *eBOSS*. It gains greater efficiency by requiring targets to be red in both *r*-*i* and *i*-*z*, providing a veto in cases where one color is affected by bad photometry. However, at redshifts $z \gtrsim 0.75$ both of the candidate selection algorithms yielded similar results.

B.0.4 DIFFERENCES BETWEEN *SEQUELS* AND *eBOSS* TARGETS

Post SEQUELS, we made a few improvements in our target selection algorithm. These changes are expected to improve our secure redshift measurement rate by removing objects whose counterparts yielded extremely low signal-to-noise spectra in *SEQUELS*. For *eBOSS* LRGs, we add two additional criteria to the *SEQUELS r/i/z/WISE* selection:

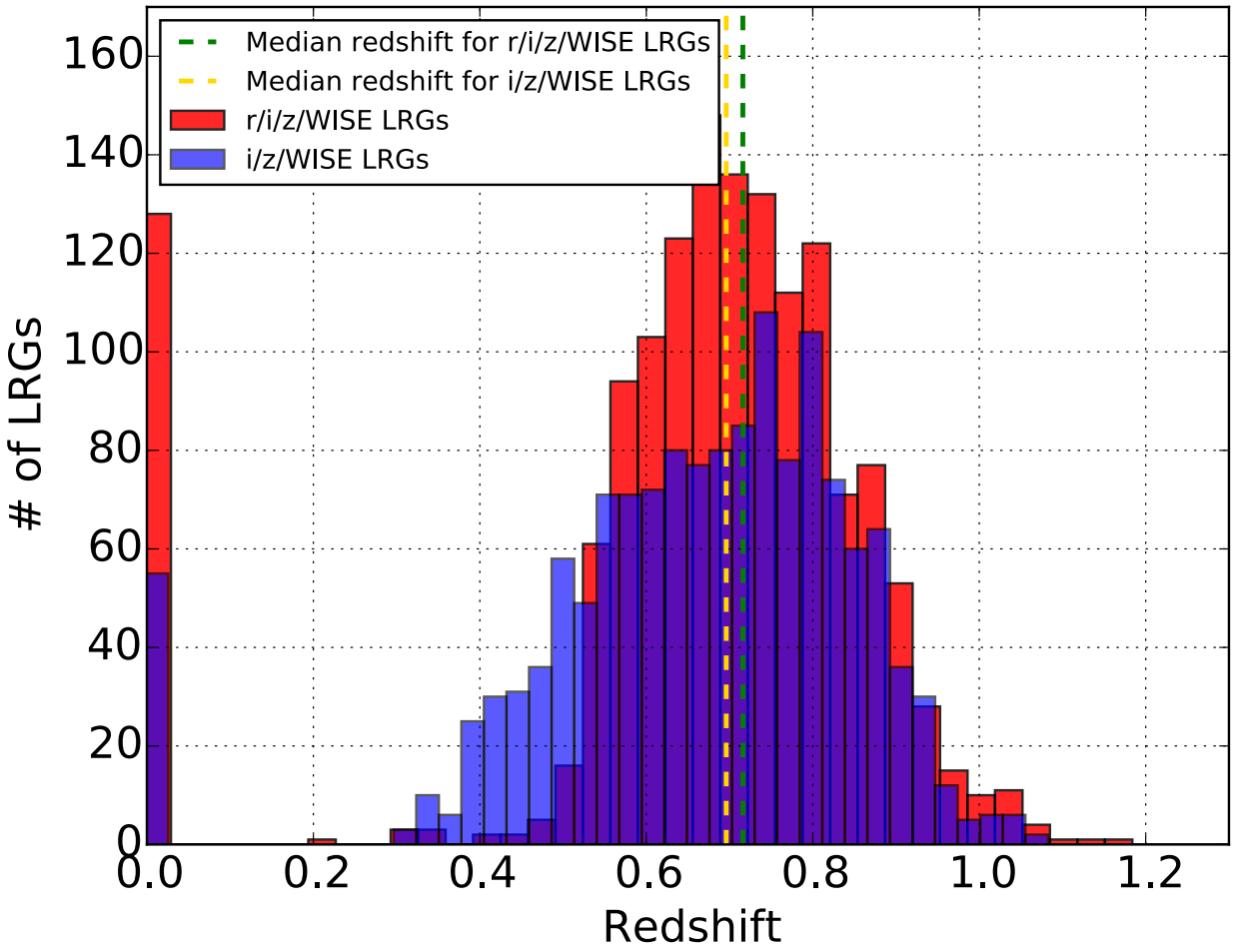


Figure B1: Redshift histogram of ∼1500 visually inspected LRGs targeted by two different selection algorithms as part of SEQUELS. *r/i/z/WISE* selects more LRGs at higher redshift. In contrast, *i/z/WISE* selects more LRGs at lower redshifts which are less useful for *eBOSS*. Hence, *r/i/z/WISE* is the preferred choice for the *eBOSS* LRG sample.

$$W1_{AB} \leq 20.299, \text{and} \tag{B.10}$$

$$i_{Model} \leq 21.8 \tag{B.11}$$

Equation B.10 effectively requires a 5σ detection in the first channel (*W1*) of *WISE*. In addition, we put a faint limit on $i_{Modelflux}$ through equation B.11; this was not applied in *SEQUELS*. These additional flux limits reduce the number of noise-dominated LRG spectra significantly when applied to the *SEQUELS* sample.

BIBLIOGRAPHY

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Alonso, D. 2012, ArXiv e-prints, arXiv:1210.1833

Amit, Y., & Geman, D. 1997, Neural Computation, 9, 1545

Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26

Ata, M., Baumgarten, F., Bautista, J., et al. 2017, ArXiv e-prints, arXiv:1705.06373

Aubourg, É., Bailey, S., Bautista, J. E., et al. 2014, ArXiv e-prints, arXiv:1411.1074

Baum, W. A. 1962, in IAU Symposium, Vol. 15, Problems of Extra-Galactic Research, ed. G. C. McVittie, 390

Bertin, E. 2006, in Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, 112

Bertin, E., Mellier, Y., Radovich, M., et al. 2002, in Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI, ed. D. A. Bohlender, D. Durand, & T. H. Handley, 228

Beutler, F., Saito, S., Brownstein, J. R., et al. 2014a, MNRAS, 444, 3501

Beutler, F., Saito, S., Seo, H.-J., et al. 2014b, MNRAS, 443, 1065

Blake, C., Kazin, E. A., Beutler, F., et al. 2011, MNRAS, 418, 1707

Blanton, M. R., Lin, H., Lupton, R. H., et al. 2003, AJ, 125, 2276

Blanton, M. R., & Roweis, S. 2007, AJ, 133, 734

Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144

Boroson, T. A., & Lauer, T. R. 2009, Nature, 458, 53

Boyle, B. J., Fong, R., Shanks, T., & Peterson, B. A. 1990, MNRAS, 243, 1

Breiman, L. 1996, Machine Learning, 24, 123

Buitinck, L., Louppe, G., Blondel, M., et al. 2013, in ECML PKDD Workshop: Languages for Data Mining and Machine Learning, 108–122

Cannon, R., Drinkwater, M., Edge, A., et al. 2006, MNRAS, 372, 425

Comparat, J., Delubac, T., Jouvel, S., et al. 2015a, ArXiv e-prints, arXiv:1509.05045

—. 2015b, ArXiv e-prints, arXiv:1509.05045

Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, MNRAS, 356, 415

Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 392, 19

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013a, AJ, 145, 10

—. 2013b, AJ, 145, 10

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2015, ArXiv e-prints, arXiv:1508.04473

—. 2016, AJ, 151, 44

Eisenstein, D. J., & Hu, W. 1998, ApJ, 496, 605

Eisenstein, D. J., Annis, J., Gunn, J. E., et al. 2001, AJ, 122, 2267

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Fitzpatrick, E. L. 1999, PASP, 111, 63

Friedman, J., Hastie, T., & Tibshirani, R. 2000, Ann. Statist., 28, 337

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

—. 1996b, AJ, 111, 1748

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

Gerke, B. F., Newman, J. A., Faber, S. M., et al. 2007, MNRAS, 376, 1425

Giannantonio, T., Ross, A. J., Percival, W. J., et al. 2014, Phys. Rev. D, 89, 023511

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Green, R. F., Schmidt, M., & Liebert, J. 1986, ApJS, 61, 305

Griffith, R. L., Cooper, M. C., Newman, J. A., et al. 2012, ApJS, 200, 9

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

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

Gwyn, S. D. J. 2011a, ArXiv e-prints, arXiv:1101.1084

—. 2011b, ArXiv e-prints, arXiv:1101.1084

Hanley, J. A., & McNeil, B. J. 1982, Radiology, 143, 29

Hennawi, J. F., Strauss, M. A., Oguri, M., et al. 2006, AJ, 131, 1

Ho, S., Cuesta, A., Seo, H.-J., et al. 2012, ApJ, 761, 14

Ho, T. K. 1995, in Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1, ICDAR '95 (Washington, DC, USA: IEEE Computer Society), 278–

Ho, T. K. 1998, IEEE Trans. Pattern Anal. Mach. Intell., 20, 832

Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

Hu, W. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 339, Observing Dark Energy, ed. S. C. Wolff & T. R. Lauer, 215

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Ilbert, O., Salvato, M., Capak, P., et al. 2008a, in Astronomical Society of the Pacific Conference Series, Vol. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki, 169

Ilbert, O., Salvato, M., Capak, P., et al. 2008b, in Astronomical Society of the Pacific Conference Series, Vol. 399, Panoramic Views of Galaxy Formation and Evolution, ed. T. Kodama, T. Yamada, & K. Aoki, 169

Inada, N., Oguri, M., Pindor, B., et al. 2003, Nature, 426, 810

Jarrett, T. H., Cohen, M., Masci, F., et al. 2011a, ApJ, 735, 112

—. 2011b, ApJ, 735, 112

Jennings, E., Baugh, C. M., Li, B., Zhao, G.-B., & Koyama, K. 2012, MNRAS, 425, 2128

John, T. L. 1988, A&A, 193, 189

Johri, V. B., & Rath, P. K. 2007, International Journal of Modern Physics D, 16, 1581

Jones, D. 1968, Leaflet of the Astronomical Society of the Pacific, 10, 145

Kaiser, N. 1984, ApJ, 284, L9

Kaiser, N., Burgett, W., Chambers, K., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7733, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0

Kaspi, S., Brandt, W. N., Maoz, D., et al. 2007, ApJ, 659, 997

Kauffmann, G., & Haehnelt, M. 2000, MNRAS, 311, 576

Kauffmann, G., White, S. D. M., Heckman, T. M., et al. 2004, MNRAS, 353, 713

Kennefick, J. D., de Carvalho, R. R., Djorgovski, S. G., et al. 1995, AJ, 110, 78

Kerscher, M., Szapudi, I., & Szalay, A. S. 2000, ApJ, 535, L13

Laher, R. R., Surace, J., Grillmair, C. J., et al. 2014, PASP, 126, 674

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

Lang, D., Hogg, D. W., & Schlegel, D. J. 2014, ArXiv e-prints, arXiv:1410.7397

Laurent, P., Eftekharzadeh, S., Le Goff, J.-M., et al. 2017, J. Cosmology Astropart. Phys., 7, 017

Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395

Leistedt, B., Peiris, H. V., Mortlock, D. J., Benoit-Lévy, A., & Pontzen, A. 2013, MNRAS, 435, 1857

Limber, D. N. 1954, ApJ, 119, 655

Lin, Y.-T., & Mohr, J. J. 2003, ApJ, 582, 574

Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406

Matthews, D. J., Newman, J. A., Coil, A. L., Cooper, M. C., & Gwyn, S. D. J. 2013, ApJS, 204, 21

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

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

Mobasher, B., Capak, P., Scoville, N. Z., et al. 2007, ApJS, 172, 117

Morganson, E., Green, P. J., Anderson, S. F., et al. 2015, ApJ, 806, 244

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

Netzer, H., & Trakhtenbrot, B. 2007, ApJ, 654, 754

Newberg, H. J., & Yanny, B. 1997, ApJS, 113, 89

Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, ApJS, 208, 5

Ofek, E. O., Laher, R., Law, N., et al. 2012, PASP, 124, 62

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

—. 1983b, ApJ, 266, 713

Padmanabhan, N., Schlegel, D. J., Seljak, U., et al. 2007, MNRAS, 378, 852

Padmanabhan, N., Schlegel, D. J., Finkbeiner, D. P., et al. 2008, ApJ, 674, 1217

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

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

Peebles, P. J. E., & Ratra, B. 1988, ApJ, 325, L17

Peebles, P. J. E., & Yu, J. T. 1970, ApJ, 162, 815

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16

Postman, M., & Geller, M. J. 1984, ApJ, 281, 95

Postman, M., & Lauer, T. R. 1995, ApJ, 440, 28

Prakash, A., Licquia, T. C., Newman, J. A., & Rao, S. M. 2015, ApJ, 803, 105

Prakash, A., Licquia, T. C., Newman, J. A., et al. 2016, ApJS, 224, 34

Rau, A., Kulkarni, S. R., Law, N. M., et al. 2009, PASP, 121, 1334

Reid, B., Ho, S., Padmanabhan, N., et al. 2016, MNRAS, 455, 1553

Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Ross, A. J., Ho, S., Cuesta, A. J., et al. 2011a, MNRAS, 417, 1350

—. 2011b, MNRAS, 417, 1350

Ross, A. J., Percival, W. J., Sánchez, A. G., et al. 2012a, MNRAS, 424, 564

Ross, A. J., Samushia, L., Burden, A., et al. 2014, MNRAS, 437, 1109

Ross, N. P., Shanks, T., Cannon, R. D., et al. 2008, MNRAS, 387, 1323

Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012b, ApJS, 199, 3

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104

Samushia, L., Reid, B. A., White, M., et al. 2014, MNRAS, 439, 3504

Sandage, A., & Luyten, W. J. 1969, ApJ, 155, 913

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

—. 2011b, ApJ, 737, 103

Schlafly, E. F., Finkbeiner, D. P., Jurić, M., et al. 2012, ApJ, 756, 158

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998a, ApJ, 500, 525

—. 1998b, ApJ, 500, 525

Schmidt, M. 1963, Nature, 197, 1040

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

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

Seo, H.-J., & Eisenstein, D. J. 2003, ApJ, 598, 720

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

Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013a, AJ, 146, 32

—. 2013b, AJ, 146, 32

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485

Sunyaev, R. A., & Zeldovich, I. B. 1980, ARA&A, 18, 537

Totsuji, H., & Kihara, T. 1969, PASJ, 21, 221

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549

Warren, S. J., Hewett, P. C., Irwin, M. J., McMahon, R. G., & Bridgeland, M. T. 1987, Nature, 325, 131

Weinberg, S. 1987, Physical Review Letters, 59, 2607

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010a, AJ, 140, 1868

—. 2010b, AJ, 140, 1868

Yan, L., Donoso, E., Tsai, C.-W., et al. 2013, AJ, 145, 55

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

—. 2000b, AJ, 120, 1579

York, D. G., Khare, P., Vanden Berk, D., et al. 2006, MNRAS, 367, 945

Zhao, G.-B., Wang, Y., Ross, A. J., et al. 2016, Monthly Notices of the Royal Astronomical Society, 457, 2377