Optical Detection of Galaxy Clusters

Samuel Farrens

Thesis submitted for the Degree of Doctor of Philosophy
of the University of London

Department of Physics & Astronomy
UNIVERSITY COLLEGE LONDON

November 2010

I, Samuel Farrens, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
Specifically:

- **Chapter 2:** This chapter is based on work presented in Farrens et al. (2010), which includes contributions by Filipe Abdalla, Eduardo Cyrpriano, Cristiano Sabiu and Chris Blake. In §2.4, the mock catalogue described and fig.2.3, and in §2.6, the correlation function calculations, fig.2.14 and fig.2.15 were produced by Cristiano Sabiu.

- **Chapter 3:** This chapter is based on work to be presented in Farrens et al. (*in prep*), which includes contributions by Filipe Abdalla, Eduardo Cyrpriano, Cristiano Sabiu and Chris Blake. In §3.4, the mock catalogues described and fig.3.4, and in §3.6, the correlation function calculations and fig.3.23 were produced by Cristiano Sabiu.

- **Chapter 4:** In §4.4.1, the reduction of the data obtained using the Anglo-Australian Telescope using the 2dfdr software package was performed by Chris Blake, Sarah Brough and Kevin Pimbblet. In §4.4.2, additional redshifting of the reduced data was carried out by Filipe Abdalla, Chris Blake, Sarah Brough and Kevin Pimbblet.

- **Chapter 5:** In §5.3.1, fig.5.3 and fig.5.4 were provided by Jorge Carretero Palacios. In §5.4.3.2, cluster catalogues are provided by Marcelle Soares-Santos and Jiangang Hao for comparison.

To my father who has instilled in me a love of learning.

“Realists do not fear the results of their study.”

— Fyodor Mikhaylovich Dostoyevsky

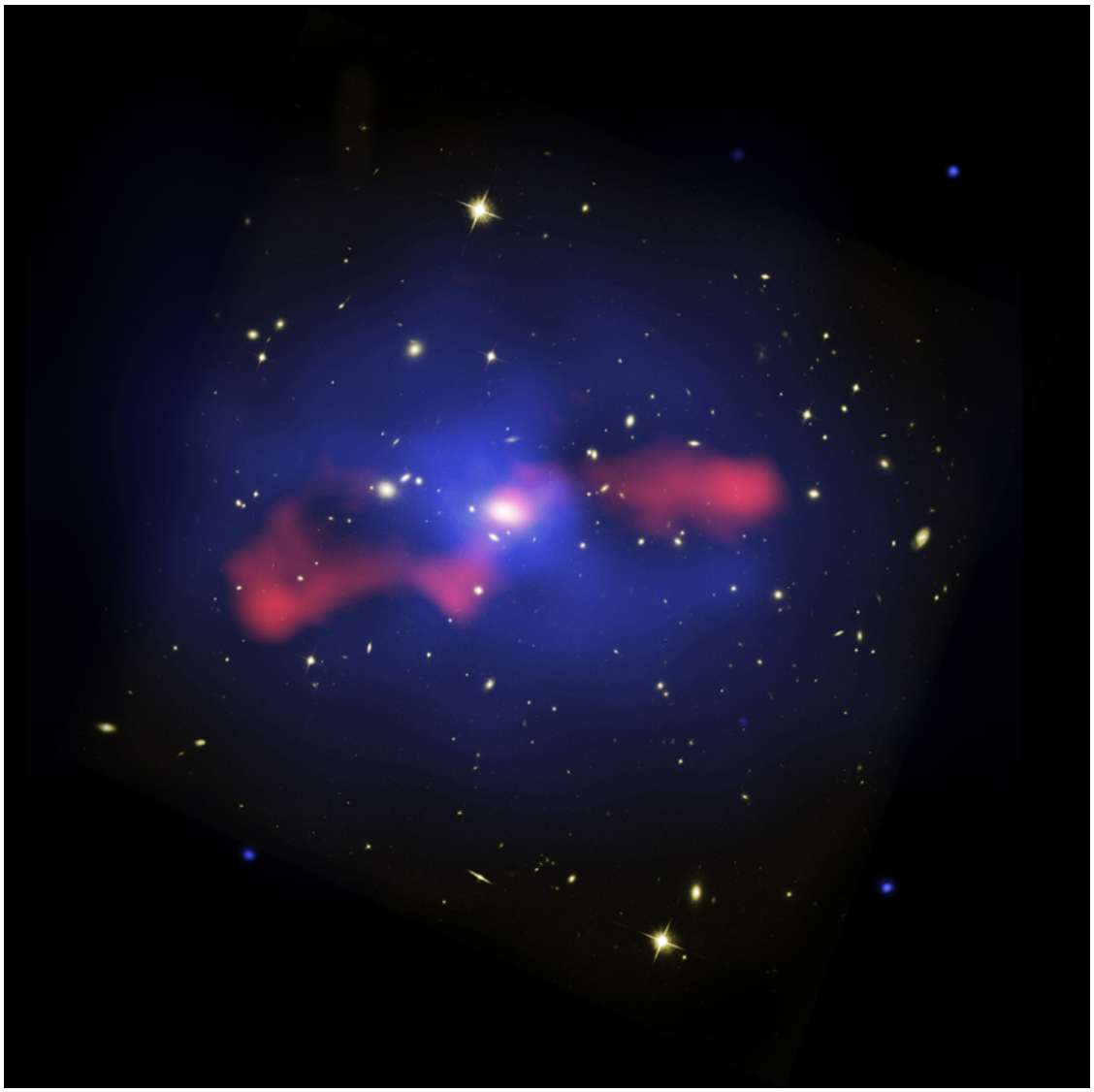
ABSTRACT

This thesis first presents a relatively straight forward approach for detecting galaxy clusters using spectroscopic data. A friend-of-friends algorithm based on that of Huchra & Geller (1982) is implemented with linking parameters that take into account selection effects on the the 2dF-SDSS and QSO (2SLAQ) Luminous Red Galaxy Survey (Cannon *et al.* 2006). The linking parameters are constrained using a mock halo catalogue. The galaxy groups and clusters found with this method have an average velocity dispersion of $\overline{\sigma}_v = 467.97$ kms^{-1} and an average size of $\overline{R}_{\rm clt} = 0.78$ h^{-1}Mpc. Cluster masses are estimated using the cluster velocity dispersions and appear consistent with values expected for genuine structures.

The spectroscopic cluster catalogue is then used to calibrate and compare with a more complex method for detecting clusters using photometric redshifts based on the method of Botzler *et al.* (2004). The spectroscopic cluster catalogue can be reproduced by around 38% and up to 80% if matching is made only to groups and clusters with six or more members. This code is also applied to the Mega$z$-LRG DR7 catalogue (Collister & Lahav 2004) producing two catalogues. One that appears to have a good level of completeness relative to the 2SLAQ spectroscopic catalogue.

A spectroscopic follow up of some preliminary results from the photometric cluster finder was made using the Anglo-Australian Telescope, which show that the majority of the clusters analysed are genuine and approximate masses can be estimated from the cluster velocity dispersions.

Finally, some initials results from on going work in the Dark Energy Survey collaboration are presented, which cover simulated galaxy photometric redshift and colour analysis as well as cluster detection.



Composite Image of Galaxy Cluster MS 0735. Hubble and Chandra Image Credit: NASA, ESA, CXC, STScI, and B. McNamara (University of Waterloo). Very Large Array Telescope Image Credit: NRAO, and L. Birzan and team (Ohio University).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Prelude

Isaac Newton is famously quoted as writing that *"If I have seen a little further it is by standing on the shoulders of Giants"* in a letter to Robert Hooke. Although it is debated wether or not Newton's comment was intended as derogatory or philosophical, it is a rather appropriate metaphor to the world of science. Science is about contribution and collaboration. Despite the fact that there have been certain special names who have made a rather significant contribution to science, science is not about individuals. The work of any scientist would be impossible without thousands of years of wisdom that has come before and science will continue to grow and evolve regardless of any revolutionary individual contributions. The work presented in this thesis is a contribution, perhaps a small one, but a contribution none the less.

This section gives a brief summary of the work of some of those 'giants' upon whose shoulders we stand everyday, perhaps without even realising it.

1.2 Standard Model of Cosmology

This section outlines some of the principal concepts and equations that make up the standard model of cosmology. It begins with a look at an expanding Universe (§1.2.1), followed by the relations between the density and geometry of the Universe (§1.2.2), it then briefly describes current measurements of the cosmological parameters (§1.2.3) and finishes by explaining some of the methods that allow us to measure distance in a dynamic Universe (§1.2.4).

1.2.1 The Expanding Universe

One of the major discoveries of the 20^{th} Century was that our galaxy is not alone in the Universe. Not only are there countless other galaxies beyond the Milky Way, but they are all appear to be moving away from each other. This led to incredible realisation that Universe itself was expanding (Hubble 1929).

1.2.1.1 Redshift

In 1848, French physicist Armand-Hippolyte-Louis Fizeau noticed that the shift in stellar spectral lines could be attributed to the stretching of the light waves emitted by a star as it moves away from us via the Doppler effect. In 1908, the term 'red-shift' was coined by American astronomer Walter S. Adams to describe this process (Adams 1908).

Redshift is generally described as an effect whereby the light emitted by an object receding along the line of sight appears to be shifted towards the red end of the electromagnetic spectrum. The reverse process for objects moving towards us is called blueshift. We define the redshift, z, of an object as

$$z = \frac{\lambda_o}{\lambda_e} - 1, \tag{1.1}$$

where λ_o is the wavelength of light observed for the object and and λ_e is the wavelength of light emitted by the object.

In 1868, British astronomer William Huggins used the redshift of a star moving away from the Earth to determine its velocity (Huggins 1868). For small redshifts a linear proportionality to velocity v can be used,

$$z = \frac{v}{c}, \tag{1.2}$$

where c is the speed of light in a vacuum. This relation allows us to relate the redshifts that we can observe for astronomical objects to their recession velocities.

1.2.1.2 Hubble's Law

In first half of the 20^{th} century, while working with the Hooker telescope at Mount Wilson in California, American astronomer Edwin Powell Hubble made a series of discoveries that greatly changed the way we view the Universe. Hubble discovered, first, that there exist galaxies outside of the Milky Way and, second, these galaxies appear to be receding from us. This was a shocking

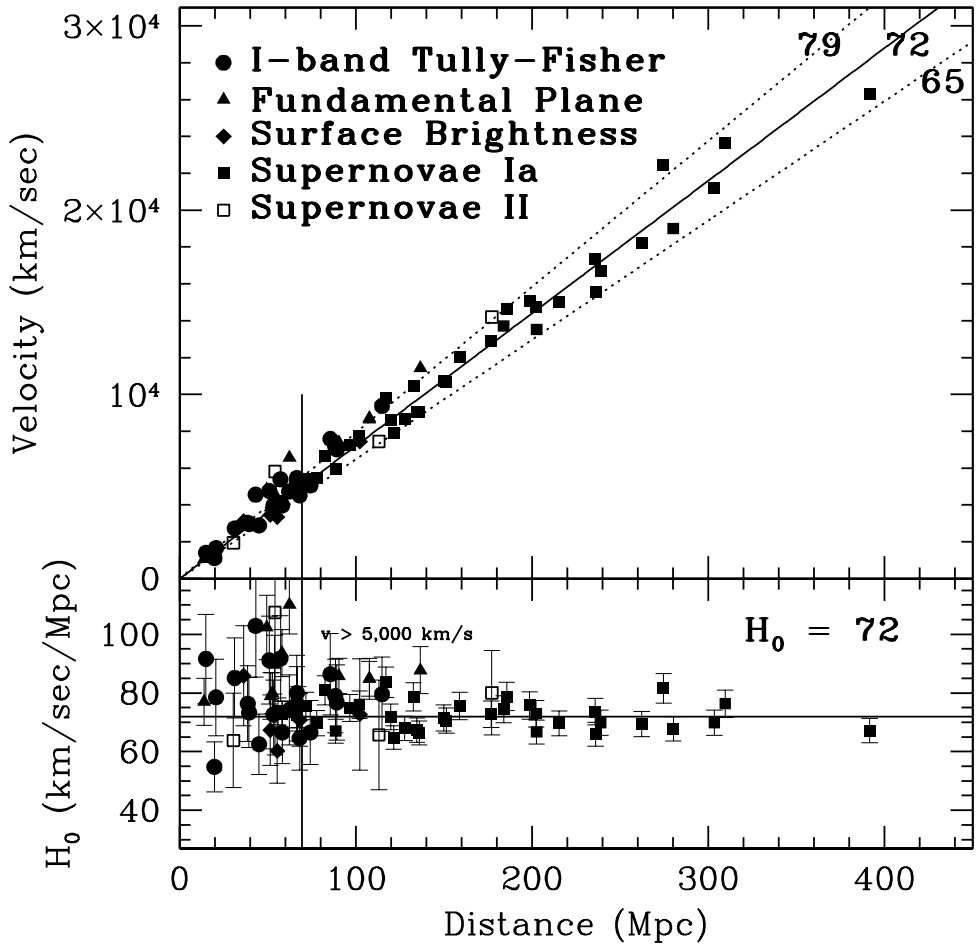


Figure 1.1: Hubble diagram of distance (in Mpc) vs. velocity using five different measures of distance (Freedman *et al.* 2001). The bottom panel shows the value of H_0 as a function of distance.

revelation that contradicted the widely accepted belief that the Universe was static. Using a sample of only 24 galaxies, Hubble (1929) found a linear relation between the distance and the velocities of the galaxies. Recent studies using various different distance measures show this same trend seen in fig.1.1 (Freedman *et al.* 2001).

Hubble's law relates the recession speed, v, to the distance, d, of an object in the expanding Universe via

$$v = H_0 d, \tag{1.3}$$

where H_0 is the Hubble constant,

$$H_0 = 100\ h\ \text{km s}^{-1}\text{Mpc}^{-1}. \tag{1.4}$$

Hubble's law can also be rearranged to relate the redshift of an object to its distance using 1.2,

$$z = \frac{H_0 d}{c}, \tag{1.5}$$

where the redshift is owing to the expansion of the Universe itself. This relation is only valid to $z \sim 0.2$ after which a more complicated approximation is required,

$$d = \frac{cz}{H_0}\Big(1 + \frac{1}{2}(1 - q_0)z\Big), \tag{1.6}$$

(Zeilik & Gregory 1998) where q_0 is the current value of the deceleration parameter, $q(t) = -\ddot{a}a/\dot{a}^2$, and $a(t)$ is the scale factor.

The scale factor is a term that represents the relative expansion of the Universe and can be related to redshift via

$$z = \frac{\lambda_o}{\lambda_e} - 1 = \frac{a(t_o)}{a(t_e)} - 1, \tag{1.7}$$

where t_o and t_e are the times of observation and emission respectively.

1.2.1.3 The Friedmann Equation

In 1922, Russian cosmologist and mathematician Alexander Friedmann derived a solution to Einstein's field equations for an expanding Universe (Friedmann 1922). This solution corroborated Hubble's later discovery that galaxies outside of our own appear to be moving away from us.

If one assumes a sphere expanding in an homogenous universe inside which the mass, M, remains constant. The acceleration due to gravity takes the form

$$\ddot{R} = -\frac{\mathrm{G}M(R)}{R^2} = -\frac{4\pi}{3}\mathrm{G}\rho R, \tag{1.8}$$

where G is the gravitational constant, R is the radius of the sphere and ρ is the density within the sphere. If the mass is conserved then

$$\frac{dM}{dt} = -\frac{4\pi}{3}\Big(R^3\frac{d\rho}{dt} + 3\rho R^2\frac{dR}{dt}\Big) = 0 \tag{1.9}$$

and therefore

$$\frac{d\rho}{dt} + 3\rho\frac{\dot{R}}{R} = 0. \tag{1.10}$$

We define the Hubble rate, which measures how rapidly the scale factor a changes, as

$$H(t) = \frac{\dot{a}}{a} = \frac{\dot{R}}{R}. \tag{1.11}$$

Therefore substituting 1.11 into 1.10 gives

$$\frac{d\rho}{dt} + 3\rho H(t) = 0, \tag{1.12}$$

which is the Newtonian form of the fluid equation.

Multiplying equation 1.8 by $2\dot{R}$ leads to

$$2\dot{R}\ddot{R} = -\frac{8\pi}{3}\mathrm{G}\rho R\dot{R}. \tag{1.13}$$

Integrating equation 1.13 gives

$$\dot{R}^2 = \frac{8\pi}{3}\mathrm{G}\rho R^2 + \mathrm{A}, \tag{1.14}$$

where A is a constant (Milne 1934). Dividing 1.14 by R^2 gives

$$H(t)^2 = \frac{8\pi}{3}\mathrm{G}\rho + \frac{\mathrm{B}}{a^2}, \tag{1.15}$$

where B=Aa^2/R^2. For B=0 we can define the critical density as

$$\rho_{\mathrm{crit}} = \frac{3H(t)^2}{8\pi\mathrm{G}}. \tag{1.16}$$

The critical density is a measure of the density required to make a flat universe (§1.2.2). We can now define the density parameter, $\Omega(t)$, as

$$\Omega(t) = \frac{\rho}{\rho_{\mathrm{crit}}}. \tag{1.17}$$

Hence we can see that

$$\rho(t) = \Omega(t)\frac{3H(t)^2}{8\pi\mathrm{G}}. \tag{1.18}$$

Substituting equation 1.18 at $t = 0$ into equation 1.15 gives

$$\mathrm{B} = H_0^2(1 - \Omega_0). \tag{1.19}$$

Which leads to

$$H(t)^2 = \frac{8\pi}{3}\mathrm{G}\rho + \frac{H_0^2(1 - \Omega_0)}{a^2}. \tag{1.20}$$

Thus the Friedmann Equation can be expressed as

$$H(t)^2 = \frac{8\pi}{3}\mathrm{G}\Big(\rho(t) + \frac{\rho_{\mathrm{crit}} - \rho_0}{a(t)^2}\Big), \tag{1.21}$$

(Dodelson 2003). This is arguably one of the most important equations in cosmology. This equation determines the form of the scale factor, and thus the evolutionary history and fate of the Universe.

When derived from general relativity the Friedmann equation is often expressed in the more general form,

$$H(t)^2 = \frac{8\pi}{3}\mathrm{G}\rho(t) - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \tag{1.22}$$

and is usually accompanied by the acceleration equation

$$\dot{H}(t) + H(t)^2 = \frac{\ddot{a}}{a} = -\frac{4\pi}{3}\mathrm{G}\Big(\rho(t) + \frac{3P}{c^2}\Big) + \frac{\Lambda c^2}{3}, \tag{1.23}$$

where k is term that describes the global curvature of the Universe, Λ is the cosmological constant (§1.2.2) and P is the pressure. Combing 1.22 and 1.23 leads to the full fluid equation,

$$\dot{\rho} + 3\frac{\dot{a}}{a}\Big(\rho + P\Big) = 0. \tag{1.24}$$

This is the equation that describes the evolution of density in the Universe.

1.2.2 Density and Geometry

With the realisation that the Universe is expanding, we are led to question what the fate of the Universe will be. To answer this question we must first ask what the overall density and geometry of the Universe is.

1.2.2.1 Density Parameters

The density parameter, $\Omega(t)$, which is defined by equation 1.17, can decomposed into individual parts,

$$\Omega(t) = \Omega_R(t) + \Omega_M(t) + \Omega_k(t) + \Omega_\Lambda(t), \tag{1.25}$$

where Ω_R is the radiation density, Ω_M is the matter density, which consists of ordinary baryonic matter and dark matter, Ω_k is the spacial curvature density and Ω_Λ is the vacuum density or cosmological constant. Each of these components, which are discussed in the following subsections, make up the Universe as we know it and therefore contribute to the total density. The Friedmann equation (1.21) can also be expressed in terms of the full density parameter,

$$H(t)^2 = H_0^2(\Omega_R a^{-4} + \Omega_M a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda), \tag{1.26}$$

(Peacock 1999). At the present epoch, Ω_R can be expressed as

$$\Omega_R \equiv \frac{\rho_{R0}}{\rho_{cr0}}, \tag{1.27}$$

Ω_M can be expressed as

$$\Omega_M \equiv \frac{\rho_{M0}}{\rho_{cr0}} \equiv \frac{8\pi G\rho_0}{3H_0^2}, \tag{1.28}$$

Ω_Λ can be expressed as

$$\Omega_\Lambda \equiv \frac{\rho_{\Lambda 0}}{\rho_{cr0}} \equiv \frac{\Lambda c^2}{3H_0^2}, \tag{1.29}$$

and Ω_k, can be expressed as

$$\Omega_k \equiv \frac{kc^2}{H_0^2 a_0^2}. \tag{1.30}$$

In general Ω_k can be expressed in terms of Ω_R, Ω_M and Ω_Λ via

$$\Omega_k = 1 - \Omega_R - \Omega_M - \Omega_\Lambda, \tag{1.31}$$

which shows that for a flat Universe (*i.e.* with no curvature, $k = 0$) $\sum_i \Omega_i = 1$ at all times.

1.2.2.2 Radiation

Soon after the Big Bang the Universe was dominated by radiation. Since we can observe a universe full of matter, we need to understand how and when the transition from a radiation dominated to a matter dominated universe occurred.

We can define the energy density of a relativistic gas, ε, as the sum over all frequencies of the number density of photons of energy $h\nu$, $N(\nu)$, via

$$\varepsilon = \sum_\nu h\nu N(\nu). \tag{1.32}$$

(Longair 1998). If the total number of photons is conserved, as the Universe expands, then the number density varies as $N = N_0(1+z)^3 = N_0 a^{-3}$ and the energy of each photon varies as $\nu = \nu_0(1+z) = \nu_0 a^{-1}$. Therefore 1.32 becomes

$$\varepsilon = \sum_{\nu_0} h\nu_0 N_0(\nu_0)(1+z)^4 = \sum_{\nu_0} h\nu_0 N_0(\nu_0) a^{-4}. \tag{1.33}$$

From 1.33 we can see that $\rho_R \propto a^{-4}$. This is an important relation that tells us how the density of radiation scales with time and since it scales as a^{-4} we can see that it becomes very small quite quickly.

The StefanBoltzmann law, which relates the total energy radiated by a black-body to its thermodynamic temperature, is given by

$$F = \sigma T^4, \tag{1.34}$$

where F is the flux and σ is the StefanBoltzmann constant, $\sigma = 5.6704 \times 10^{-8}\text{Js}^{-1}\text{m}^{-2}\text{K}^{-4}$. Therefore $\rho_R \propto T^4$, which tells us that as the Universe expands, a black-body spectrum maintains its form but the temperature changes as $T \propto a^{-1}$. The radiation density parameter can thus be expressed as

$$\Omega_R = \frac{\rho_R}{\rho_{cr}} = \frac{\pi^2 T^4}{15\rho_{cr}} = \frac{\pi^2 T_0^4}{15\rho_{cr}a^4}, \tag{1.35}$$

(Dodelson 2003).

In 1964, while working at Bell Laboratories, Arno Penzias and Robert Woodrow Wilson fortuitously detected the microwave emission left behind by the Big Bang (Penzias & Wilson 1965). The temperature of this emission, now formally known as the Cosmic Microwave Background (CMB) radiation, can be accurately measured as $T_0 = 2.728\text{K}$. From 1.35 we can see that at the present epoch $\Omega_R \simeq 5 \times 10^{-5}$. Therefore a late times the radiation density parameter is often neglected.

1.2.2.3 Baryons

The matter content of the Universe that is made up of atoms (*i.e.* in the form of stars, galaxies, clusters *etc.*) is referred to as baryonic. This is an important definition in cosmology because the majority of matter in the Universe cannot be observed directly and is referred to as dark matter. In fact, the baryonic content of the Universe is only $\sim 5\%$.

In 1905, Albert Einstein showed that the energy of body can be related to its rest mass via

$$E^2 = m^2c^4 + p^2c^2. \tag{1.36}$$

where m is the rest mass and p is the momentum (Einstein 1905). As the Universe expands and cools objects lose energy and exert less pressure, P, on their surroundings. Therefore, at late times with low densities and non-relativistic matter, one can assume $P \simeq 0$ and thus from 1.24 we get

$$\frac{\dot{\rho}}{\rho} = -3\frac{\dot{a}}{a}. \tag{1.37}$$

Equation 1.37 shows us that $\rho_M \propto a^{-3}$, which simply implies that the matter density scales with volume. If we compare this to $\rho_R \propto a^{-4}$ we can see that matter density drops more slowly than the radiation density, which explains the transition from a radiation dominated universe to a matter dominated universe.

1.2.2.4 Dark Matter

In 1933, Swiss astrophysicist Fritz Zwicky noticed the effect of some unseen mass in the Coma cluster (Zwicky 1933). He estimated that there should be ~ 400 times more mass than is observed. Around 40 years later Vera Rubin and Kent Ford measured the velocity curves of edge-on spiral galaxies and found that most stars orbit the centre of the galaxy at the same speed (Rubin & Ford 1970). This discovery implied that there was a large amount of mass in the outer parts of galaxies that was not observed. This missing mass is generally referred to as dark matter and is thought to account for $\sim 23\%$ of the Universe.

Since its discovery, there have been several models that have attempted to explain the nature of dark matter, namely: Hot Dark Matter (HDM) models (Zel'Dovich 1970), which consist of ultra-relativistic nonbaryonic particles, Warm Dark Matter (WDM) models (Pagels & Primack 1982), which consist of relativistic nonbaryonic particles, and Cold Dark Matter (CDM) models (Peebles 1982), which consist of non-relativistic nonbaryonic particles. Today the most widely accepted model is Cold Dark Matter. The term 'cold' is in reference to the thermal velocities of the dark matter particles, which in this model are effectively zero.

CDM makes no predictions as to the actual nature of the dark matter particles. The three principal candidates are WIMPS (Weakly Interacting Massive Particles, Freese 1986; Krauss *et al.* 1986), MACHOs (Massive Compact Halo Objects, Alcock *et al.* 1993*a*,*b*) and Axions (Peccei & Quinn 1977; Efstathiou 1990; Kolb & Turner 1990). WIMPS interact through the weak nuclear force and gravity, but since they do not interact through the electromagnetic force they cannot be directly observed and therefore are very difficult to detect. MACHOs are objects composed of normal baryonic matter, but emit little or no light (*e.g* black holes, neutron stars, brown dwarfs, etc.) and are therefore also very difficult to detect. Axions are elemental particles, which are predicted to have no electric charge, a very small mass and very little interaction with the strong and weak forces. Axions are predicted to change to and from photons in the presence of strong magnetic fields.

1.2.2.5 Curvature

One of the most important factors in determining the fate of the Universe is determining its geometry. Knowing if we live in an open or closed Universe will tell us if it will expand forever or stop expanding at some point and begin to contract.

The curvature term, k, in equation 1.30 can take one of three possible values, as seen in fig.1.2:

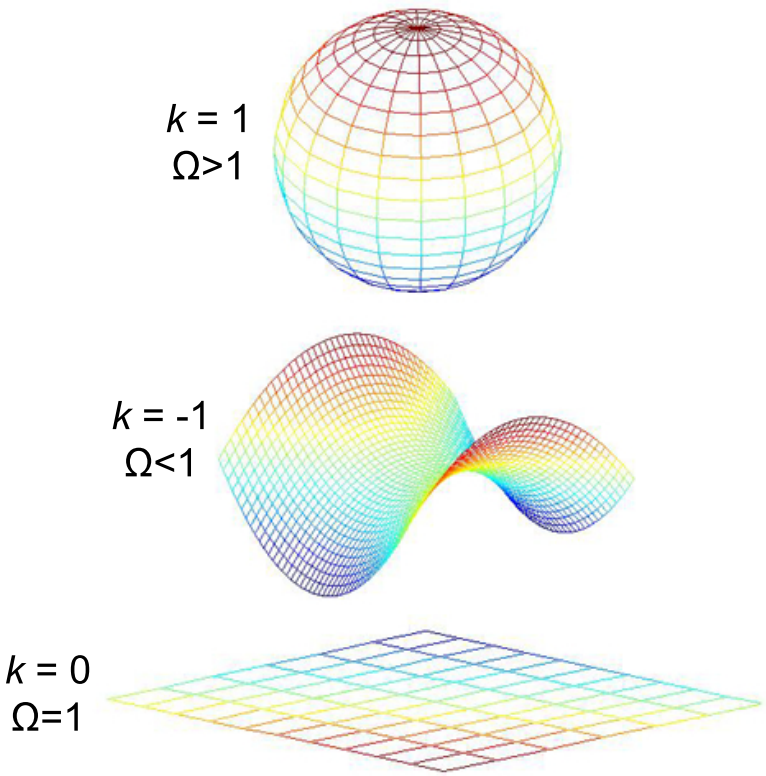


Figure 1.2: Different possible geometries of the Universe for different values of the curvature term, k. $k = 1$ ($\Omega > 1$) corresponds to a closed spherical Universe, $k = -1$ ($\Omega < 1$) corresponds to an open hyperbolic Universe and $k = 0$ ($\Omega = 1$) corresponds to a flat Euclidean Universe.

$k = 0$ ($\Omega = 1$), corresponding to a flat Euclidean Universe in which the sum of the angles of a triangle is 180° and parallel lines remain parallel, $k = 1$ ($\Omega > 1$), corresponding to a spherical Universe in which the sum of the angles of a triangle is greater than 180° and parallel lines eventually converge, and $k = -1$ ($\Omega < 1$), corresponding to a hyperbolic Universe in which the sum of the angles of a triangle is less than 180° and parallel lines eventually diverge.

1.2.2.6 Dark Energy and the Cosmological Constant

Recent observations of Type Ia supernovae have provided evidence that the Universe is accelerating in its expansion (Perlmutter *et al.* 1998; Schmidt *et al.* 1998) as seen in fig.1.3. This is one the most significant discoveries in cosmology, as far as what it implies. Several theories have been proposed to explain the supernovae results, such as a modified theory of general relativity (*e.g.* Carroll *et al.* 2004) or the introduction of a cosmological constant, Λ.

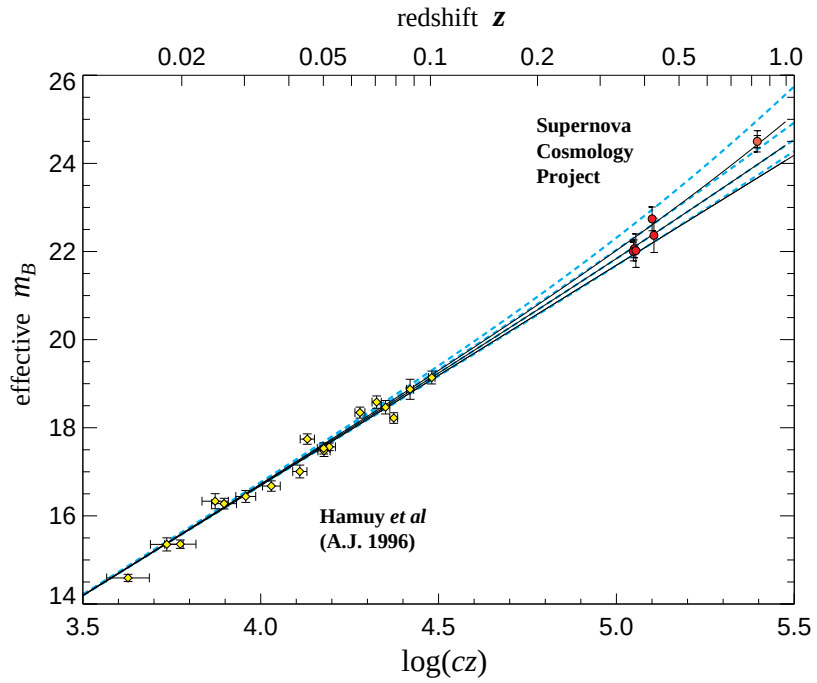


Figure 1.3: SN 1997ap at z = 0.83 plotted on a Hubble diagram with the five of the first seven high-redshift supernovae that could be width-luminosity corrected and the 18 of the lower-redshift supernovae from the Calán/Tololo Supernova Survey that were observed earlier then 5 days after maximum light. Magnitudes have been K-corrected and corrected for the width-luminosity relation. The inner error bar on the SN 1997ap point corresponds to the photometry error alone while the outer error bar includes the intrinsic dispersion of SNe Ia after stretch correction. The solid curves are theoretical mB for (Ω_M,Ω_Λ) = (0,0) on top, (1, 0) in middle, and (2, 0) on bottom. The dotted curves are for the flat universe case, with (Ω_M,Ω_Λ) = (0, 1) on top, (0.5, 0.5), (1, 0), and (1.5,0.5) on bottom (Perlmutter *et al.* 1998).

In addition to the supernovae problem, we live in a Universe that appears very much flat (or critical) and yet, combing ordinary baryonic matter and dark matter, we can only account for 1/3 of the critical matter density (Hinshaw *et al.* 2009). We can, therefore, only infer that there is some other unknown entity that makes up the remaining 2/3 of the density in the Universe. This entity has been dubbed dark energy and is believed to permeate all of space. Dark energy is believed to take the form of either a cosmological constant or some type of scalar field (Peebles & Ratra 1988).

The cosmological constant was originally introduced by Albert Einstein to achieve a static Universe with general relativity (Einstein 1917). He later abandoned this idea after Hubble's discovery that the Universe was expanding. More recently, Λ has been reintroduced and is currently believed to be the most probable explanation for the accelerated expansion of the Universe and the

missing matter density (Riess *et al.* 1998).

Adopting the cosmological constant solves several problems but presents a new question in understanding the source of this constant. One possible solution comes in the form of the vacuum energy. In empty space (*i.e.* devoid of any particles) one would assume the energy to be zero. This, however, would require us to know exactly the energy at a given time, which would violate the uncertainty principle, $\Delta E \Delta t \geq \hbar/2$. Therefore, in realty, empty space must be full of virtual particles that exist in particle-antiparticle pairs for extremely short amounts of time. The total energy of these particles in called the vacuum energy.

Although we are not certain as to the true nature of Λ, we are sure about some of the properties it must have. We observe that energy density in the Universe today is very close to critical. From equation 1.24 we can see that the condition for constant energy density can only be satisfied if $P = -\rho$ (*i.e.* negative pressure). We can thus define the equation of state parameter w as

$$w \equiv \frac{P}{\rho}, \tag{1.38}$$

(Dodelson 2003). From 1.38 we can see that for a cosmological constant we require $w = -1$, for matter we require $w = 0$ and for radiation we require $w = 1/3$. Using the relation $\rho_\Lambda \propto a^{-3(1+w)}$, we can constrain the value of Ω_Λ by constraining w.

1.2.3 ΛCDM

The combination of a cosmological constant with cold dark matter (ΛCDM) is known as the concordance model because, at present, it offers a picture the Universe that is consistent with observations.

Observations of the cosmic microwave background by the Wilkinson Microwave Anisotropy Probe (WMAP) provide the current constraints on the ΛCDM model parameters (Hinshaw *et al.* 2009). The basic results for WMAP 5-year data release are listed in table 1.1 and fig.1.4.

1.2.4 Distance Measures

With the knowledge that we live in an expanding universe, measuring the distances to distant objects becomes very complicated. To this end, we employ various cosmological distance measures that account for this fact. The following subsections list some of the methods commonly used to measure the distance between points on cosmological scales as described in Hogg (1999).

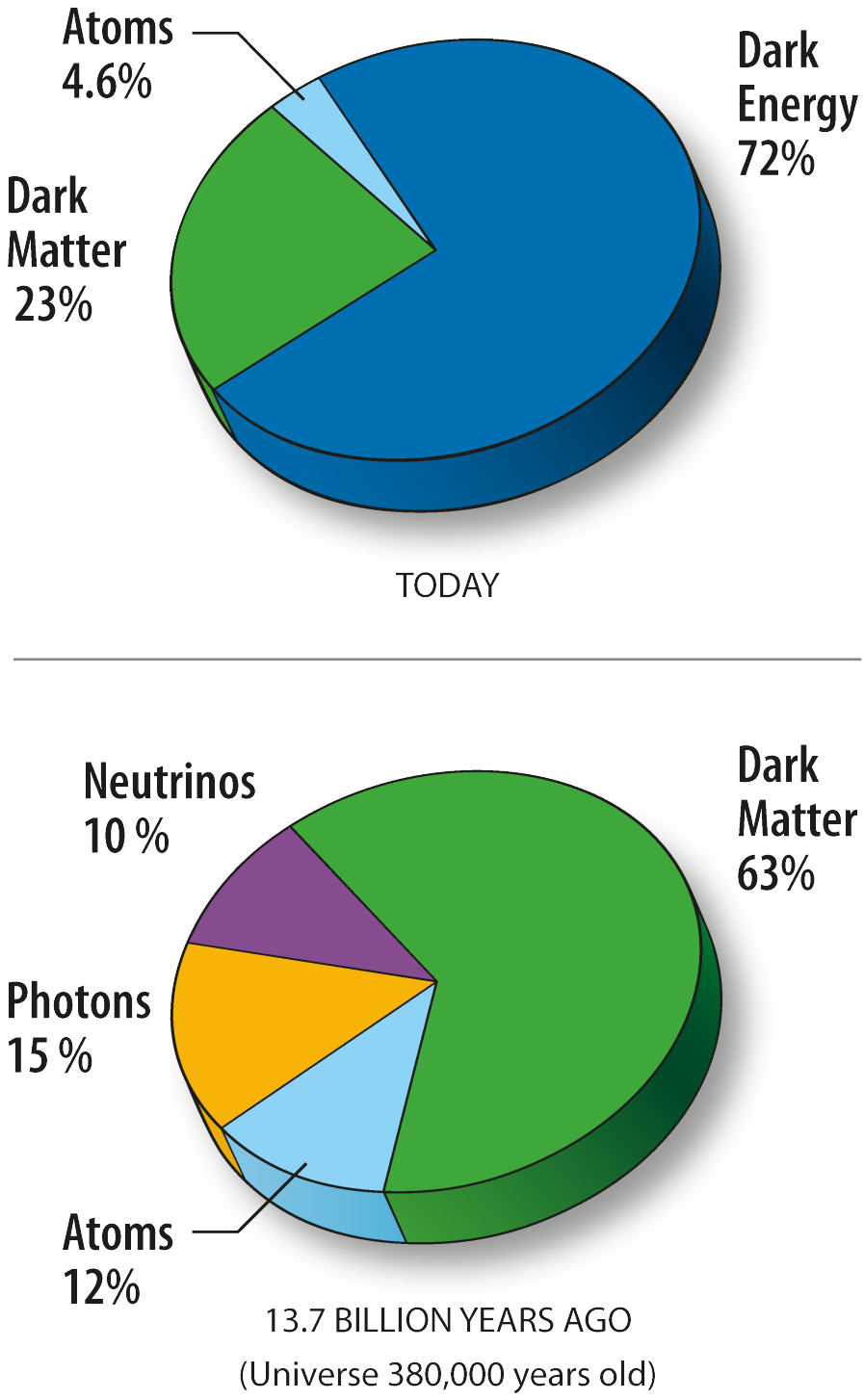


Figure 1.4: Pie charts of the constituents of the Universe today (top panel) and 13.7 Gyr ago (bottom panel) based on WMAP data.

Table 1.1: WMAP 5-year Data Results (Hinshaw *et al.* 2009). t_0 is the age of the Universe and z_* is the redshift of decoupling.

Parameter	Best Fit (WMAP)	Best Fit (WMAP+SNe+BAO)
H_0 (kmMpc^{-1}s^{-1})	$71.9^{+2.6}_{-2.7}$	70.5±1.3
σ_8	0.796±0.036	0.812±0.026
Ω_b	0.0441±0.003	0.0456±0.0015
Ω_c	0.214±0.027	0.228±0.013
Ω_Λ	0.742±0.030	0.726±0.015
Ω	$1.099^{+0.100}_{-0.085}$	$1.0050^{+0.0060}_{-0.0061}$
t_0 (Gyr)	13.69±0.13	13.72±0.12
z_*	1090.51±0.95	1090.88±0.72

1.2.4.1 Comoving Distance

We define the line-of-sight comoving distance, D_C, as the distance between two objects that remains constant with the expansion of the Universe. This distance can be expressed as

$$D_C = \frac{c}{H_0}\int_o^z \frac{dz'}{\sqrt{\Omega_M(1+z)^3+\Omega_k(1+z)^2+\Omega_\Lambda}}. \tag{1.39}$$

We define the transverse comoving distance, D_M, as the distance separating two objects at the same redshift but separated on the sky. This distance can be expressed as

$$D_M = \begin{cases} \frac{c}{H_0\sqrt{\Omega_k}}\sinh[\sqrt{\Omega_k}\frac{D_C H_0}{c}] & \text{for } \Omega_k > 0 \\ D_C & \text{for } \Omega_k = 0 \\ \frac{c}{H_0\sqrt{\Omega_k}}\sin[\sqrt{\mid \Omega_k \mid}\frac{D_C H_0}{c}] & \text{for } \Omega_k < 0 \end{cases}. \tag{1.40}$$

1.2.4.2 Comoving Volume

We define the comoving volume, V_C, as the volume measure in which the number density of objects remains constant with redshift. We can express the comoving volume element, dV_C, as

$$dV_C = \frac{c(1+z)^2 D_A^2}{H_0\sqrt{\Omega_M(1+z)^3+\Omega_k(1+z)^2+\Omega_\Lambda}} d\Omega dz, \tag{1.41}$$

where $d\Omega$ is the solid angle element. For $\Omega_k = 0$, we can see that

$$V_C = \frac{4\pi}{3} D_M^3. \tag{1.42}$$

1.2.4.3 Angular Diameter Distance

We define angular diameter distance, d_A, as ratio of the transverse physical size of an object to its angular size in radians,

$$d_A = \frac{l}{\theta} = \frac{D_M}{1+z}. \tag{1.43}$$

where l is the size of an object and θ is the angle it subtends on the sky.

1.2.4.4 Luminosity Distance

The luminosity distance, D_{lum}, is defined as

$$D_{\text{lum}} \equiv \sqrt{\frac{L}{4\pi S}}, \tag{1.44}$$

where L is the bolometric luminosity and S is the bolometric flux. D_{lum} can also be expressed as a function of comoving distance or angular diameter distance,

$$D_{\text{lum}} = (1+z)D_M = (1+z)^2 d_A. \tag{1.45}$$

1.3 Growth of Structure

Having discussed how the Universe as a whole grows and evolves, we can now turn out attention to the formation of the structure we observe within the Universe. This section discusses the mechanisms by which the initial perturbations in the density field grow, starting with a brief look at inflation (§1.3.1), followed by a description of the linear model of growth (§1.3.2) and finishing with look at the non-linear regime (§1.3.3).

1.3.1 Inflation

In 1981, American physicist Alan Guth proposed a theory in which the early Universe underwent a short period of accelerated expansion or inflation (Guth 1981). It is predicted that this inflationary period would have lasted some where between 10^{-36}-10^{-34}s and in this time the scale factor, a, would have increased by a factor of as much as 10^{50}. Guth (1981) attributes the source of inflation to a scalar field that behaves like the vacuum energy (see §1.2.2.6) under certain circumstances, which, if it had a much higher energy density than it does today, would cause the Universe to expand exponentially. This theory addresses several key problems in cosmology, several of which will be discussed in the following subsections.

1.3.1.1 Flatness Problem

Current WMAP observations (see §1.2.3) show that we live in a Universe with $\Omega \simeq 1$, which implies at very early times (*e.g.* $t < 10^{-6}$s) $\Omega \simeq 1$ also. The problem we have is understanding why immediately after the Big Bang the Universe would be so flat.

From equation 1.30, which shows that $\Omega_k \propto a^{-2}$, we see that inflation offers a rather simple solution. Increasing the scale factor by 10^{50} would make $\Omega_k \approx 0$ (Jones *et al.* 2006). Therefore, inflation essentially flattens the Universe regardless of the initial value of Ω_k.

1.3.1.2 Horizon Problem

Measurements of the CMB temperature show that it is uniform to one part in 10^5. The problem that arises from this is that regions on the sky more than two degrees apart were separated by a distance greater than the horizon distance at the time of last scattering (Jones *et al.* 2006). The horizon distance is the distance a signal could travel through space in the time that the Universe has existed and the time of last scattering or decoupling is the moment at which the expansion of the Universe overtook the rate of Compton scattering producing the CMB we observe today. The question we are left with is understanding how regions that have never been in contact are so uniform.

Inflation also provides a solution to this problem in suggesting that the regions we observe today that appear homogenous were inflated beyond the horizon distance. Before inflation these regions would have been in contact and thus been able to reach thermal equilibrium.

The comoving horizon is defined by

$$\eta = \int_0^a \frac{da'}{a'} \frac{1}{a'H(a')}, \tag{1.46}$$

where $\frac{1}{aH}$ is the distance particles can move in the period of one expansion and is called the comoving Hubble radius. Most inflationary models operate at energy scales $\simeq 10^{15}$ GeV, which means that the scale factor at the end of inflation is given by $a_e \simeq T_0/10^{15}$ GeV $\simeq 10^{-28}$. The number of time intervals in which the scale factor can grow exponentially by a factor of e, or e-folds is given by

$$N = \ln\Big(\frac{a_2}{a_1}\Big), \tag{1.47}$$

which for $a_e \simeq 10^{-28}$ corresponds to $N \simeq 60$ e-folds.

1.3.1.3 Monopole Problem

Grand Unified Theories (GUTs) are theories that attempt to unify electromagnetism, the strong nuclear force and the weak nuclear force in the early Universe. One of the predictions of GUTs is that at high temperatures heavy stable particles called magnetic monopoles should be produced in large quantities. The problem that arises is that these particles have never been detected and the standard Big Bang model cannot account for this fact.

Through inflation magnetic monopoles would be separated from each other as the Universe expands making them extremely difficult to detect, potentially explaining the observed discrepancy with GUTs.

1.3.1.4 Perturbation Problem

From the measurements of the CMB, we have learned that the early Universe was very smooth with initial perturbations of only one part in 10^5. These perturbations are thought to be the seeds of structure formation. What we need to understand is what the source of the perturbations is and how they grew large enough for structure to form.

The proposed answer is that quantum fluctuations that occurred during inflation would have grown until they surpassed the horizon scale at which point the fluctuation would no longer be able smooth itself from one side to the other. This would cause perturbations to become 'frozen in', while the Universe continued to grow (Jones *et al.* 2006). After inflation, the expansion rate of the Universe slowed dramatically allowing the 'frozen in' density perturbations to re-enter the horizon. Interestingly, the resulting density pattern from this process would give rise to the observed range of clusters and superclusters.

This prediction of inflation is of most interest for this section as it provides a starting place from which we can investigate the growth of structure in the Universe.

1.3.2 The Linear Regime

After the transition from a radiation dominated Universe to a matter dominated Universe we can begin to model the matter content of the Universe as a fluid. This assumption allows us to trace the evolution of the primordial density perturbations left over from inflation.

1.3.2.1 Growth of Linear Perturbations

We model the Universe as a fluid with density $\rho(x,t)$ and pressure $P \ll \rho$ (Liddle & Lyth 2000). An element of gas then has position $r(t) = a(t)x(t)$ and velocity $u = \dot{r}$. The acceleration of the element is provided by the Euler equation

$$\Big(\frac{\partial}{\partial t} + u \cdot \bigtriangledown\Big)u = -\frac{1}{\rho} \bigtriangledown P - \bigtriangledown\Phi, \tag{1.48}$$

where Φ is the gravitational potential, which satisfies the Poisson equation,

$$\bigtriangledown^2\Phi = 4\pi \mathrm{G}\rho. \tag{1.49}$$

In spherical polar coordinates 1.49 becomes

$$\frac{1}{r^2}\frac{\partial}{\partial r}\Big(r^2\frac{\partial\Phi}{\partial r}\Big) = 4\pi \mathrm{G}\rho. \tag{1.50}$$

We can define the differential form of of the continuity equation as

$$\frac{\partial\rho}{\partial t} + (\bigtriangledown \cdot u)\rho = 0. \tag{1.51}$$

In an unperturbed Universe starting from $x = 0$, from 1.50 we find

$$\Phi = \frac{2}{3}\pi \mathrm{G}\rho r^2. \tag{1.52}$$

With $P = 0$ and using 1.51, 1.52 and 1.23, 1.48 gives

$$\frac{\partial u}{\partial t} = -\frac{4\pi}{3}\mathrm{G}\rho r = \frac{\ddot{a}}{a}r. \tag{1.53}$$

This corresponds to a universe of pressureless matter (*i.e.* dark matter) undergoing Hubble expansion. We can introduce small linear perturbations with

$$\rho(x,t) = \rho(t) + \delta\rho(x,t), \tag{1.54}$$

$$P(x,t) = P(t) + \delta P(x,t), \tag{1.55}$$

$$\Phi(x,t) = \Phi(t) + \delta\Phi(x,t), \tag{1.56}$$

and

$$u(x,t) = u(t) + v = H(t)r + v. \tag{1.57}$$

where v is the peculiar velocity. Introducing these perturbations into 1.48, 1.49 and 1.51 gives

$$\Big[\frac{\partial}{\partial t} + (u+v)\cdot\bigtriangledown\Big](u+v) = -\frac{1}{\rho}\bigtriangledown(P+\delta P) - \bigtriangledown(\Phi + \delta\Phi), \tag{1.58}$$

$$\bigtriangledown^2(\Phi + \delta\Phi) = 4\pi \mathrm{G}(\rho + \delta\rho), \tag{1.59}$$

and

$$\frac{\partial(\rho + \delta\rho)}{\partial t} + [\bigtriangledown \cdot (u + v)](\rho + \delta\rho) = 0. \tag{1.60}$$

Subtracting the unperturbed quantities and ignoring the products of perturbations (*e.g.* $\delta x \cdot \bigtriangledown x$), we can see that the perturbed quantities satisfy

$$\dot{v} + Hv = - \bigtriangledown \delta\Phi - \frac{1}{\rho} \bigtriangledown \delta P, \tag{1.61}$$

$$\bigtriangledown^2 \delta\Phi = 4\pi \mathrm{G}\delta\rho, \tag{1.62}$$

and

$$\dot{\delta\rho} = -(\bigtriangledown \cdot v)\rho - (\bigtriangledown \cdot u)\delta\rho. \tag{1.63}$$

We can define the density contrast as $\delta(x) = \delta\rho/\rho$ and using the reciprocal rule, $\frac{d}{dx}\left(\frac{1}{f(x)}\right) = -\frac{1}{f(x)^2}\frac{df(x)}{dx}$, 1.63 becomes

$$\dot{\delta}(x) = -(\bigtriangledown \cdot v). \tag{1.64}$$

Since we are dealing with an expanding Universe, it is convenient to use comoving coordinates to express these equations. In comoving coordinates $x = a(t)r$ and therefore

$$\dot{x} = \dot{a}r + a\dot{r}. \tag{1.65}$$

Comparing this with 1.57 shows that

$$u(t) = \dot{a}r, \tag{1.66}$$

and

$$v = a\dot{r} = au. \tag{1.67}$$

Inserting 1.67 into 1.61 a using $\bigtriangledown = \bigtriangledown_c a^{-1}$ gives

$$\dot{u} + 2Hu = -\frac{1}{a^2} \bigtriangledown_c \delta\Phi - \frac{1}{a^2\rho} \bigtriangledown_c \delta P. \tag{1.68}$$

Inserting 1.67 into 1.64 gives

$$\dot{\delta}(x) = -(\bigtriangledown_c \cdot u), \tag{1.69}$$

and 1.62 becomes

$$\bigtriangledown_c^2 \delta\Phi = 4\pi \mathrm{G}\delta\rho a^2. \tag{1.70}$$

Taking the divergence of 1.68 and the derivative of 1.69 gives

$$\bigtriangledown_c \cdot \dot{u} + 2H(\bigtriangledown_c \cdot u) = -\frac{1}{a^2}\bigtriangledown_c^2 \delta\Phi - \frac{c_s^2}{a^2\rho}\bigtriangledown_c^2 \delta\rho, \tag{1.71}$$

where $c_s^2 = \delta P/\delta\rho$ is the adiabatic sound speed, and

$$\ddot{\delta}(x) = -(\bigtriangledown_c \cdot \dot{u}). \tag{1.72}$$

Combining 1.69, 1.70, 1.71 and 1.72 gives

$$\ddot{\delta}(x) + 2H\dot{\delta}(x) = 4\pi \mathrm{G}\delta\rho + \frac{c_s^2}{a^2\rho}\bigtriangledown_c^2 \delta\rho. \tag{1.73}$$

Taking a wave solution for $\delta(x)$ of the form $\delta(x) \propto \exp i(k \cdot r - \omega t)$, where k is the comoving wavevector, $\bigtriangledown^2\rho/\rho = -k^2$ and 1.73 becomes

$$\ddot{\delta}(x) + 2H\dot{\delta}(x) = \delta(x)(4\pi \mathrm{G}\rho - k^2 c_s^2). \tag{1.74}$$

This equation is essential in the study of structure formation as it describes the growth of small density perturbations.

1.3.2.2 Jeans' Instability

In 1902, British astronomer James Hopwood Jeans determined the critical radius at which the thermal energy in an interstellar dust cloud is countered by gravity, causing the cloud to collapse (Jeans 1902). This critical radius is known as the Jeans' length, $\lambda_{\mathrm{J}} = \sqrt{15k_bT/4\pi \mathrm{G}\mu\rho}$. Jeans' length also gives us a measure of the maximum scale for stable fluctuations at a given epoch, $\lambda_{\mathrm{J}} = cs(\pi/\mathrm{G}\rho)^{\frac{1}{2}}$, where c_s is the sound speed (Longair 1998).

For wavelengths smaller than λ_{J}, perturbations are sound waves with

$$\omega = \left[4\pi \mathrm{G}\rho_0\left(\frac{\lambda_{\mathrm{J}}^2}{\lambda}\right) - 1\right]^{\frac{1}{2}}. \tag{1.75}$$

For wavelengths greater than λ_{J}, the perturbations are unstable and for instabilities with $\lambda \gg \lambda_{\mathrm{J}}$, we can ignore the pressure term in equation 1.74, which gives

$$\ddot{\delta}(x) + 2H\dot{\delta}(x) = 4\pi \mathrm{G}\rho\delta(x). \tag{1.76}$$

which can also be expressed as

$$\ddot{\delta}(x) + 2H\dot{\delta}(x) = \frac{3\Omega_0 H_0^2}{2a^3}\delta(x), \tag{1.77}$$

where

$$\dot{a}^2 = H_0\Big[\Omega_0\Big(\frac{1}{a} - 1\Big) + \Omega_\Lambda(a^2 - 1) + 1\Big], \tag{1.78}$$

(Longair 1998).

The general solution to equation 1.77 is of the form

$$\delta(x,t) = A(x)\delta_g(t) + B(x)\delta_d(t), \tag{1.79}$$

where A and B are arbitrary functions of x, and $\delta_g(t)$ and $\delta_d(t)$ correspond to growing and decaying modes respectively with

$$\delta_g \propto H(t)\int_0^a \dot{a}^{-3}da, \tag{1.80}$$

and

$$\delta_d \propto H(t) \tag{1.81}$$

(Heath 1977; Carroll *et al.* 1992).

For a Universe with $\Omega_0 = 1$, $4\pi\mathrm{G}\rho = 2/3t$ and $H = 2/3t$, 1.76 becomes

$$\ddot{\delta}(x) + \frac{4}{3t}\dot{\delta}(x) - \frac{2}{3t^2}\delta(x) = 0, \tag{1.82}$$

which has a solution of the form $\delta(x) = at^n$. Substituting this into 1.82 gives

$$n(n-1) + \frac{4}{3}n - \frac{2}{3} = 0, \tag{1.83}$$

which has solutions $n = 2/3$ (growing mode) and $n = -1$ (decaying mode) (Longair 1998). For the growing mode we can see that

$$\delta(x) \propto t^{\frac{2}{3}} \propto a(t), \tag{1.84}$$

which implies that the density contrast scales with the Universe.

1.3.2.3 Baryon Acoustic Oscillations

When the pressure term dominates over gravity and $c_s = \frac{k_B T_0}{m_p a}$ (adiabatic sound speed for an ideal gas), equation 1.74 becomes

$$\ddot{\delta}(x) + 2H\dot{\delta}(x) + \Big(\frac{k_B T_0}{m_p}\Big)\Big(\frac{k^2}{a^3}\Big)\delta(x) \cong 0 \tag{1.85}$$

where k_B is the Boltzmann constant ($k_B = 1.381 \times 10^{-23}$ JK^{-1}) and m_p is the mass of the proton ($m_p = 1.673 \times 10^{-27}$ kg). 1.85 has solutions of the form

$$\delta(x) = t^{-\frac{1}{6}} e^{(\pm \frac{i\omega}{3} \ln t)} \tag{1.86}$$

where $\omega = \frac{4k^2}{H^2 a^3}\Big(\frac{k_B T_0}{m_p}\Big) \gg 1$.

Equation 1.86 shows that $\delta(x)$ scales as $t^{-\frac{1}{6}}$, which implies that oscillations in the baryon density can be thought of as acoustic vibrations in a medium (Padmanabhan 1996).

After decoupling, these baryon acoustic oscillations (BAO) give rise to shells of baryonic matter at a fixed radius, know as the sound horizon. CMB measurements provide the size of the sound horizon at recombination to good accuracy and measurements of the clustering of galaxies can be used to provide the present size of the sound horizon (Percival *et al.* 2010; Reid *et al.* 2010). In this way, BAO can be considered a standard ruler in cosmology, which can be used to gain a better understanding of the accelerated expansion of the Universe.

1.3.3 The Nonlinear Regime

Thus far, we have examined the growth of perturbations with a linear approximation, however the equations of motion are nonlinear in nature (Peacock 1999). The full treatment of the growth of perturbations or gravitational instabilities is more complicated, therefore we require more sophisticated techniques to model them. There are some simplifications that can be made for certain cases, which will be described in this section.

1.3.3.1 Spherical Collapse

An overdense sphere is a useful example for exploring the nonlinear regime, as it behaves in the same way as a closed universe.

A gravitating sphere of mass, M, and radius, R, can be described by the Newtonian equation,

$$\ddot{R} = -\frac{\mathrm{G}M(R)}{R^2}. \tag{1.87}$$

Integrating equation 1.87 and assuming that mass does not change as a function of time gives

$$\dot{R}^2 = \frac{2\mathrm{G}M(R)}{R} + \mathrm{C}, \tag{1.88}$$

where C is a constant that determines the total energy of the sphere. If C> 0, the sphere will expand forever, while if C< 0, the sphere expand for a period and at a certain value of R will begin to contract. For collapse, the solution for the C< 0 case is the cycloid,

$$t = \mathrm{A}(\theta - sin\theta), \tag{1.89}$$

and

$$R = \mathrm{B}(1 - cos\theta), \tag{1.90}$$

where θ is the development angle ($0 \leq \theta \leq 2\pi$), and A and B are constants. We define turnaround at $\theta = \pi$. This is the point at which the perturbation reaches its maximum expansion. Equations 1.89 & 1.90 can be expressed in terms of the turnaround time, t_{max}, and the maximum scale factor, a_{max}, which define the point at which expansion ceases and collapse begins via

$$\frac{t}{t_{\mathrm{max}}} = \frac{1}{\pi}(\theta - sin\theta), \tag{1.91}$$

and

$$\frac{a(t)}{a_{\mathrm{max}}} = \frac{1}{2}(1 - cos\theta). \tag{1.92}$$

Fig.1.5 shows the evolution of three different scale factors (Liddle & Lyth 2000).

Expanding equations 1.91 & 1.92 to second order gives

$$\frac{t}{t_{\mathrm{max}}} = \frac{1}{\pi}\Big(\frac{\theta^3}{6} - \frac{\theta^5}{120}\Big), \tag{1.93}$$

and

$$\frac{a(t)}{a_{\mathrm{max}}} = \frac{\theta^2}{4} - \frac{\theta^4}{48}. \tag{1.94}$$

Taking the leading order term of 1.93 gives

$$\theta^2 = \Big(\frac{6\pi t}{t_{\mathrm{max}}}\Big)^{\frac{2}{3}}. \tag{1.95}$$

Substituting 1.95 back into 1.93 gives

$$\theta^2 \simeq \Big(\frac{6\pi t}{t_{\mathrm{max}}}\Big)^{\frac{2}{3}}\Big[1 + \frac{1}{30}\Big(\frac{6\pi t}{t_{\mathrm{max}}}\Big)^{\frac{2}{3}}\Big]. \tag{1.96}$$

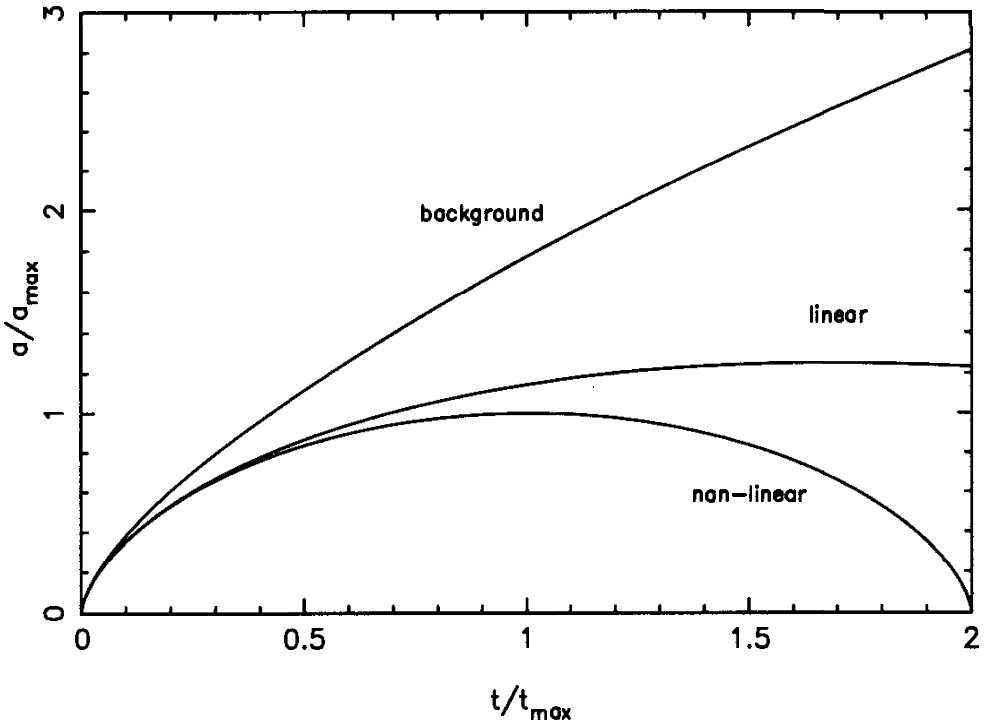


Figure 1.5: The evolution of three different scales factors, the background, linear evolution and non-linear collapse (Liddle & Lyth 2000).

Substituting 1.96 back into 1.94 gives the linearised scale factor

$$\frac{a_{\rm lin}(t)}{a_{\rm max}} \simeq \frac{1}{4}\Big(\frac{6\pi t}{t_{\rm max}}\Big)^{\frac{2}{3}}\Big[1-\frac{1}{20}\Big(\frac{6\pi t}{t_{\rm max}}\Big)^{\frac{2}{3}}\Big]. \tag{1.97}$$

The linear density contrast, $\delta_{\rm lin}$, is defined as

$$\delta_{\rm lin} = \frac{\rho-\bar{\rho}}{\bar{\rho}} = \Big(\frac{a_{\rm back}}{a_{\rm lin}}\Big)^3 - 1, \tag{1.98}$$

where $a_{\rm back}$ is the background evolution and

$$\frac{a_{\rm back}(t)}{a_{\rm max}} \simeq \frac{1}{4}\Big(\frac{6\pi t}{t_{\rm max}}\Big)^{\frac{2}{3}}. \tag{1.99}$$

Therefore dividing 1.97 by 1.99 gives

$$\Big(\frac{a_{\rm back}}{a_{\rm lin}}\Big)^3 \simeq 1+\frac{3}{20}\Big(\frac{6\pi t}{t_{\rm max}}\Big)^{\frac{2}{3}}. \tag{1.100}$$

Substituting 1.100 into 1.98 gives

$$\delta_{\rm lin} = \frac{3}{20}\Big(\frac{6\pi t}{t_{\rm max}}\Big)^{\frac{2}{3}}. \tag{1.101}$$

We define turnaround as the point at which $t = t_{\rm max}$. We can then see that the density contrast at turnaround is

$$\delta_{\rm lin}^{\rm turn} = \frac{3}{20}(6\pi)^{\frac{2}{3}} = 1.06. \tag{1.102}$$

If turnaround occurs at $t = t_{\rm max}$ then collapse occurs around $t = 2t_{\rm max}$. We find the density contrast at collapse is

$$\delta_{\rm lin}^{\rm coll} = \delta_c = \frac{3}{20}(12\pi)^{\frac{2}{3}} = 1.686. \tag{1.103}$$

This value sets the perturbation threshold for the collapse of structure in the density field.

1.3.3.2 The Virial Theorem

In reality, density perturbations will not be spherical. In addition, full spherical collapse will never occur because the kinetic energy of the infalling particles will be converted into random motions (Peacock 1999). We can relate the potential energy of the collapsing particles to their kinetic energy through the virial theorem.

The virial theorem describes objects in dynamical equilibrium under gravity. We define the crossing time as

$$t_{cr} = \frac{R}{< v >}. \tag{1.104}$$

where R is the size of the system and $< v >$ is the velocity dispersion. We can define the acceleration of the i^{th} particle due to all other particles as

$$\ddot{r_i} = \sum_{j \neq i} \frac{{\rm G}m_j(r_j - r_i)}{|r_i - r_j|^3}, \tag{1.105}$$

(Longair 1998). Multiplying 1.105 by $m_i r_i$ gives

$$m_i r_i \ddot{r_i} = \sum_{j \neq i} \frac{{\rm G}m_i m_j r_i \cdot (r_j - r_i)}{|r_i - r_j|^3}. \tag{1.106}$$

Summing over i, 1.106 can be expressed as

$$\frac{1}{2}\frac{d^2}{dt^2}\sum_i m_i {r_i}^2 - \sum_i m_i \dot{r_i}^2 = -\frac{1}{2}\sum_{\substack{i,j \\ j \neq i}} \frac{{\rm G}m_i m_j}{|r_i - r_j|}. \tag{1.107}$$

We define the total kinetic energy of the particles, T, as

$${\rm T} = \frac{1}{2}\sum_i m_i \dot{r}_i^2, \tag{1.108}$$

and the gravitational potential energy of the system, U, as

$$\mathrm{U} = -\frac{1}{2}\sum_{\substack{i,j \\ j\neq i}} \frac{\mathrm{G}m_i m_j}{|r_i - r_j|}. \tag{1.109}$$

Substituting 1.108 and 1.109 into 1.107 gives

$$\frac{1}{2}\frac{d^2}{dt^2}\sum_i m_i {r_i}^2 = 2\mathrm{T} - |\mathrm{U}|. \tag{1.110}$$

We define the condition for statistical equilibrium as

$$\frac{1}{2}\frac{d^2}{dt^2}\sum_i m_i {r_i}^2 = 0. \tag{1.111}$$

Therefore the Virial Theorem is expressed as

$$\mathrm{T} = \frac{1}{2}|\mathrm{U}|. \tag{1.112}$$

If we assume the velocity distribution of galaxies is isotropic then

$$< v^2 >= 3 < v_\parallel^2 >, \tag{1.113}$$

where $v_\parallel$ is the velocity along the line of sight. If the velocity dispersion is independent of the mass of the galaxies, the total kinetic energy becomes

$$\mathrm{T} = \frac{1}{2}\sum_i m_i r_i^2 = \frac{3}{2}M < v_\parallel^2 >, \tag{1.114}$$

where M is the total mass of the system. If the system is assumed to be spherically symmetric

$$|\mathrm{U}| = \frac{\mathrm{G}M^2}{R_{\mathrm{cl}}}, \tag{1.115}$$

where R_{cl} is the weighted mean separation of galaxies. We can therefore define the Virial Mass as

$$M = \frac{3 < v_\parallel^2 > R_{\mathrm{cl}}}{\mathrm{G}}. \tag{1.116}$$

From equation 1.112 we can see that $R_{\mathrm{vir}} = R_{\mathrm{max}}/2$, which corresponds to $\theta_{\mathrm{vir}} = 3\pi/2$. We can therefore express the density contrast at virialisation as

$$1 + \delta_{\mathrm{vir}} = \frac{(6\pi)^2}{2} \simeq 178\Omega^{-0.7} \tag{1.117}$$

(Peacock 1999).

From this analysis, we can assume that linear theory applies until $\delta(x) \sim \delta_c$ at which point virialization occurs. δ_{vir} then distinguishes collapsed systems from those in the process of infall.

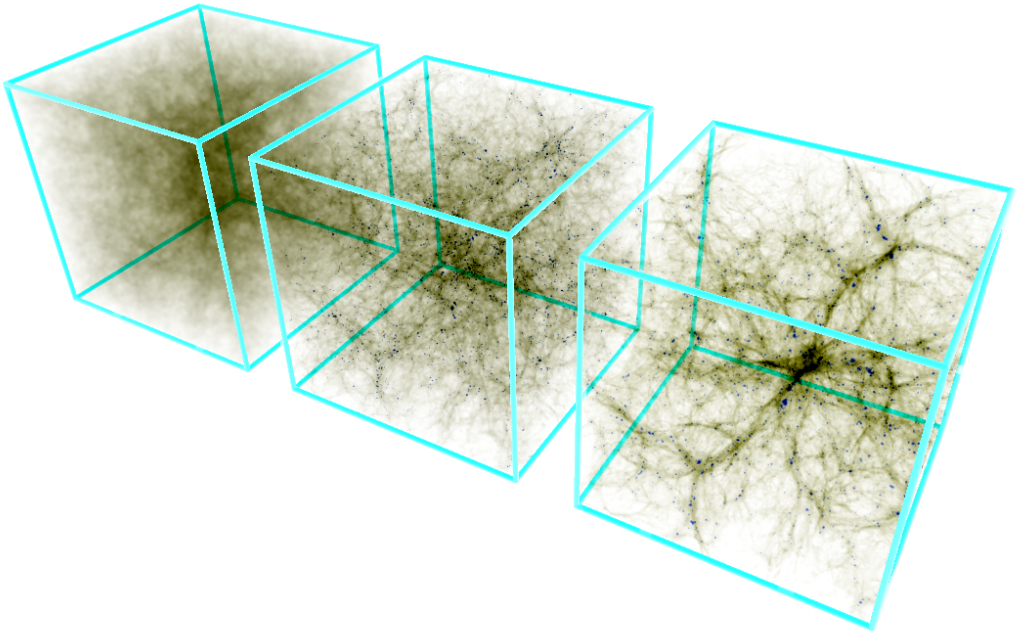


Figure 1.6: N-body simulation of gas and stars (Springel & Hernquist 2003).

1.3.3.3 N-body Simulations

The most complete treatment of the growth of perturbations is provided by N-body simulations. An N-body model simulates the effects of gravity and gas dynamics on a set of N discrete particles in a box of a given size (*e.g.* fig.1.6). This technique was pioneered by Holmberg (1941), who used 37 lightbulbs to represent galaxies. The first direct computational implementation of an N-body simulation was carried out by von Hoerner (1960).

The equation of motion for a given particle in the simulation is given by

$$\frac{d}{d\ln a}(a^2 u) = \sum F_k \exp(-ik \cdot x), \tag{1.118}$$

where

$$F_k = -ik\frac{3\Omega_m H a^2}{2k^2}\delta_k, \tag{1.119}$$

(Peacock 1999).

One of the most significant N-body simulations in recent years was the Millennium Run, which traced 2160^3 particles in a cubic region of the Universe over 2 billion light-years on a side (Springel *et al.* 2005).

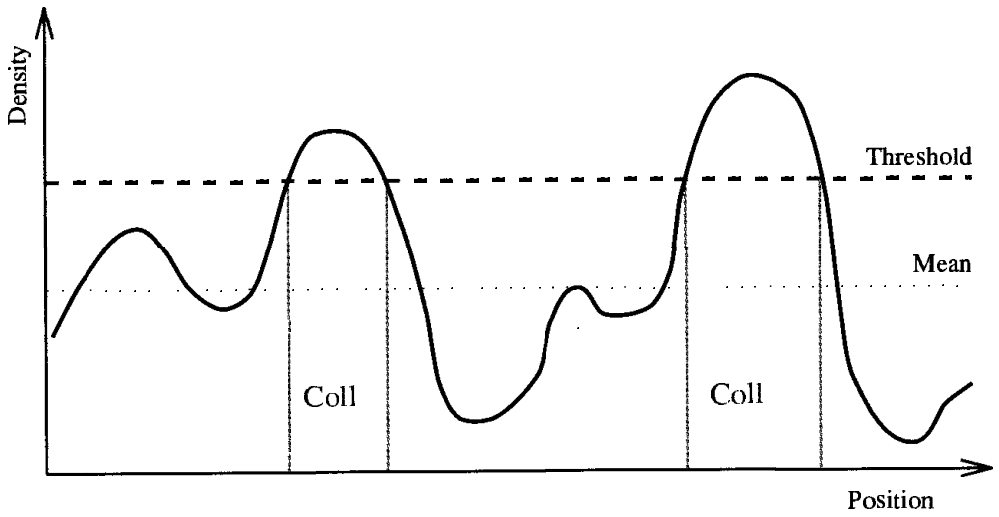


Figure 1.7: Schematic of a smoothed density field with collapsed regions (Liddle & Lyth 2000).

1.3.3.4 The Press-Schechter Formalism

In 1974, William H Press and Paul Schechter introduced a theory that is very successful at reproducing N-body simulations (Press & Schechter 1974). The theory states that if we smooth the density field in the linear regime on a mass scale M, the fraction of space with $\delta(x) > \delta_c$ will be in the form of collapsed objects of mass greater than M (Liddle & Lyth 2000).

If we assume primordial density perturbations are Gaussian fluctuations, then we can locate bound objects by smoothing (filtering) the initial linear density field. The probability that a point belongs to a collapsed object is

$$P(\delta > \delta_c|R_f) = \frac{1}{2}\Big[1 - \text{erf}\Big(\frac{\delta_c}{\sqrt{2}\sigma(R_f)}\Big)\Big], \tag{1.120}$$

where R_f is the characteristic length of the filter function and $\sigma(R_f)$ is the linear rms in the filtered version of δ. Fig.1.7 shows a schematic of a smoothed density field with collapsed regions.
The total probability from equation 1.120 is only equal to 1/2. Therefore Press & Schechter (1974) fudge in an extra factor of 2.

We define the fraction of haloes with mass greater than M as

$$F(> M) = 1 - \text{erf}\Big(\frac{\nu}{\sqrt{2}}\Big), \tag{1.121}$$

where $\nu = \delta_c/\sigma(M)$. The mass function is thus the comoving number density of haloes of mass

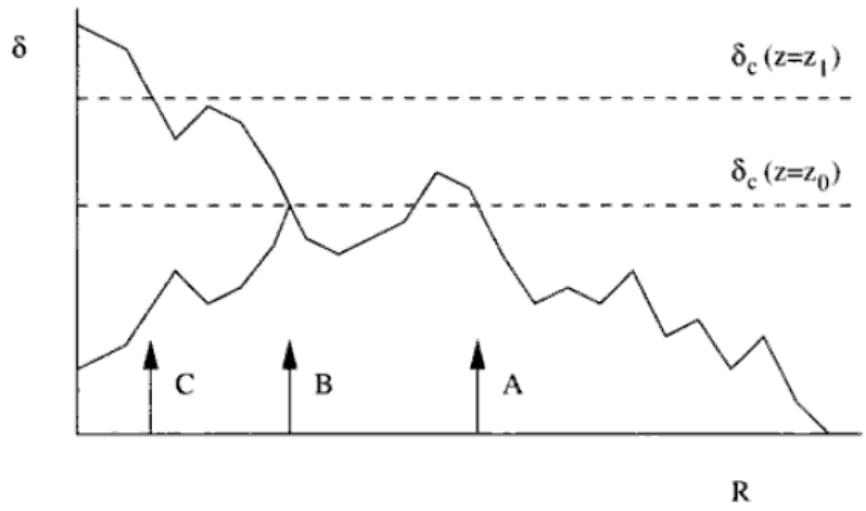


Figure 1.8: Random trajectory taken by filtered field δ and a fixed point, as a function of filtering radius R (Peacock 1999).

M,

$$n(M) = \frac{\bar{\rho}}{M}\Big|\frac{dF}{dM}\Big|, \tag{1.122}$$

where $\bar{\rho} = 3H_0^2\Omega_M/8\pi$G is the mean comoving density. Multiplying 1.122 by M leads to

$$n(M) = \frac{\bar{\rho}}{M^2}\Big|\frac{dF}{d\ln M}\Big| = \frac{\bar{\rho}}{M^2}\frac{d}{d\ln M}\Big[1 - \mathrm{erf}\Big(\frac{\nu}{\sqrt{2}}\Big)\Big], \tag{1.123}$$

which leads to the standard form of the mass function,

$$n(M) = \sqrt{\frac{2}{\pi}}\frac{\bar{\rho}}{M^2}\nu\Big|\frac{d\ln\sigma(M)}{d\ln M}\Big|\exp\Big(-\frac{\nu^2}{2}\Big). \tag{1.124}$$

In the limit of massive objects, the shape of the mass function is dominated by the exponential tail in equation 1.124. This implies that the mass function is exponentially sensitive to the choice of cosmological parameters, which makes it a powerful probe of cosmology.

If we examine the random trajectory of $\delta(R)$ in a sharp k-space filter (see fig.1.8), where R is the filtering radius, we can calculate the probability that the δ_c threshold has never be crossed via

$$\frac{dP_s}{d\delta} = \frac{1}{\sqrt{2\pi}\sigma}\Big[\exp\Big(-\frac{\delta^2}{2\sigma^2}\Big) - \exp\Big(-\frac{(\delta - 2\delta_c)^2}{2\sigma^2}\Big)\Big]. \tag{1.125}$$

Integrating 1.125 gives the probability that δ_c has been crossed at least once,

$$1 - P_s = 1 - \mathrm{erf}\Big(\frac{\delta_c}{\sqrt{2}\sigma}\Big), \tag{1.126}$$

which is twice the value of equation 1.120 (Peacock 1999). This provides an justification of the factor of 2 fudge originally introduced by (Press & Schechter 1974).

The Press & Schechter (1974) mass function has a tendency to over predict larges haloes and under predict smaller ones. Therefore, Sheth & Tormen (1999) introduced a function that better fits the observed mass distribution,

$$f_{\rm ST}(\nu) = A\sqrt{\frac{\alpha\nu^2}{2\pi}}\Big[1 + \frac{1}{(\alpha\nu^2)^p}\Big]\exp\Big(\frac{-\alpha\nu^2}{2}\Big), \tag{1.127}$$

where $\alpha = 0.707$, $p = 0.3$ and A is determined by requiring that

$$\int n_{\rm ST}(m)mdm = \bar{\rho} = 2\int_0^\infty \bar{\rho} f_{\rm ST}(\nu)d{\rm ln}\nu, \tag{1.128}$$

which leads to

$$A = \frac{1}{1 + \frac{\Gamma(0.5-p)}{2^p\pi}} \simeq 0.322184. \tag{1.129}$$

1.3.3.5 Dark Matter Haloes

The spatial distribution of dark matter, as predicted by N-body simulations, can be described by an NFW profile,

$$\rho(r|m) = \frac{\rho_{\rm crit}\delta_c}{\frac{r}{r_s}\Big(1 + \frac{r}{r_s}\Big)^2}, \tag{1.130}$$

where $\rho_{\rm crit} = \frac{3H^2}{8\pi{\rm G}}$ is the critical density, $r_s = \frac{r_{200}}{c}$, δ_c is the characteristic overdensity and c is the halo concentration. r_{200} is the radius at which $\rho = 200 \times \rho_{\rm crit}$ (Navarro *et al.* 1996). This profile shows that the density around a dark matter halo manly depends on its mass.

Spherical collapse shows that haloes form depending on the initial density field. Therefore, Mo & White (1996) have shown that haloes trace the dark matter field with bias, $b(M)$, given by

$$b(M) = 1 + \frac{\nu^2 - 1}{\delta_c}. \tag{1.131}$$

Equation 1.131 was later extended to a more general form by Sheth & Tormen (1999) to

$$b(M) = 1 + \frac{a\nu^2 - 1}{\delta_c} + \frac{2p}{\delta_c[1 + (a\nu^2)^p]}, \tag{1.132}$$

where $a = 0.707$ and $p = 0.3$ from numerical simulations.

1.4 Cosmological Probes

One of the fundamental objectives in cosmology is developing statistical tools for probing the various parameters that govern the state of the Universe. This section covers some of the techniques employed to help us constrain these parameters starting with a look at the correlation function and the power spectrum (§1.4.1), followed by an overview of the CMB (§1.4.2), then a brief description of gravitational lensing (§1.4.3), an introduction to photometric redshifts (§1.4.4) and finally a discussion about galaxy clusters (§1.4.5).

1.4.1 Correlation Function and Power Spectrum

The correlation function and power spectrum are helpful two-point statistics that describe the fluctuations in the density field in the Universe.

1.4.1.1 The Two-Point Correlation Function

One of the most useful tools for describing the clustering of galaxies in cosmology is the the two-point correlation function, ξ, which provides the excess probability of finding a galaxy at some distance r from another galaxy. This is a generally useful statistical tool that was originally applied to galactic data by *e.g.* Totsuji & Kihara (1969).

The two-point correlation function, or autocorrelation function when dealing with variations in the the density field, $\delta(x) = \rho(x)/\bar{\rho} - 1$, is defined as

$$\xi(r) = \langle \delta(x)\delta(x+r) \rangle. \tag{1.133}$$

In cosmology, $\xi(r)$ is defined by the probability of finding a pair of galaxies within the volume elements dV_1 and dV_2 according to

$$dP = \rho_0^2[1 + \xi(r)]dV_1 dV_2, \tag{1.134}$$

where ρ_0 is the average density of the sample (Peacock 1999).

There are several techniques for statistically estimating $\xi(r)$, each of which averages the counts of pairs of n galaxies at a given scale in a box of a given size (Martínez & Saar 2003). The problem that each of these techniques needs to address is the underestimation of the number of galaxy pairs for those galaxies near the boundary of the box. The simplest estimator that incorporates edge-correction is the natural estimator of Peebles & Hauser (1974),

$$\hat{\xi_n} = \frac{DD}{RR} - 1, \tag{1.135}$$

where $DD(r)$ is the number of pairs of galaxies and $RR(r)$ is the number of pairs of random points, each in the range $[r, r + dr]$. The standard estimator of Davis & Peebles (1983) was introduced to improve edge-correction with

$$\hat{\xi_s} = \frac{DD}{DR} - 1, \tag{1.136}$$

where $DR(r)$ is the number of pairs between the data and random samples.

Later estimators were introduced by Hamilton (1993),

$$\hat{\xi}_{\text{Ham}} = \frac{DD\ RR}{DR^2} - 1, \tag{1.137}$$

and Landy & Szalay (1993),

$$\hat{\xi}_{\text{LS}} = \frac{DD - 2DR + RR}{RR} - 1, \tag{1.138}$$

which only suffer from second-order bias effects.

Fig.1.9 shows and an example of the the two-point correlation function produced from SDSS data by Eisenstein *et al.* (2005) using the Landy & Szalay (1993) estimator. The bump at 100 h^{-1} Mpc is a detection of the baryon acoustic oscillation (BAO) peak (see §1.3.2.3). This illustrates one of the important uses of the correlation function as a cosmological probe.

1.4.1.2 Redshift-Space Distortions

When dealing with galaxy catalogues, it is convenient to separate the components of the distance between galaxy pairs, r, into radial and perpendicular components π and σ respectively. Thus, $\xi(\sigma, \pi)$ shows the correlation in redshift space as shown in fig.1.10. Examining the correlation function in this way enables one to identify the effects of redshift-space distortions, which are caused by the peculiar velocities of galaxies not associated to the Hubble flow. The two significant effects in redshift space are fingers of God (Jackson 1972), where long filamentary structures pointing along the line of sight are observed as a result of the random velocities of galaxies in clusters, and the Kaiser effect (Kaiser 1987), where a subtle flatting is observed in the correlation function due to the velocities of galaxies bound to a central mass undergoing infall. The Kaiser effect is a much smaller effect than fingers of God, but it can be distinguished by the fact that it occurs on larger scales.

1.4.1.3 The Power Spectrum

Another useful cosmological tool is the power spectrum, which describes the power of fluctuations on different scales in k-space as shown in fig.1.11. The power spectrum is directly related to the

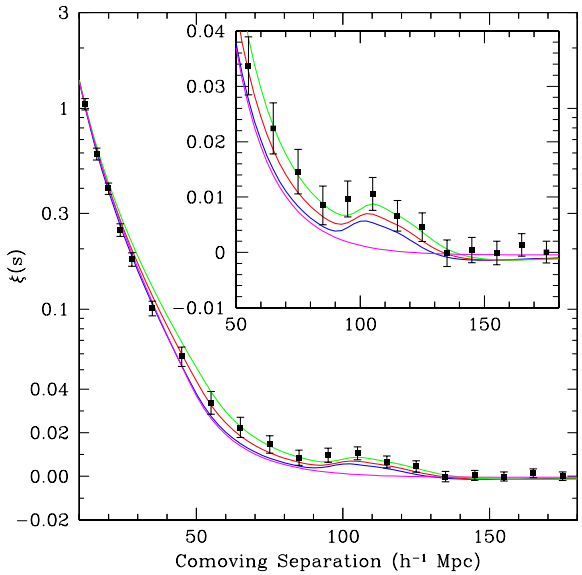


Figure 1.9: Large-scale redshift-space correlation function of the SDSS LRG sample. The error bars are from the diagonal elements of the mock-catalog covariance matrix; however, the points are correlated. The inset shows an expanded view with a linear vertical axis. The models are $\Omega_m h^2 = 0.12$ (top line), 0.13 (second line), and 0.14 (third line), all with $\Omega_b h^2 = 0.024$ and $n = 0.98$ and with a mild nonlinear prescription folded in. The bottom line shows a pure CDM model ($\Omega_m h^2 = 0.105$), which lacks the acoustic peak. $\xi(s)$ was computed using the Landy & Szalay (1993) estimator. The bump at 100 h^{-1} Mpc is the acoustic peak (Eisenstein *et al.* 2005).

correlation function.

The Fourier transform pair in n-dimensions for x are defined as

$$F(x) = \left(\frac{L}{2\pi}\right)^n \int F_k(k) e^{(-i\mathrm{k}\cdot\mathrm{x})} d^n k \tag{1.139}$$

and

$$F_k(k) = \left(\frac{1}{L}\right)^n \int F(x) e^{(i\mathrm{k}\cdot\mathrm{x})} d^n x, \tag{1.140}$$

(Peacock 1999).

The density field can be expressed as a sum via

$$\delta(x) = \sum_k \delta_k e^{(i\mathrm{k}\cdot\mathrm{x})} \tag{1.141}$$

and since $\delta(x)$ is real, the correlation function (equation 1.133) can be expressed as

$$\xi(r) = \left\langle \sum_k \sum_{k'} \delta_k \delta^*_{k'} e^{i(\mathrm{k'}-\mathrm{k})\cdot\mathrm{x}} e^{(i\mathrm{k}\cdot\mathrm{r})} \right\rangle \tag{1.142}$$

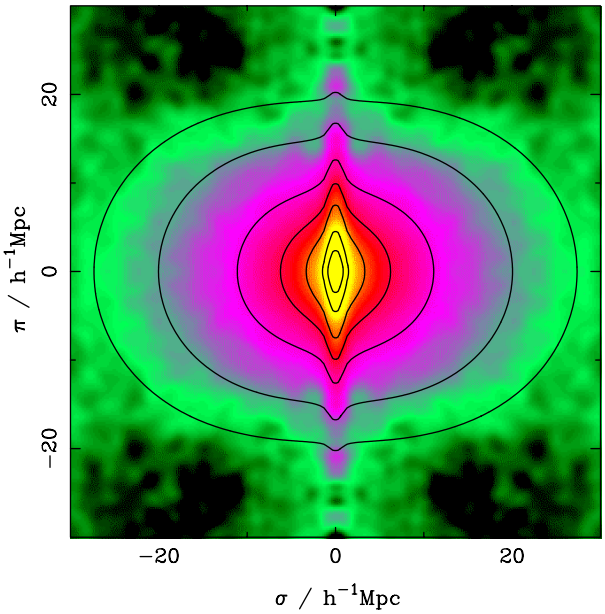


Figure 1.10: The redshift-space correlation function for the 2dFGRS, $\xi(\sigma, \pi)$, plotted as a function of transverse (σ) and radial (π) pair separation. The function was estimated by counting pairs in boxes of side 0.2 h^{-1} Mpc (assuming an Ω = 1 geometry), and then smoothing with a Gaussian of rms width $0.5h^{-1}$ Mpc. To illustrate deviations from circular symmetry, the data from the first quadrant are repeated with reflection in both axes. This plot clearly displays redshift distortions, with fingers of God elongations at small scales and the coherent Kaiser flattening at large radii (Peacock *et al.* 2001).

(Longair 1998). All cross terms with k'≠k average to zero and 1.142 becomes

$$\xi(r) = \sum |\delta_k|^2 e^{(i\mathbf{k}\cdot\mathbf{r})}, \tag{1.143}$$

which can be expressed in integral form as

$$\xi(r) = \frac{V}{2\pi^2} \int |\delta_k|^2 e^{(i\mathbf{k}\cdot\mathbf{r})} d^3k. \tag{1.144}$$

Since $\xi(r)$ is real and defining the power spectrum $P(k) = |\delta_k|^2$, 1.144 becomes

$$\xi(r) = \frac{V}{2\pi^2} \int P(k) \frac{\sin(kr)}{kr} 4\pi k^2 dk. \tag{1.145}$$

Comparing equations 1.139 and 1.145 shows that the power spectrum is the Fourier transform of the correlation function. Therefore from equation 1.140, the power spectrum is defined as

$$P(k) = \frac{1}{V} \int \xi(r) \frac{\sin(kr)}{kr} 4\pi r^2 dr. \tag{1.146}$$

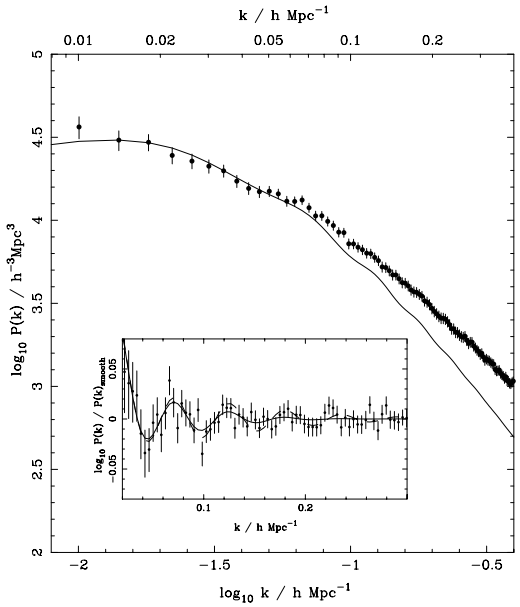


Figure 1.11: The redshift-space power spectrum recovered from the combined SDSS main galaxy and LRG sample, optimally weighted for both density changes and luminosity dependent bias (solid circles with 1-σ errors) (Percival *et al.* 2007).

1.4.2 Cosmic Microwave Background

Soon after the Big Bang, the Universe was in a hot dense state in which photons regularly collided with electrons. This process maintained the photons in a state of thermal equilibrium that corresponds to uniform blackbody radiation. Around 380,000 years later ($z = 1090.88 \pm 0.72$, Hinshaw *et al.* 2009) the Universe had cooled sufficiently for the photons to decouple from the electrons and travel freely through space. These photons left over from last scattering are generally referred to as the cosmic background radiation. At the present epoch this radiation corresponds to wavelengths in the range $10^{-4} \leq \lambda \leq 10^{-1}$m and therefore is most commonly known as the cosmic microwave background (CMB). The CMB radiation was first detected in 1965 by Arno Penzias and Robert Woodrow Wilson (Penzias & Wilson 1965).

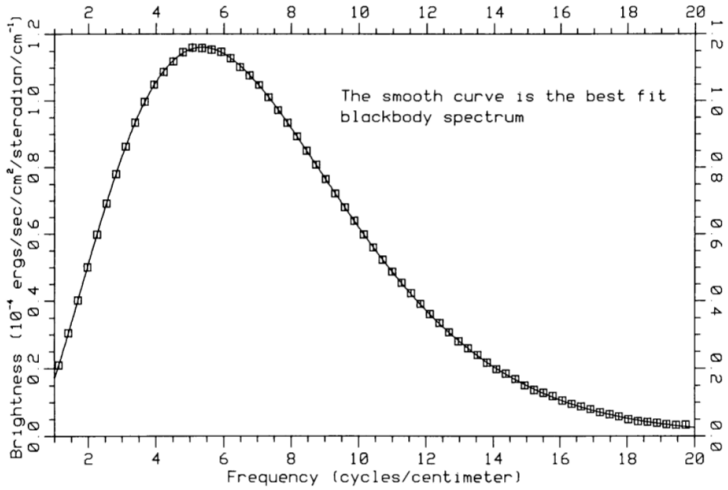


Figure 1.12: Preliminary spectrum of the cosmic microwave background from the FIRAS instrument at the north Galactic pole, compared to a blackbody. Boxes are measured points and show size of assumed 1% error band. The units for the vertical axis are 10^{-4} ergs s^{-1} cm^{-2} sr^{-1} cm (Mather *et al.* 1990).

1.4.2.1 CMB Temperature

The temperature of a blackbody spectrum can be related to the wavelength at which the maximum intensity is emitted through Wien's displacement law,

$$\lambda_{\text{max}} = \frac{2.898 \times 10^{-3}}{T}. \tag{1.147}$$

Therefore, from equation 1.7 we can see that temperature of the CMB is related to the scale factor via

$$T \propto \frac{1}{a(t)}. \tag{1.148}$$

Equation 1.148 indicates that the temperature of the CMB was much higher in the past than it is today.

Measuring the spectrum of the CMB is difficult using ground based instruments as it peaks around 1 mm and for wavelengths shorter than this atmospheric absorption becomes a major obstacle. The Cosmic Background Explorer (COBE) was a NASA space based mission launched in 1989 to measure the CMB across the whole sky using the DIRBE (Diffuse InfraRed Background Experiment), DMR (Differential Microwave Radiometer) and FIRAS (Far InfraRed Absolute Spectrometer) instruments (Boggess *et al.* 1992). The COBE FIRAS results revealed a near perfect blackbody spectrum with temperature $T = 2.725$ K as shown in fig.1.12 (Mather *et al.* 1990).

1.4.2.2 CMB Anisotropies

As mentioned in earlier sections the CMB appears smooth to one part in 10^5, however the tiny anisotropies that can be observed below this scale are of the most interest for cosmology.

There are several sources for the anisotropies seen in the CMB temperature, which can be divided into two classes: those that occurred at the time of last scattering or primary anisotropies and those that occurred after the time of last scattering or secondary anisotropies. Primary anisotropies include acoustic oscillations (see §1.3.2.3), the Sachs-Wolfe effect (Sachs & Wolfe 1967), which is an effect whereby CMB photons are gravitationally redshifted, and Silk damping (Silk 1968), which is an effect whereby acoustic oscillations with short wavelengths are effectively smoothed out. The secondary anisotropies included effects such as the integrated Sachs-Wolfe (ISW) effect (Hu & Dodelson 2002; Scranton *et al.* 2003), which is the result of gravitational shifting of CMB photons after last scattering, and the Sunyaev-Zeldovich (SZ) effect (see §1.4.5.4).

CMB anisotropies were observed by COBE (Bennett *et al.* 1996), but improved resolution from the Wilkinson Microwave Anisotropy Probe (WMAP) revealed a much richer distribution of temperature fluctuations as seen in fig.1.13 (Bennett *et al.* 2003).

Analysis of the CMB anisotropies can be made by observing the angular power spectrum as shown in fig.1.14, which shows the best fit ΛCDM model (solid red line) and the cosmic variance expected for this model (grey shaded region) from Bennett *et al.* (2003). The x-axis in the angular power spectrum is the multipole number, l, where $l = \frac{180^\circ}{\theta}$, and y-axis is the angular power given by $l(l+1)C_l/2\pi$.

Cosmic variance or sample variance is the statistical uncertainty introduced into cosmological observations by the fact that we can only observe a fraction of the Universe at any given time, which is particularly problematic when making observations on scales approaching the horizon distance. Therefore, observations of the CMB power spectrum on large scales (low l) are dominated by sample variance as shown in fig.1.14.

The first peak seen in the CMB anisotropy power spectrum is believed to be the result of acoustic oscillations of the photon-baryon plasma in the Universe that occurred before decoupling.

The predicted shape of the angular power spectrum depends on cosmological parameters, which is significant because it means that measurements of the CMB angular power spectrum can essentially be used to constrain all the cosmological parameters simultaneously (Jones & Lambourne 2004). The locations of the peaks in the angular power spectrum are particularly sensitive to the curvature of the Universe. This is owing to the fact that if the Universe is flat, CMB

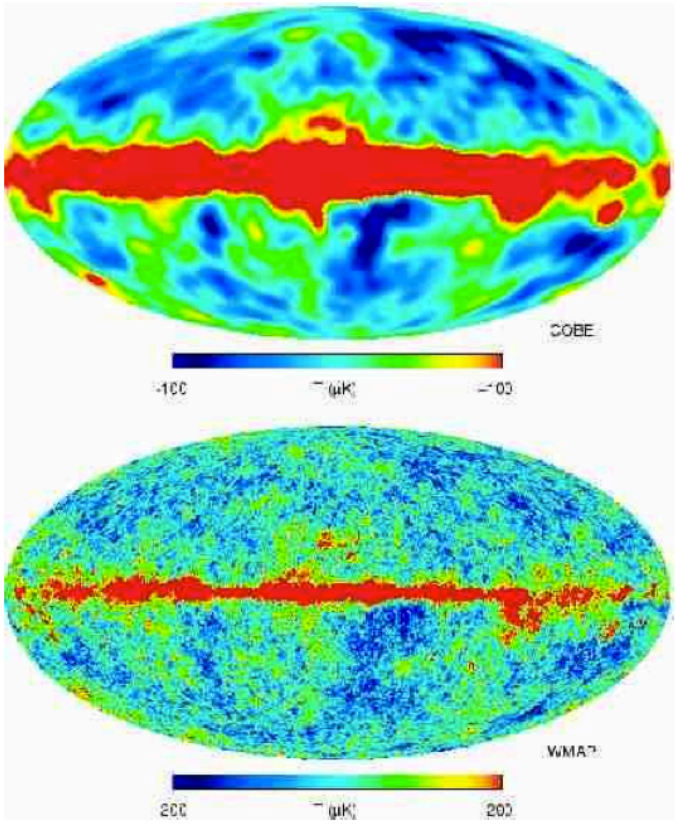


Figure 1.13: A comparison of the COBE 53 GHz map (Bennett *et al.* 1996) with the W-band WMAP map. The WMAP map has 30 times finer resolution than the COBE map (Bennett *et al.* 2003).

photons that start traveling along geodesics parallel to each other will remain parallel always, while if the Universe is not flat, CMB photons that start traveling parallel to each other will eventually diverge. If the Universe is closed this effect will shift the peak to larger angular scales (lower l) or if the Universe is open the peak will be shifted to smaller angular scales (higher l). The magnitude of this effect depends on the comoving angular diameter distance to the surface of last scattering, which is smaller for a closed universe and larger for an open universe (Dodelson 2003). Current measurements of the angular power spectrum of the CMB suggest that the Universe is in fact extremely close to flat.

1.4.3 Gravitational Lensing

In the early part of the 20$^{\text{th}}$ century, Albert Einstein predicted, from his general theory of relativity, that the path that light travels could be bent near massive objects like the sun by more than twice the amount predicted by Newtonian physics (Einstein 1915*a*,*b*). This is caused by the curvature

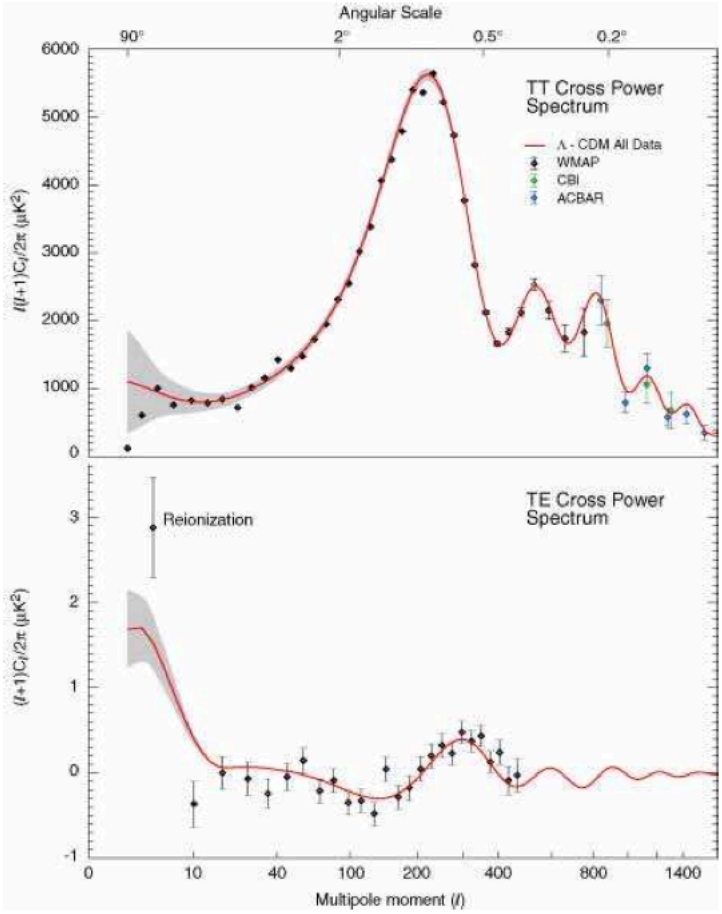


Figure 1.14: The WMAP angular power spectrum. (top:) The WMAP temperature (TT) results are consistent with the ACBAR and CBI measurements, as shown. The best fit running index ΛCDM model is shown. The grey band represents the cosmic variance expected for that model. The quadrupole has a surprisingly low amplitude. Also, there are excursions from a smooth spectrum (e.g., at $l \approx 40$ and $l \approx 210$) that are only slightly larger than expected statistically. (bottom:) The temperature-polarization (TE) cross-power spectrum, $(l+1)Cl/2\pi$ (note that this is not multiplied by the additional factor of l). The peak in the TE spectrum near $l \sim 300$ is out of phase with the TT power spectrum, as predicted for adiabatic initial conditions. The antipeak in the TE spectrum near $l \sim 150$ is evidence for superhorizon modes at decoupling, as predicted by inflationary models (Bennett *et al.* 2003).

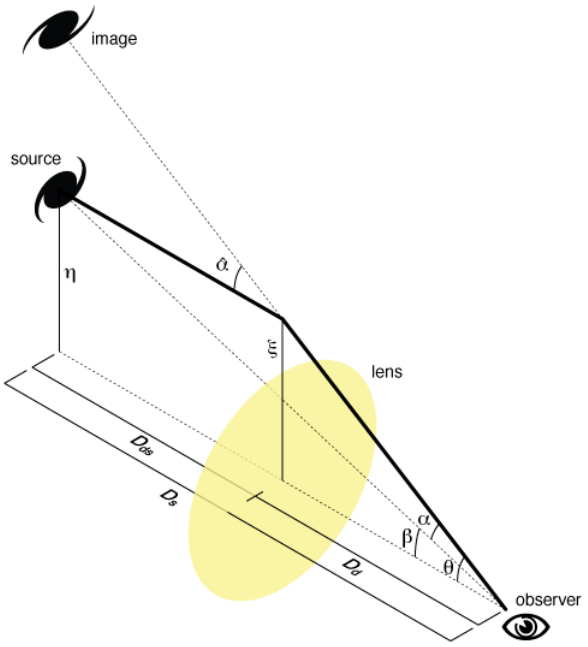


Figure 1.15: Diagram of angles involved in gravitational lensing.

of space-time around the massive object. This prediction was confirmed a few years later by Dyson *et al.* (1920), which gave rise to the idea that large bodies could act as gravitational lenses. Gravitational lensing is divided into two classes, strong gravitational lensing and weak gravitational lensing, which are discussed in the following subsections.

1.4.3.1 Strong Lensing

From general relativity the deflection of the position of a background object due to the bending of light by a point mass, M, is

$$\alpha = \frac{4\mathrm{G}M}{\xi c^2} \tag{1.149}$$

where $\xi = D_d\theta$ is the distance of closest approach as shown in fig.1.15. This figure shows the angle separating the true position of an object from it apparent position on the sky, where D_s is the distance to the object being lensed, D_d is the distance to the lensing source and D_{ds} is the distance from the lens to the object.

Given that all angles are small, θ can be expressed as

$$\theta = \left(\frac{4\mathrm{G}M}{c^2 D}\right)^{\frac{1}{2}}, \tag{1.150}$$

where $D = \left(\frac{4\mathrm{G}M}{c^2}\right)$.

Equation 1.150 implies that objects with masses $M \sim 10^{15} M_{\odot}$, which is the size of large galaxy clusters, at cosmological distances can result in large ring shaped images around the lensing source known as Einstein rings (Longair 1998). This effect was originally detected back in the 19$^{\text{th}}$ century (Lynds & Petrosian 1986; Soucail *et al.* 1987*a*,*b*). Observations of these rings can reveal information about the dark matter content and halo structure of lensing objects, and help constrain cosmological parameters (Suyu *et al.* 2010).

1.4.3.2 Weak Lensing

We define the surface mass density, Σ, which is the mass density projected onto a plane perpendicular to the incoming light ray as

$$\Sigma(\vec{\xi}) = \int \rho(\vec{\xi}, z) dz \tag{1.151}$$

(Bartelmann & Schneider 2001). For a thin lens approximation, the deflection angle then becomes

$$\vec{\hat{\alpha}} = \frac{4\mathrm{G}}{c^2} \int \frac{(\vec{\xi} - \vec{\xi}')\Sigma(\vec{\xi}')}{|\vec{\xi} - \vec{\xi}'|^2} d^2\xi'^2. \tag{1.152}$$

We can then define the dimensionless convergence, κ, as

$$\kappa(\vec{\theta}) = \frac{\Sigma(D_d\vec{\theta})}{\Sigma_{\mathrm{cr}}}, \tag{1.153}$$

where Σ_{cr} is the critical density defined as

$$\Sigma_{\mathrm{cr}} = \frac{c^2 D_s}{4\pi \mathrm{G} D_{ds} D_d}. \tag{1.154}$$

Equation 1.152 can thus be expressed as

$$\vec{\alpha}(\vec{\theta}) = \frac{1}{\pi} \int \kappa(\vec{\theta}) \frac{\vec{\theta} - \vec{\theta}'}{|\vec{\theta} - \vec{\theta}'|^2} d^2\theta'^2, \tag{1.155}$$

and we can define the deflection potential, ψ, as

$$\psi(\vec{\theta}) = \frac{1}{\pi} \int \kappa(\vec{\theta}) \ln|\vec{\theta} - \vec{\theta}'| d^2\theta'^2. \tag{1.156}$$

Finally, we define the lens equation from fig.1.15 as

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}(\vec{\theta}) = \vec{\theta} - \frac{D_{ds}}{D_s} \vec{\hat{\alpha}}(\vec{D_d\theta}). \tag{1.157}$$

The distortion between a lensed and an unlensed systems is described by a Jacobian matrix of the form

$$A_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial \alpha_i}{\partial \theta_j} = \delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}, \tag{1.158}$$

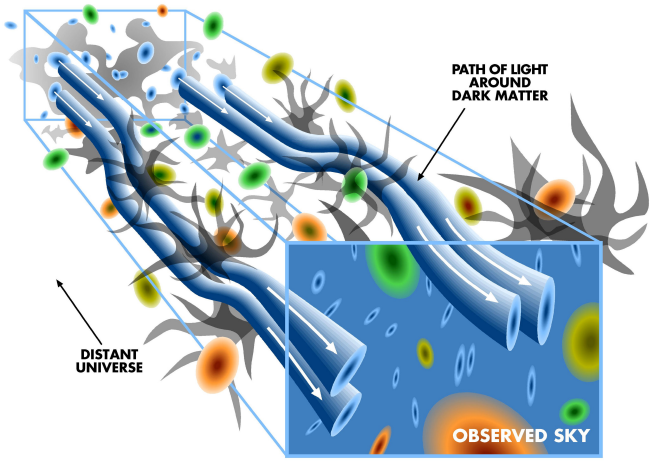


Figure 1.16: Distortion of images due to gravitational fields.

where δ_{ij} is the Kronecker delta and $\gamma = \gamma_1 + i\gamma_2$ are the components of the shear, which are related to the deflection potential via

$$\gamma_1 = \frac{1}{2}(\psi_{11} - \psi_{22}) \tag{1.159}$$

and

$$\gamma_2 = \psi_{12}. \tag{1.160}$$

The images are thus distorted in shape and size as they travel through gravitational fields. Fig.1.16 shows a diagram of the distortion of the path light travels due to gravitational fields. The shape distortion is due to the tidal gravitational field, described by the shear, γ, whereas the magnification is caused by both isotropic focusing caused by the local matter density, κ, and anisotropic focusing caused by shear (Bartelmann & Schneider 2001).

1.4.4 Photometric Redshifts

Redshifts are an important tool for probing cosmology (§1.2.1). A redshift can be used to convert the apparent properties of an object to its intrinsic properties. Normally spectroscopic redshifts are used to study objects, however measuring redshifts spectroscopically for a large number of galaxies is an arduous and time consuming process. This task becomes even more difficult when taking spectra of distant objects that require long exposure times. For this reason a method for estimating redshifts using broad-band photometry is often employed when spectroscopic data is not available.

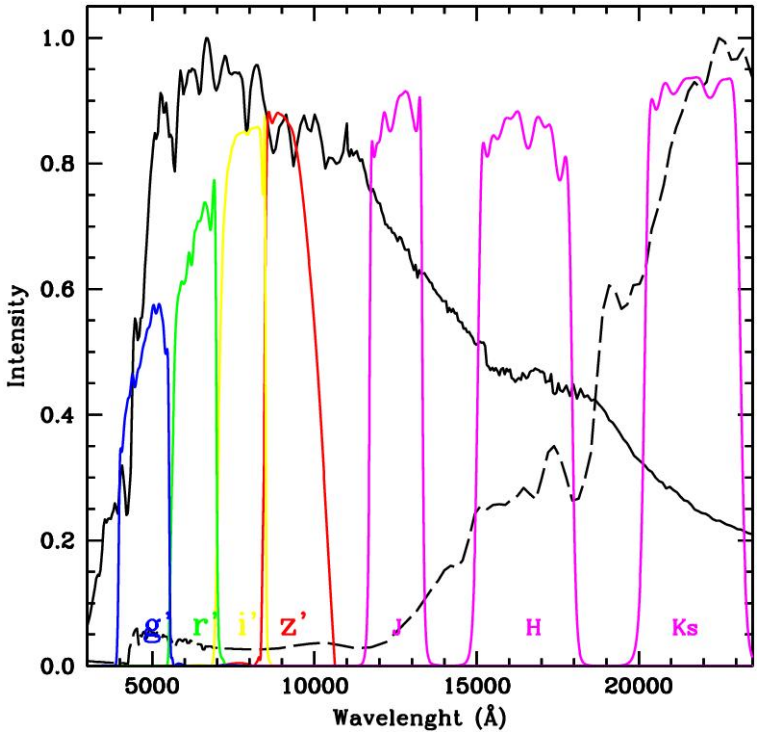


Figure 1.17: DES+Vista JHK photometric filters.

The redshifts obtained by this method are referred to as photometric redshifts and are advantageous in that they can be measured significantly faster than spectroscopic redshifts. More importantly, the imaging detectors used to measure photometric redshifts also observe a larger fraction of the sky than spectrographs allowing a greater number of objects to be measured simultaneously. These benefits do, however, need to be weighed against the drawback that uncertainties in photometric redshifts are around two orders of magnitude higher than in spectroscopic redshifts. This difference in accuracy stems from the fact that spectroscopic redshifts are measured by observing the shift of an absorption line, emission line or broad band feature of known wavelength. The photometric passband for this same region would cover a much broader range of wavelengths and this makes determining the the shift in wavelength about a hundred times more uncertain as shown in fig.1.17.

1.4.4.1 Estimation Methods

Baum (1962) conceived the first method for measuring photometric redshifts. He observed the magnitudes, in several different filters, of a sample of elliptical galaxies in the Virgo cluster and Abel 0801. These magnitudes were converted to fluxes using

$$F_f = F_0 10^{\frac{-m_f}{2.5}}, \tag{1.161}$$

where m_f is the apparent magnitude, F_f is the flux (units W Å^{-1} m^{-2}) and F_0 is the zero-point flux of the filter system (units W Å^{-1} m^{-2}), (Gwyn & Hartwick 1996). The resulting fluxes were then plotted against the wavelengths for the corresponding filters to produce low resolution spectral energy distributions (SEDs). By measuring the displacement in the SEDs relative to each other, Baum was able to determine the photometric redshifts for Abel 0801 and the Virgo cluster. Different approaches to those used by Baum were later devised by many other authors (*e.g.* Koo 1985; Loh & Spillar 1986; Connolly *et al.* 1995).

Koo (1985) devised a method that consists in plotting the photometric data in a colour-colour diagram along with a series of lines of constant redshift or iso-z lines. He was then able to determine the redshift for a galaxy by finding the closest matching iso-z line.

The two principal techniques that have been developed for measuring photometric redshifts require either fitting the observed photometric data for the galaxies to template spectra or calibrating the photometric data against a spectroscopic training set.

In the template fitting technique, originally proposed by Loh & Spillar (1986), the photometric data obtained are converted into SEDs. A selection of model SEDs or template spectra are produced spanning a range of galaxy types and redshifts relevant to the galaxies being studied. The closest matching model spectrum for a particular galaxy is determined by minimising χ^2. The redshift for that galaxy can then be defined as the redshift of its corresponding model spectrum. This method is widely used because it requires only a small sample of template spectra, however the accuracy of the results obtained is directly related to the degree to which the template spectra represent the target data. Some examples of photometric redshift estimators currently in use that employ template fitting are Hyper$Z$ (Bolzonella *et al.* 2000), Le Phare (Ilbert *et al.* 2006) and ZEBRA (Feldmann *et al.* 2006).

The most empirical method for measuring photometric redshifts was developed by Connolly *et al.* (1995). This method requires a relatively large training set of galaxies with accurately known photometry and spectroscopic redshifts. The redshift z can then be expressed as a function of the galaxy magnitudes, m, and weights, w, via

$$z = f(\bar{m}, \bar{w}). \tag{1.162}$$

Once a relation between the magnitudes and redshifts is established the photometric data from galaxies with unknown spectroscopic redshifts can be used to estimate their redshifts. This technique has the advantage of being very simple and not making any assumptions regarding a galaxys SED or evolution. The difficulty encountered when using this technique is that it can only be applied

to galaxies that are sufficiently sampled in the training set. An example of a code that uses this technique is ANNz (Collister & Lahav 2004).

1.4.4.2 ANNz

ANNz is a software package that uses an Artificial Neural Networks (ANN) to estimate photometric redshifts. An ANN is a method of non-linear statistical data modelling. The most common ANN, which is used in ANNz, is known as a multi-layer perceptron (MLP). Typical MLPs consist of a set of nodes forming an input layer, a series of hidden layers of computation nodes and an output layer (see fig.1.18). The architecture of a typical ANN can be denoted by $N_{in} : N_1 : N_2 : ... : N_{out}$ where N_{in} is the number of nodes in the input layer, N_{out} is the number of nodes in the output layer and N_n is the number of nodes in the n^{th} hidden layer. ANNz takes as inputs the magnitudes, m_i, of a given galaxy in a number of different filters and produces as a single output, the photometric redshift of that galaxy.

The nodes in each layer are connected to all the nodes in adjacent layers. These connections carry a weight, w_{ij}, and the conjoint weights comprise a vector, **w**, which is to be optimised. The output of the network is determined by calculating the activation function, $g_j(u_j)$, that is defined at each node with argument

$$u_j = \sum_i w_{ij} g_i(u_i). \tag{1.163}$$

Before an ANN can output a desired quantity from the given inputs it must be trained to model the dependency between the inputs and the output. ANNz is trained by minimising the cost function,

$$E = \sum_i (z_{phot}(\mathbf{w}, m_k) - z_{spec,k})^2, \tag{1.164}$$

where z_{phot} is the photometric redshift, z_{spec} is the spectroscopic redshift, m is the magnitude and **w** are the weights. An additional quadratic cost term is added to equation 1.164 to ensure that the weights do not become too large. The extra term is

$$E_w = \beta \sum_{i,j} w_{ij}^2. \tag{1.165}$$

To avoid over-fitting to the training set the cost function is evaluated on a separate validation set after each iteration. ANNz also assess the variance imposed by the network and photometric noise. This ensures that ANNz is an effective tool for estimating photometric redshifts. It should be noted that ANNz requires a large training set, which is representative of the objects being studied.

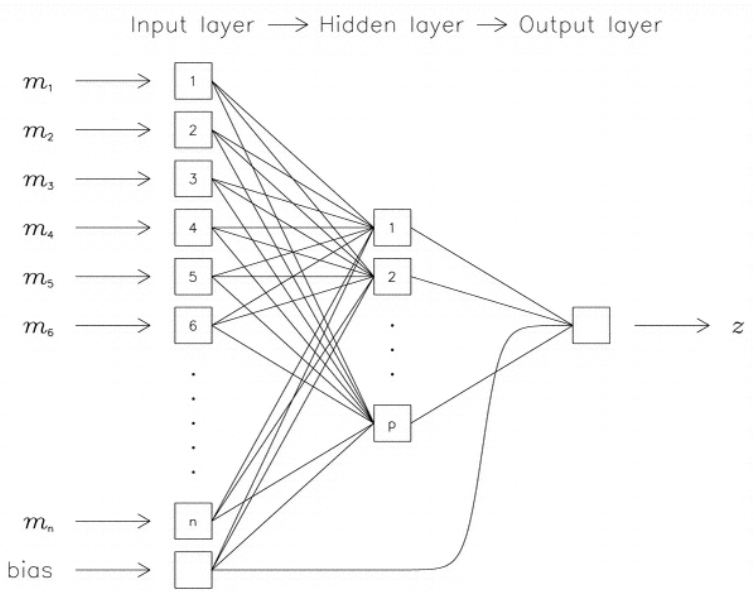


Figure 1.18: A schematic diagram of a multi-layer perceptron, as implemented by ANNz, with input nodes taking magnitudes $m_i = -2.5 \log_{10} f_i$ in various filters, a single hidden layer, and a single output node giving redshift z (Collister & Lahav 2004). Each connecting line carries a weight w_{ij} . The bias node allows for an additive constant in the network function defined at each node. More complex networks can have additional hidden layers and/or outputs.

1.4.5 Galaxy Clusters

Groups and clusters of galaxies are the largest gravitationally bound structures in the Universe and because of this they are of particular interest in cosmology. By detecting and examining galaxy clusters we gain insight into the content of matter and its distribution in the Universe. Investigating how the mass function of galaxy clusters changes with time provides an effective method for constraining cosmological parameters (Press & Schechter 1974; Peebles 1993). One could even point out that clusters are of interest as our own Milky Way galaxy forms part of the Local Group along with objects such as Andromeda, and the large and small Magellanic clouds.

Clusters can range from groups of only a few galaxies to massive structures with thousands of galaxies like the Vigo and Coma clusters. In general clusters tend to have sizes of the order $r_{\rm clt} \leq 10$ Mpc regardless of the number of galaxy members, with richer clusters being more densely packed.

Table 1.2: Abell richness classes.

Richness Class	Number of Members
0	30-49
1	50-79
2	80-129
3	130-199
4	200-299
5	$\geq$300

1.4.5.1 Early Catalogues

One of the most significant figures in cluster science was the american astronomer George Ogden Abell. Using the 48-inch Schmidt telescope at the Palomar Observatory in California, Abell surveyed around three quarters of the sky. The Palomar Sky Survey obtained 879 pairs of photometric plates over a seven year period, which Abell used to search for rich clusters by eye, where rich clusters were defined as reliable bound structures (Abell 1958). A later survey was conducted in the southern hemisphere using as twin 48-inch Schmidt telescope at the Siding Springs Observatory in New South Wales (Abell *et al.* 1989). There are 4073 clusters in the combined survey results. Abell used the following selection criteria to define a rich cluster:

- **Richness Criterion.** Clusters must have 50 members in the magnitude range m_3 to $m_3 + 2$, where m_3 is the magnitude of the 3^{rd} brightest galaxy member in the cluster. Clusters satisfying this criterion were divided into six richness classes as shown in table 1.2.

- **Compactness Criterion.** Galaxies are only counted towards the cluster richness if they are within a distance of $1.5h^{-1}$Mpc of the cluster centre. The cluster redshift is estimated from the magnitude of the 10^{th} brightest member, m_{10}.

- **Distance Criterion.** Clusters should be in the redshift range $0.02 \leq z \leq 0.2$. The lower limit is set because clusters with redshifts lower than this spanned more than one photographic plate and the upper limit is set by the magnitude limit of the survey, $m_r = 20$. Abell also divided clusters into seven distance classes as shown in table 1.3.

Another import cluster catalogue was produced by Swiss astronomer Fritz Zwicky while working at the California Institute of Technology in the 1960s (Zwicky *et al.* 1961, 1963, 1965,

Table 1.3: Abell distance classes.

Distance Class	Magnitude Range (m_{10})
0	13.3-14.0
1	14.1-14.8
2	14.9-15.6
3	15.7-16.4
4	16.5-17.2
5	17.3-18.0
6	$\geq$18.1

1966; Zwicky & Kowal 1968; Zwicky *et al.* 1968; Zwicky & Zwicky 1971). Zwicky's catalogue contained 9700 clusters that were selected using criteria similar to those employed by Abell. Clusters were required to have 50 members in the magnitude range m_1 to $m_1 + 2$, where m_1 is the magnitude of the brightest galaxy member in the cluster. Galaxies were required to lie within the range where the projected density of galaxies is twice that of the neighbouring field. Zwicky did not place any constraint on the cluster redshifts.

Both of these catalogues were produced by eye and therefore include many bias effects. Later methods, which incorporate automated algorithms, were developed to remove subjectivity in cluster detection. Some of these techniques are described in the following subsection.

1.4.5.2 Optical Detection Methods

Because of the vast amount of photometric data that is available now and that will be obtained by current and future surveys, the development of optical detection methods of clusters of galaxies is essential. There are, however, numerous different ways to approach this problem, some of which are discussed below. A good review of optical cluster detection is provided by Gal (2006).

One of the simplest techniques for finding clusters in optical data is **counts in cells**. This method has been applied to cluster detection by Couch *et al.* (1991) and Lidman & Peterson (1996) with data at $z \sim 0.5$. Counts in cells looks for enhancements above the mean background of the galaxy surface density in a cell of a given area. This enhancement or contrast is defined as

$$\sigma_{\rm cl} = \frac{N_{\rm cluster} - N_{\rm field}}{\sigma_{\rm field}}, \tag{1.166}$$

where $N_{\rm cluster}$ is the number of galaxies in the cell corresponding to the cluster, $N_{\rm field}$ is the number

of mean background counts and $\sigma_{\rm field}$ is the variance of the field counts for the same area. Clusters are identified by setting a detection threshold on the density contrast, and constraining the magnitude range and cell size. The counts in cells method is simple in concept and application, but cannot be applied to deep surveys.

Another straightforward approach is the **smoothing kernel**, which takes a weighted average of all the points in a sample. A simple version of this technique, which did not take into account redshift effects, was applied to cluster finding by Shectman (1985). A more sophisticated adaptive kernel method was later implemented by Silverman (1986) in which the kernel size varies as a function of the local density.

The Silverman (1986) method initially produces a pilot estimate, $f(x)$, of the galaxy density at each point, x, in the map and applies the smoothing kernel via

$$\lambda_i = \left[\frac{f(x)}{g}\right]^{-\alpha}, \tag{1.167}$$

where g is the geometric mean of $f(t)$ and α sets the variation of the kernel size with density. The adaptive kernel estimate is then obtained via

$$\hat{f}(x) = n^{-1}\sum_{i=1}^{n} h^{-2}\lambda_i^{-2}K\{h^{-2}\lambda_i^{-2}(x - X_i)\}, \tag{1.168}$$

where h is the bandwidth and K is the kernel (Gal 2006). Clusters can then be identified from peaks in the density map using a source extraction code such as SExtractor (Bertin & Arnouts 1996). The smoothing kernel technique is most applicable to shallow surveys with poor photometry.

One of the most successful cluster finding algorithms is the **matched filter**. This technique examines the distribution of cluster galaxies in space and luminosity at different redshifts. Postman *et al.* (1996) used this method to find 79 clusters in a region of 51deg^2 in the Palomar Distant Cluster Survey (PDCS). In the Postman *et al.* (1996) code the distribution of galaxies per unit area and magnitude is expressed as

$$D(r, m) = b(m) + \Lambda_{\rm cl} P\Big(\frac{r}{r_c}\Big)\phi(m - m^*), \tag{1.169}$$

where $b(m)$ is the background density, $\Lambda_{\rm cl}$ is the cluster contribution, $\phi(m - m^*)$ is the differential luminosity function, $P\Big(\frac{r}{r_c}\Big)$ is the projected radial profile, r_c is the characteristic cluster radius and m^* is the characteristic galaxy luminosity. Using this model, a likelihood, $\mathcal{L}$, can be defined of the form

$$\ln\mathcal{L} \propto \int P\Big(\frac{r}{r_c}\Big)\frac{\phi(m - m^*)}{b(m)}D(r, m)d^2r dm. \tag{1.170}$$

Clusters are detected by finding the redshift at which this likelihood is maximised. One of the benefits of this approach is that the richnesses are provided in the process of detecting the clusters. In addition, this method can be applied to deep surveys when the photometry is good, however it is limited in that it relies on fixed analytic luminosity functions (Gal 2006).

An extension to the matched filter method was later introduced by Kepner *et al.* (1999), which could be applied to data sets with spectroscopic or photometric redshifts. This **Adaptive Matched Filter (AMF)** method evaluates the likelihood at each galaxy position to determine if there is a cluster centred on this point at a given redshift. AMF was found to be better a determining cluster properties but less good at actually detecting them than the standard matched filter technique. Kepner *et al.* (1999) proposed a hybrid approach whereby clusters are found using the standard matched filter and the properties of these clusters are then determined with AMF.

An important observational effect that can be used for cluster detection is the **cluster red sequence (CRS)**. The CRS is a tight relation seen in colour-magnitude space, which shows that cluster ellipticals are the reddest objects at all redshifts as illustrated in fig.1.19. Gladders & Yee (2000, 2005) used this relation to find galaxy clusters by assigning survey galaxies to colour slices and applying a smoothing kernel to produce surface density maps in which they could identify the peaks.

Another approach that also takes into account the CRS relation is **maxBCG**, which was principally developed to find clusters in SDSS data (Annis *et al.* 1999; Hansen *et al.* 2005; Koester *et al.* 2007). Brightest cluster galaxies (BCGs) have predictable colours and magnitudes out to $z \sim 1$. The likelihood as a function of redshift is calculated for each BCG in a survey via

$$\mathcal{L}_{\rm max} = \max\mathcal{L}(z) = \max(\mathcal{L}_{\rm BCG+\log\mathit{N}_{gal}}), \tag{1.171}$$

where $\mathcal{L}_{\rm BCG}$ is the likelihood, at z, that a given galaxy is a BCG, based on its colours and luminosity, and $N_{\rm gal}$ is the number of galaxies within $1h^{-1}$Mpc with colours and magnitudes consistent with the CRS. Cluster candidates are selected from peaks in the $\mathcal{L}_{\rm max}$ distribution. This method produces very reliable results out to $z \sim 0.3$ after which the CRS becomes less prominent.

Techniques that make no prior assumptions about the colours, size or shapes of the clusters are **Voronoi tessellation** and **friends-of-friends (or percolation)**. These two methods are only interested in the distribution of galaxies relative to each other in a given sample.

The Voronoi tessellation technique (hereafter VT) is named after the Russian mathematician Georgy Voronoi (also known as Dirichlet tessellation after German mathematician Lejeune Dirichlet) is a method of decomposing a set of points in a plane. The decomposition is performed by

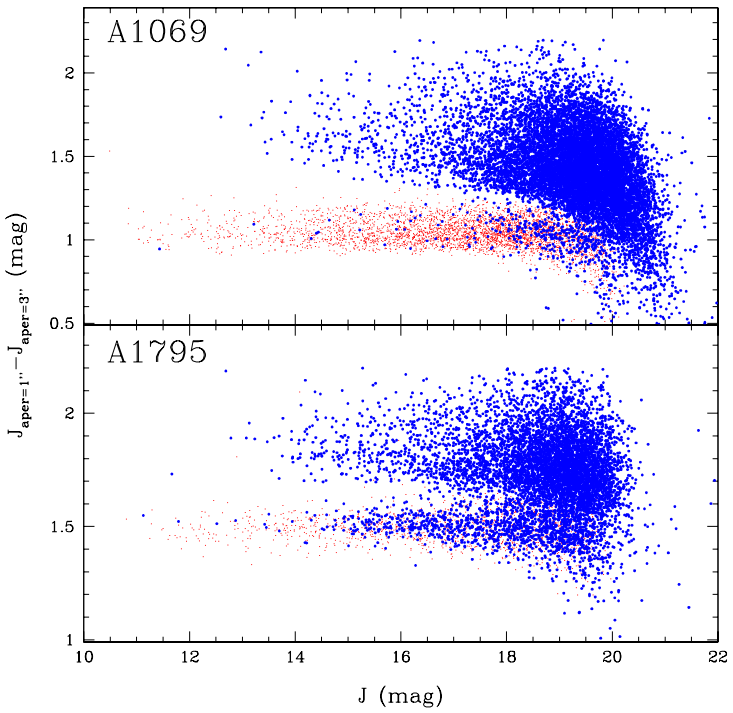


Figure 1.19: Colour magnitude diagrams of classified galaxies for a selection of WINGS-NIR clusters. The relatively thin red cluster sequences confirm the accuracy of the photometry provided by SExtractor. There are also indications of red sequences for the field galaxies at bright magnitudes, indicating the presence of background clusters at higher redshift (Valentinuzzi *et al.* 2009).

defining lines that bisect each of the points, which will result in any one point surrounded by three or more distinct points being enclosed in a cell as can be seen in fig.1.20. VT can be applied to a variety of different fields for many purposes. Ramella *et al.* (2001), Kim *et al.* (2002) and Lopes *et al.* (2004) used VT to search for clusters of galaxies. This is possible by taking a two-dimensional selection of galaxies (*i.e.* galaxies within a limited redshift interval) and using their positions in right ascension and declination to define points on a plane. Cells are defined around each of the points in the plane, the areas of which are can be inverted to provide the local densities. Galaxy clusters can then be identified by setting a density threshold and by determining the probability that an overdense region is just a random fluctuation. The regions of high density will consist of a group of small adjacent cells (*e.g.* the dark patches seen in fig.1.20).

One of the benefits of using VT to find galaxy clusters is that VT is non-parametric and does not smooth data therefore clusters are identified irrespective of their shape. This means that VT is sensitive to both symmetric and irregular clusters (Ramella *et al.* 2001).

The friends-of-friends (FoF) method simply links together galaxies that are separated by a

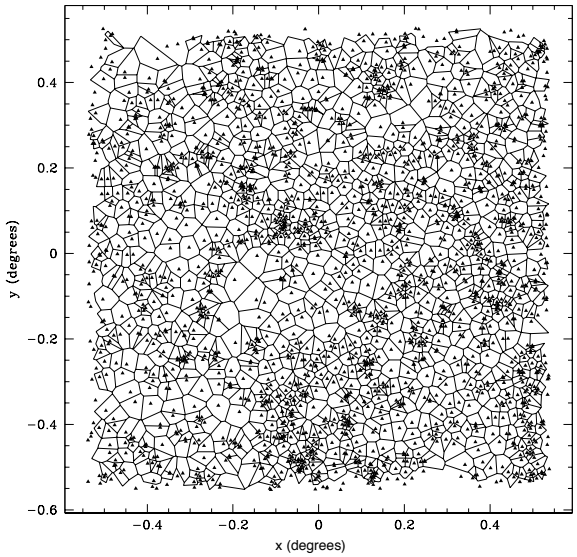


Figure 1.20: Voronoi tessellation of a galaxy field (Ramella *et al.* 2001).

distance less than some threshold. This technique has been applied to find structure in N-body simulations by (Davis *et al.* 1985; Efstathiou *et al.* 1988) and redshift surveys (Huchra & Geller 1982; Ramella *et al.* 2001). The Huchra & Geller (1982) algorithm was created to search for clusters of galaxies in the magnitude limited CFA1 redshift survey (Huchra & Geller 1982; Huchra *et al.* 1983; Geller & Huchra 1983). This code calculates the projected distance between all galaxies in a given survey via

$$D_{ij} = 2\sin(\frac{\theta_{ij}}{2})\frac{v_i + v_j}{2H_0}, \tag{1.172}$$

where θ is the angular separation between two galaxies. Galaxies are linked to each other if they satisfy

$$D_{ij} \leq D_{\rm L}(v_i, v_j, m_i, m_j), \tag{1.173}$$

and

$$|v_i - v_j| \leq V_{\rm L}(v_i, v_j, m_i, m_j), \tag{1.174}$$

where $D_{\rm L}$ and $V_{\rm L}$ are the thresholds in distance and velocity respectively, and m are the galaxy magnitudes. Clusters are identified when the number of friends in a group candidate exceeds some threshold. The threshold values are chosen depending on the properties of the galaxies in the catalogue. For simplicity D_L and V_L can be set as fixed values, however this may lead

to some selection effects being ignored. Huchra & Geller (1982) chose to adopt a method that compensates for the variation in the sampling of the galaxy luminosity function as a function of redshift. Assuming the luminosity function is independent of redshift and position, and also assuming that at larger distances only fainter galaxies are missing, the thresholds can be taken as

$$D_L = D_0 \left[\frac{\int_{-\infty}^{M_{ij}} \Phi(M) dM}{\int_{-\infty}^{M_{lim}} \Phi(M) dM} \right]^{-\frac{1}{3}} \tag{1.175}$$

and

$$V_L = V_0 \left[\frac{\int_{-\infty}^{M_{ij}} \Phi(M) dM}{\int_{-\infty}^{M_{lim}} \Phi(M) dM} \right]^{-\frac{1}{3}}, \tag{1.176}$$

where $\Phi(M)$ is the luminosity function. This algorithm is commutative which means that a unique group catalogue is created from any galaxy catalogue using any set of selection parameters.

A variation of the Huchra & Geller (1982) algorithm was later developed by Botzler *et al.* (2004), which can be applied to surveys with photometric redshifts. This code compensates for the large errors associated with photometric redshifts by dividing the entire survey into thin redshift slices and running the FoF on each slice independently. The group candidates in each redshift slice with galaxies in common are then merged to identify the galaxy clusters.

1.4.5.3 Galaxy Cluster Masses

In order to constrain cosmological parameters using galaxy clusters it is first necessary to determine their masses, however the mass of a cluster is not directly observable and therefore we have to resort to some form of observable mass proxy. Various methods exist for the non-trivial problem of estimating the masses of clusters of galaxies each of which have certain advantages and disadvantages. This subsection discusses how cluster masses can be obtained from the dynamics of the cluster galaxies, the cluster X-ray temperatures and the lensing of background galaxies by the cluster.

Dynamical mass estimates of galaxy clusters have been made as early as the 1930s (Smith 1936; Zwicky 1937). In section 1.3.3.2 it was shown that under the assumption of virial equilibrium (equation 1.112) an estimate for the mass of a cluster can be obtained using the velocity dispersion of the cluster galaxies and their mean separation as shown in equation 1.116, which can also be expressed as

$$M = \frac{3\pi N}{2G} \frac{\sum_i V_{zi}^2}{\sum_{i<j} 1/R_{\perp,ij}} \tag{1.177}$$

where V_{zi} is the line-of-sight velocity relative to the centre of the cluster and $R_{\perp,ij}$ is the projected separation of the cluster galaxies (Heisler *et al.* 1985). Although this approach is relatively straight forward and easy to implement, its reliability is questionable as interloping galaxies can significantly change the cluster velocity dispersion and hence the cluster mass. Another problem with using dynamical mass estimates is that it is difficult to know if one is obtaining a uniform sampling of the cluster potential with just the galaxies (Borgani 2006).

Another important tracer of cluster masses are **X-Ray temperature** maps. In the 1970s the UHURU X-ray observatory detected strong X-ray emission from rich galaxy clusters, the source of which is believed to be Bremsstrahlung emission from hot intracluster gas. This emission is significant because it is a powerful probe of cluster properties.
The intracluster gas can be assumed to be in hydrostatic equilibrium according to

$$\frac{dp}{dr} = \frac{\mathrm{G}M(<r)\rho}{r^2}, \tag{1.178}$$

where p is the gas pressure given by the ideal gas law,

$$p = \frac{\rho k_B T}{\mu m_{\mathrm{H}}}, \tag{1.179}$$

m_{H} is the mass of hydrogen and μ is the mean molecular weight of the gas (Longair 1998). Combining 1.178 and 1.179, and differentiating gives

$$M(<r) = -\frac{k_B T r^2}{\mathrm{G}\mu m_{\mathrm{H}}}\left[\frac{d(\log\rho)}{dr} + \frac{d(\log T)}{dr}\right], \tag{1.180}$$

which shows that the mass of the cluster can be determined if density and temperature as a function of r are known. These quantities can be obtained from the Bremsstrahlung spectral emissivity,

$$\kappa_\nu = \frac{1}{3\pi^2}\frac{Z^2 e^6}{\varepsilon_0^3 c^3 m_e^2}\left(\frac{m_e}{k_B T}\right)^{\frac{1}{2}} g(\nu, T) N N_e e^{-\frac{h\nu}{k_B T}}, \tag{1.181}$$

where N_e and N are the number densities of electrons and nuclei respectively, Z is the charge of the nuclei, ε_0 is the vacuum emissivity and $g(\nu, T) = \frac{\sqrt{3}}{\pi}\ln\left(\frac{k_B T}{h\nu}\right)$ is the Gaunt factor. Therefore, X-ray maps like those obtained by Jones & Forman (1984) (fig.1.21) and Böhringer (1995) can be used to determine the mass of galaxy clusters. This method is advantageous in that it probes the full cluster potential. Drawbacks for using X-ray emission are that obtaining accurate temperature measurements can be difficult because of effects like cooling flows in the central regions of galaxy clusters, which give rise to a mix of temperatures and densities at a given radius (*i.e.* multiphase IGM) (Peacock 1999). Additionally, X-ray telescopes tend to have either high angular resolution

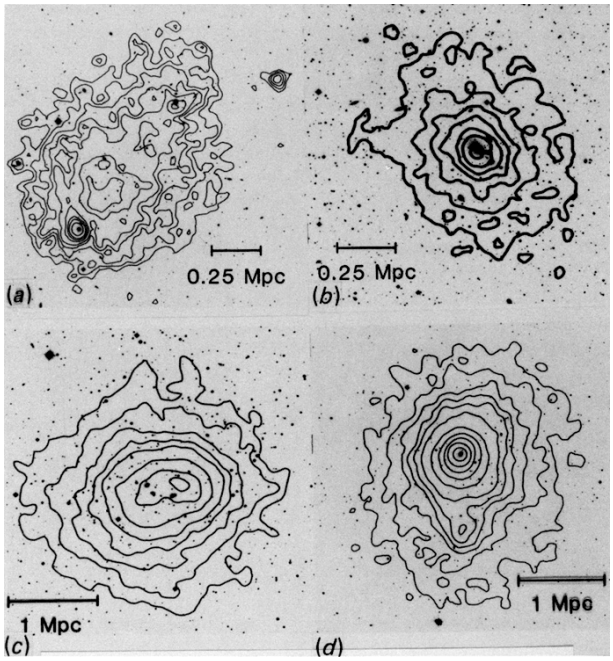


Figure 1.21: The X-ray morphology of several clusters of galaxies. Contours of constant X-ray surface brightness are shown superimposed on optical images of the clusters. (Top left), the prototypical irregular nXD cluster A1367. (Top right), the irregular XD cluster A262. (Bottom left), the regular nXD cluster A2256. (Bottom right), the regular XD cluster A85, showing the X-ray emission centred on the cD galaxy (Jones & Forman 1984).

and low spectral resolution or vice versa, where one would ideally prefer high resolution for both in order to precisely measure the X-ray emission along many lines of sight (Longair 1998).

Probably the simplest method for measuring cluster masses is through **gravitational lensing**. If a cluster of galaxies produces a large arc through strong lensing (see §1.4.3.1) the mass can be estimated via

$$M \approx 1.1 \times 10^{14} M_{\odot} \Big(\frac{\theta_{\rm arc}}{30"}\Big)^2 \Big(\frac{D_{\rm d} D_{\rm ds}}{D_{\rm s} \times 1 \text{ Gpc}}\Big), \tag{1.182}$$

where $\theta_{\rm arc}$ is the distance of the arc from the cluster centre, and $D_{\rm d}$, $D_{\rm ds}$ and $D_{\rm s}$ are the angular diameter distances from the observer to the lens, from the observer to the source and from the lens to the source respectively (Bartelmann & Narayan 1995). It should be noted that equation 1.182 assumes a smooth spherically symmetric mass distribution and therefore does not take into account the effects of asymmetry or substructure in clusters.

Another way to use gravitational lensing to estimate cluster mass is by measuring the weak lensing shear signal (see §1.4.3.2). Kaiser & Squires (1993) showed that the convergence, κ, and

shear components, γ_1 and γ_2, are related via *e.g.*

$$\kappa(\theta) = -\frac{1}{\pi}\int d^2\theta' \Big[D_1(\theta-\theta')\gamma_1(\theta') + D_2\theta-\theta')\gamma_2(\theta')\Big], \tag{1.183}$$

where $D_1(\theta) = (\theta_1^2 - \theta_2^2)/|\theta|^4$ and $D_2(\theta) = 2\theta_1\theta_2/|\theta|^4$ (Bartelmann & Narayan 1995). The convergence can be related to the surface mass density, $\Sigma(R\theta)$, as shown in equation 1.153 and hence an estimate of the mass within r_{200} (the radius at which the mean density is $200 \times \rho_{\rm crit}$) can be made via

$$M_{200} = \int_0^{2\pi} d\theta \int_0^{r_{200}} RdR\ \Sigma(R,\theta), \tag{1.184}$$

(Metzler *et al.* 1999). Therefore, cluster mass estimates can be made simply by measuring the shear of background galaxies gravitationally lensed by the cluster. This approach does not make any assumptions about the dynamical or thermodynamical state of the cluster and therefore maps the mass distribution directly. Weak lensing mass estimates, however, can be biased by the effects of infalling matter onto the cluster. Another significant limitation for using gravitational lensing to determine cluster masses is the mass-sheet degeneracy. This degeneracy arises from the fact that, in the absence of redshift information, the convergence, κ, can only be determined up to a transformation $\kappa \rightarrow \lambda\kappa + (1-\lambda)$, where λ is an arbitrary constant (Falco *et al.* 1985; Schneider & Seitz 1995). This means that the lens observables do not indicate how much of κ is due to the mass-sheet. Some methods have been proposed to break this degeneracy, but in general only for supercritical lenses (*e.g.* Bradač *et al.* 2004). This indicates that better mass estimates can be obtained for stronger lensing systems.

1.4.5.4 The Sunyaev-Zeldovich Effect

An independent method for detecting and estimating the mass of galaxy clusters is through their hot gas component via the Sunyaev-Zeldovich (SZ) effect (Sunyaev & Zeldovich 1970, 1980). The SZ effect is the result of CMB photons being Compton scattered by hot electrons in the intracluster medium. To second order there is a net gain of energy for these photons, which causes distortions in the CMB spectrum as it is shifted to slightly higher energies (Longair 1998). The observed distortion in the Rayleigh-Jeans tail of the CMB spectrum is given by

$$\frac{\Delta I_\nu}{I_\nu} = -2y, \tag{1.185}$$

where y is the Compton scattering optical depth,

$$y = \int \Big(\frac{k_B T_e}{m_e c^2}\Big)\sigma_T N_e dl. \tag{1.186}$$

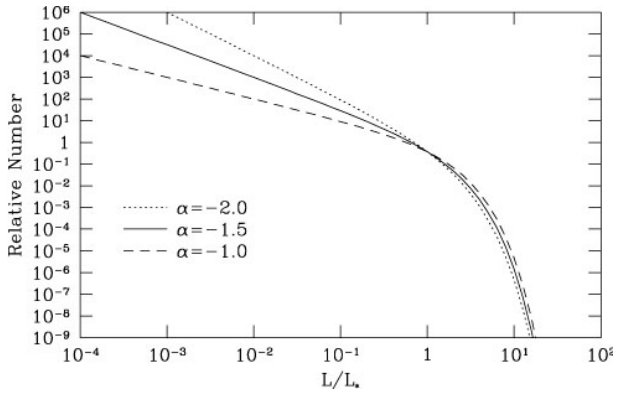


Figure 1.22: The distribution of galaxy luminosities or masses as modelled by the Schechter luminosity function assuming different power law indices (Schneider 1996).

Current surveys such as that being carried out at the South Pole Telescope (SPT) are searching for clusters of galaxies using the SZ effect in the hope of constraining the dark energy equation of state (Vanderlinde *et al.* 2010).

1.4.5.5 Mass-to-Light Ratio

The luminosity function, ϕ, is an important astronomical property that describes the comoving number density of galaxies as a function of luminosity. The Schechter luminosity function is defined via

$$\phi = \phi^* \Big(\frac{L}{L^*}\Big)^{-\alpha} e^{-(\frac{L}{L^*})} \frac{dL}{L^*}, \tag{1.187}$$

where L^* is the characteristic luminosity of a galaxy, ϕ^* is a normalisation factor, where *e.g.* $\phi = (0.0140 \pm 0.0017)h^3$ (Stromlo-APM redshift survey, Loveday *et al.* 1992), and α sets the slope of the luminosity function as seen in fig.1.22 (Schechter 1976).

The Schechter function is a good description of the luminosities of cluster galaxies and it has been shown that there is no variation in the luminosity function as a function of richness (Colless 1989). There are, however, some departures from the Schechter function for the brightest and faintest ends of cluster luminosity functions.

Mass-to-light ratios give a simple measure of how well the luminosity of an object traces it's total mass. This relationship is essential for cosmology as we can only infer the mass of objects from their observable properties and in order to determine the matter content of the Universe we require accurate mass estimates for astronomical objects, particularly large structures like clusters. The mass-to-light ratios of clusters have been used to constrain $\Omega_{\rm m}$ using the Oort limit

(*e.g.* Carlberg *et al.* 1996). Typical virial mass-to-light ratios of galaxy clusters are in the range $200 \leq M_v/L \leq 400\ hM_{\odot}L_{\odot}^{-1}$ (*e.g.* Zwicky 1933).

1.4.5.6 Observable-Mass Distribution

In order to use galaxy clusters in the context of cosmology, one generally aims to measure the mass function, dn/dM, of these structures, which due to its exponential sensitivity to cluster masses is a potentially powerful probe of cosmological parameters (see §1.3.3.4). The mass of galaxy clusters, however, is not directly observable. Therefore, what one can actually measure in a given survey is the surface density of clusters (or cluster counts), dN/dz, which can be related to the mass function via *e.g.*

$$\frac{dN}{dz} = \Delta\Omega \frac{dV}{dzd\Omega}(z) \int_0^{\infty} \phi(M,z) \frac{dn}{dM} dM \tag{1.188}$$

where $dV/(dzd\Omega)$ is the comoving volume, $\Delta\Omega$ is the angular sky coverage of the survey and $\phi(M,z)$ is a selection function that describes how the number of objects observed relates to the mass given by

$$\phi(M,z) = \frac{1}{2}\Big(\mathrm{erf}\Big[\frac{M - M_{\mathrm{lim}}(z)}{\delta M_{\mathrm{lim}}(z)}\Big]\Big), \tag{1.189}$$

where $M_{\mathrm{lim}(z)}$ is the limiting mass of the survey (Battye & Weller 2003). Battye & Weller (2003) relate this limiting mass to the limiting flux of an SZ survey to obtain a relationship between cluster masses and the temperature of the gas contained within the cluster, which unlike the mass is a directly observable property.

This relationship between cluster observables and cluster mass is known as the observable mass distribution, $P(M^{\mathrm{obs}}|M)$, which can be assumed to be given by a Gaussian distribution in $\ln M$,

$$P(M^{\mathrm{obs}}|M) = \frac{1}{\sqrt{2\pi\sigma^2_{\ln M}}} \mathrm{exp}[-x^2(M^{\mathrm{obs}})], \tag{1.190}$$

where

$$x(M^{\mathrm{obs}}) \equiv \frac{\ln M^{\mathrm{obs}} - \ln M - \ln M^{\mathrm{bias}}}{\sqrt{2\pi\sigma^2_{\ln M}}}, \tag{1.191}$$

M^{obs} is the observed mass, M is the true mass, M^{bias} is the mass bias and $\sigma^2_{\ln M}$ is the mass variance (Lima & Hu 2005).

In order to exploit the exponential sensitivity of the mass function, $P(M^{\mathrm{obs}}|M)$ has to be determined to high accuracy. Therefore, the uncertainty in $P(M^{\mathrm{obs}}|M)$ limits the power of clusters

to constrain cosmological parameters. Methods for constraining the observable-mass distribution involve employing computational simulations, which suffer from missing gas, star formation and AGN physics, and cross-correlation between lensing, microwave, optical and X-ray surveys, which propagate systematic errors in all of the surveys used (Lima & Hu 2004).

1.4.5.7 Fisher Information

One way of obtaining information about cosmological parameters from cluster counts is with a Fisher matrix. The Fisher information matrix is a method for measuring the amount of information that an observable variable X contains about an unknown parameter θ upon which the probability of X depends. The general form of the Fisher matrix is

$$\mathrm{F}_{ij} = \Big\langle \frac{\partial^2 \log \mathrm{L}(X|\theta)}{\partial\theta_i \partial\theta_j} \Big\rangle. \tag{1.192}$$

For cluster counts, if one ignores mass observable relation errors and scatter, the Fisher matrix can take the form

$$\mathrm{F}_{ij} = \sum_{\alpha=1}^{n} \frac{\partial \Delta N_\alpha}{\partial\theta_i} \frac{\partial \Delta N_\alpha}{\partial\theta_j} \frac{1}{\partial \Delta N_\alpha}, \tag{1.193}$$

where ΔN_α is the number of clusters in the interval Δz and θ are cosmological parameters (*e.g.* Tang *et al.* 2008).

In order to use the cluster counts in this way the observable mass distribution has to be margianalised over and hence becomes a nuisance parameter in the Fisher matrix. It is therefore essential to calibrate this parameter in some way.

1.4.5.8 Self-Calibration

The observable-mass distribution can be self-calibrated using a counts-in-cells analysis, where the volume of a given survey is divided into small cells and the sample variance from these cells is used to calculate the halo bias (Lima & Hu 2004; Wu *et al.* 2010). The mean number of cluster counts, $\bar{m}$, and the mean cluster bias, $\bar{b}$, for a given bin are calculated using the mass function, $dn/d\ln m$, the halo bias, $b(M)$, and the observable-mass distribution, $P(\ln M^{\rm obs}|\ln M)$, via *e.g.*

$$\bar{m} = V_c \int d\ln M \frac{dn}{d\ln M} \langle \phi | \ln M \rangle \tag{1.194}$$

and

$$\bar{b} = \frac{V_c}{\bar{m}} \int d\ln M \frac{dn}{d\ln M} b(M) \langle \phi | \ln M \rangle, \tag{1.195}$$

where V_c is the volume of the cell,

$$\langle\phi|\ln M\rangle = \int d\ln M^{\rm obs} P(\ln M^{\rm obs}|\ln M)\phi(\ln M^{\rm obs}), \tag{1.196}$$

and $\phi(\ln M^{\rm obs})$ is a binning function equal to one inside the bin and zero outside the bin (Wu *et al.* 2010). From this, the sample variance for a given cell i can be calculated using

$$S_{ij} = \langle (m_i - \bar{m}_i)(m_j - \bar{m_j})\rangle = \bar{m}_i\bar{m}_j\bar{b_i}\bar{b_j}\sigma^2_{V_c}, \tag{1.197}$$

where $\sigma^2_{V_c}$ is the variance of the dark matter density fluctuation in volume V_c. The model parameters can then be constrained using a Fisher matrix of the form

$$F_{\alpha\beta} = \bar{\mathbf{m}}^T_{,\alpha}\mathbf{C}^{-1}\bar{\mathbf{m}}_{,\beta} + \frac{1}{2}{\rm Tr}[\mathbf{C}^{-1}\mathbf{S}_{,\alpha}\mathbf{C}^{-1}\mathbf{S}_{,\beta}], \tag{1.198}$$

where the indices α and β correspond to cosmological parameters as in the previous subsubsection plus the nuisance parameters arising from self-calibration of the mass observable relation, and $\mathbf{C} = {\rm diag}(\bar{\mathbf{m}}) + \mathbf{S}$ (Lima & Hu 2004). Thus, since the sample variance depends on both the cosmology and the observable-mass distribution, the source of "noise" actually contributes to the signal and the constraints on $P(\ln M^{\rm obs}|\ln M)$ are improved.

1.4.5.9 Scaling Relations

Various techniques are employed to determine what the observable-mass distribution for galaxy clusters should be. One relatively straightforward method is to obtain scaling relations that show how a given cluster observable changes as a function of cluster mass. Two standard scaling relations used in cluster cosmology are the mass-luminosity relation and the mass-richness relation.

The **mass-luminosity relation** is a trend that relates the mass of a cluster to its luminosity, which in most cases will be the X-ray luminosity. There is a well established relation $L_{\rm X} - T^{\alpha}$, where $\alpha \simeq 2.5 - 3$, between the X-ray luminosity and temperature of a cluster as shown in the left panel of fig.1.23 (Xue & Wu 2000; Borgani 2006). As shown in section 1.4.5.3, X-ray temperatures can be used to estimate cluster masses and therefore a similar trend is seen between cluster mass and X-ray luminosity as shown in the right panel of fig.1.23 (Reiprich & Böhringer 2002). This relation is useful because we can only observe the temperature profiles of the brightest and largest clusters. One potential difficulty with this approach is that other sources of X-ray emission such as active galactic nuclei (AGN) can be very difficult to remove from data depending upon the resolution of the instrument used. This can add uncertainty into the mass-luminosity relation, although it is usually believed to be on the few percent level (Mantz *et al.* 2008).

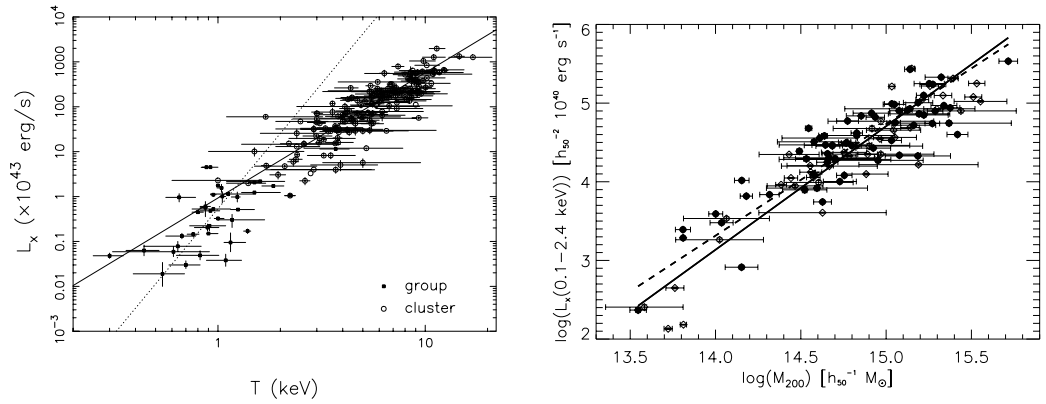


Figure 1.23: Left Panel: $L_X - T$ relations for 184 clusters (open circles) and 38 groups (filled squares). The dotted and solid lines are the best ODR fitted relations to the group and cluster samples, respectively (Xue & Wu 2000). Right Panel: Gravitational mass−X-ray luminosity relation (solid line) for the extended sample of 106 galaxy clusters. The dashed line gives the best fit relation for the 63 clusters included in HIFLUGCS (filled circles only). The bisector fit results are shown. One-σ statistical error bars are plotted for both axes, however, only the mass errors are larger than the symbol sizes (Reiprich & Böhringer 2002).

A similar relationship can be found between the flux of the SZ effect, Y, and the X-ray luminosity as shown in fig.1.24 (Diaferio *et al.* 2005). Therefore, Y can be related to cluster mass in the same way as the X-ray luminosity.

The **mass-richness** relation simply relates the mass of a cluster to the number of galaxies contained within the cluster. The difficulty with this is obtaining an accurate richness measurement for each cluster. One method, proposed by Longair & Seldner (1979), to determine cluster richnesses involves measuring the amplitude of the galaxy-cluster centre correlation function, $B_{\rm gc}$, for each cluster. This parameter scales as the net counts of galaxies normalised by the luminosity function and spatial distribution of the cluster galaxies. The mass then relates to the richness via $M \propto B_{\rm gc}^{3/\gamma}$, where γ accounts for the possible unknown evolution of the mass-richness relation, as shown in fig.1.25 (Yee & Ellingson 2003). Employing this approach requires choosing an appropriate galaxy luminosity-function and taking into account background corrections.

The principal difficulties with using any scaling relation are: gaining a clear understanding on how the given relation evolves with redshift, taking into account dynamical processes such as on-going mergers and incomplete virialization and non-gravitational heating mechanisms such as AGNs and SNe. This is one of the main systematic issues when using clusters for cosmology.

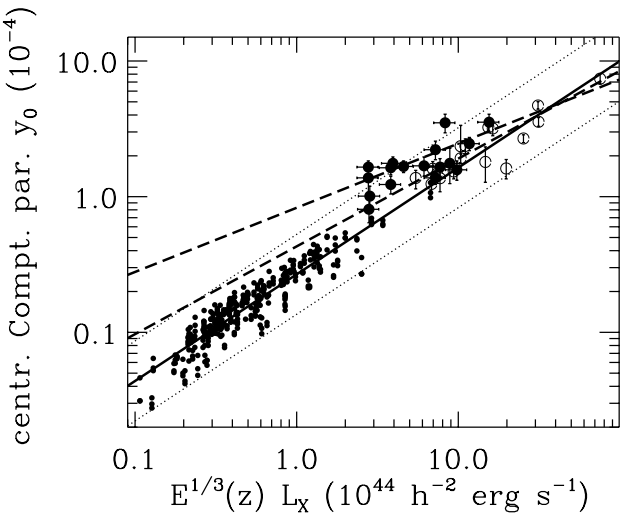


Figure 1.24: Central peak of the Comptonization parameter map vs. the X-ray bolometric luminosity. The dots are the simulated clusters, the open and solid circles with error bars the cluster sample of McCarthy *et al.* (2003) and Cooray (1999), respectively. The solid line is the best fit to the simulated clusters and the two dotted lines show the $\pm 3\sigma$ range from the best fit. The two dashed lines are the best fits to the observed samples. The steeper dashed line is the fit to the sample of McCarthy *et al.* (2003). (Diaferio *et al.* 2005).

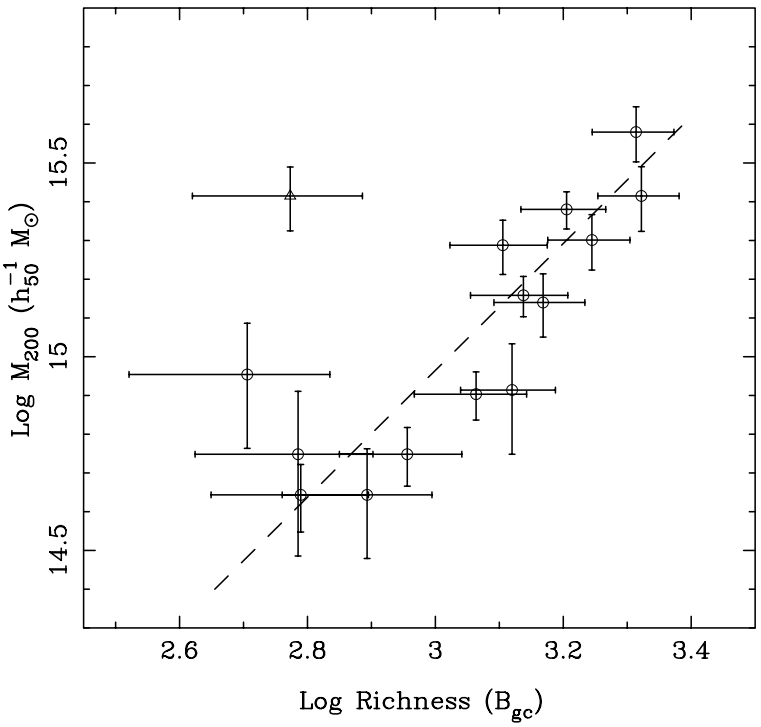


Figure 1.25: M200 vs Bgc. The dashed line shows the best fitting power law. MS1455+22 is indicated as an open triangle (Yee & Ellingson 2003).

1.4.5.10 Cosmological Constraints from Galaxy Clusters

In the unperturbed universe, scales less than $\sim 10\ h\text{Mpc}^{-1}$ have already gone non-linear (Dodelson 2003). This means that estimating the abundance of objects like galaxies, which arise from fluctuations on scales $\sim\ 1\ h\text{Mpc}^{-1}$, is very difficult. Clusters of galaxies, on the other hand, arise from scales of the order of $10\ h\text{Mpc}^{-1}$, which implies that predicting the abundance of clusters can tell us a lot about cosmology as their growth is predominantly determined by linear gravitational processes (Mantz *et al.* 2008). It has been shown that the virial mass of a rich cluster (*e.g.* $\sim 5 \times 10^{14}$ $\text{M}_{\odot}$) is very similar to the mass enclosed within a sphere of radius $8h^{-1}$Mpc in the unperturbed universe. Therefore, the current abundance of clusters gives a measure of σ_8, the rms density fluctuations in spheres of $8h^{-1}$Mpc.

In addition to σ_8, the cluster mass function also gives a measure of the total density of matter in the Universe, Ω_{m}. Recent constraints on these two parameters from galaxy clusters using the mass function are provided by *e.g* Gladders *et al.* (2007); Mantz *et al.* (2008); Rozo *et al.* (2010). Mantz *et al.* (2008) obtain values of $\Omega_m = 0.28^{+0.11}_{-0.07}$ and $\sigma_8 = 0.78^{+0.11}_{-0.13}$ using X-ray data from the Massive Cluster Survey (MACS), the local Bright Cluster Sample (BCS) and the ROSAT-ESO Flux Limited X-ray (REFLEX) sample as shown in fig.1.26. Gladders *et al.* (2007) use the Red-Sequence Cluster Survey (RCS) to derive values of $\Omega_m = 0.30^{+0.12}_{-0.11}$ and $\sigma_8 = 0.70^{+0.27}_{-0.15}$. Rozo *et al.* (2010) use the Sloan Digital Sky Survey (SDSS) maxBCG cluster catalogue to obtain values of $\Omega_m = 0.265 \pm 0.016$ and $\sigma_8 = 0.807 \pm 0.020$, which are consistent with current WMAP constraints $\Omega_m = 0.274 \pm 0.013$ and $\sigma_8 = 0.812 \pm 0.026$ (Hinshaw *et al.* 2009).

One of the principal goals of modern cosmology is to determine the value of the dark energy equation of state parameter, w (see §1.2.2.6). The evolution of the the mass function, $n(M, z)$, is governed by the growth of structure in the Universe and is therefore sensitive to w. Current and future surveys plan to use baryon acoustic oscillations, weak gravitational lensing, type Ia supernovae and galaxy clusters to constrain w. The largest weakness in using clusters as a cosmological probe is the uncertainty in the mass calibration, which at present is worse than $\sim 10\%$. In order for clusters to compete with the other cosmological probes mentioned, this uncertainty needs to be improved to $\sim 1\%$ (Albrecht *et al.* 2006; Peacock *et al.* 2006).

This thesis is focused on the optical selection of galaxy clusters, which is useful for studying the properties of the structures individually or as potential probe of cosmology.

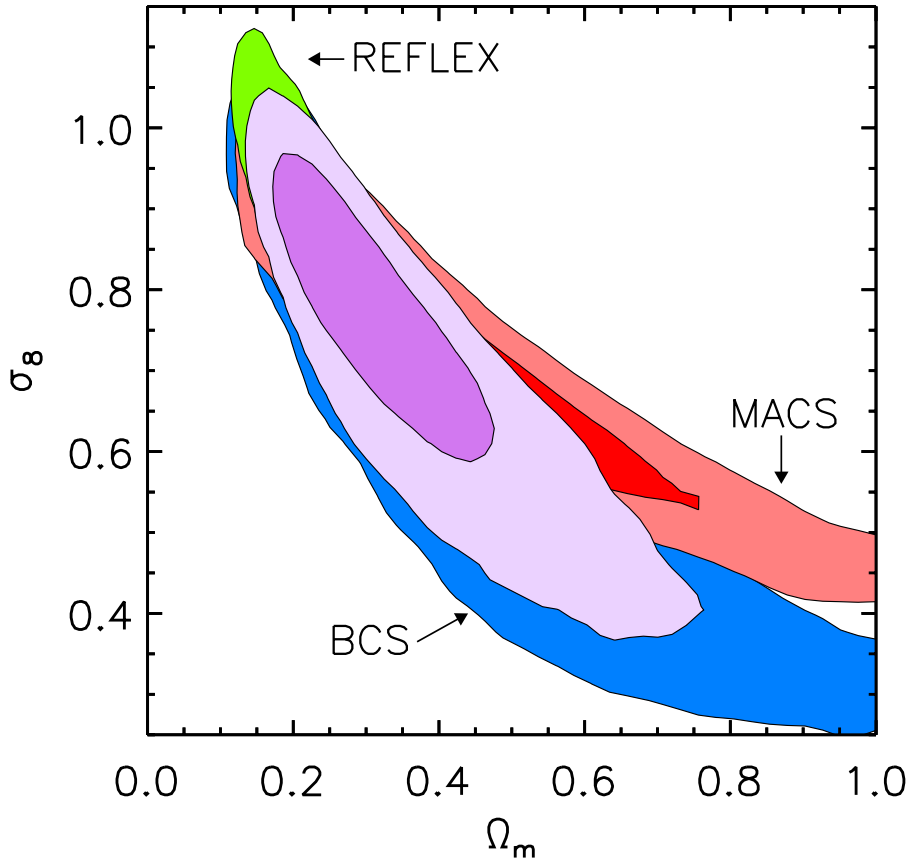


Figure 1.26: Joint 68.3 and 95.4 per cent confidence constraints on $\Omega_{\rm m}$ and σ_8 for a ΛCDM model from MACS (red), BCS (blue) and REFLEX (green) individually, and their combination (purple) using standard priors. Note that only the 95.4 per cent confidence regions are visible for the individual BCS and REFLEX data sets (Mantz *et al.* 2008).

1.5 Thesis Outline

Chapter two introduces a cluster finding algorithm based on that of Huchra & Geller (1982), used to find structure in the 2dF-SDSS and QSO (2SLAQ) Luminous Red Galaxy Survey. This chapter is based on Farrens et al. (2010). A description is provided of the simulated catalogues used to constrain the free parameters in the code and justification of the choices made. Finally, an analysis of the resulting catalogue is given, which includes looking at the two-point correlation function, the mass estimates of the clusters and an inspection of SDSS images.

Chapter three follows up on chapter two by presenting the ability of the cluster finder to be applied to 2SLAQ photometric data using a method based on that of Botzler *et al.* (2004) and comparing these results with the spectroscopic results from the previous chapter. This chapter is based on Farrens et al. (in prep). Having deduced the reliability of the photometric version of the

code, it is then run of the Mega$z$-LRG catalogue, a larger sample of LRGs. Finally, the results for both photometric 2SLAQ and Mega$z$-LRG catalogues are analysed.

Chapter four presents a description of the spectroscopic observations obtained at the Anglo-Australian Observatory in September 2007 of a selection of Mega$z$-LRG cluster candidates. This is followed by a description of the reduction and redshifting of the data. The chapter concludes with an analysis of the reduced data.

Chapter five provides a brief introduction to the Dark Energy Survey (DES). This is followed by a description of my involvement in the project, which consists mainly with the analysis of the colours, photometric redshifts and clustering in DES simulated catalogues.

Chapter six presents the conclusions of the thesis and my plans for future work.

CHAPTER 2

OPTICAL CLUSTER DETECTION I: SPECTROSCOPIC FRIENDS-OF-FRIENDS

Abstract

A catalogue of galaxy groups and clusters is presented that was selected using a friends-of-friends algorithm with a dynamic linking length from the 2dF-SDSS and QSO (2SLAQ) luminous red galaxy survey. The linking parameters for the code are chosen through an analysis of simulated 2SLAQ haloes. The resulting catalogue includes 313 clusters containing 1,152 galaxies. The galaxy groups and clusters have an average velocity dispersion of $\overline{\sigma}_v = 467.97$ kms^{-1} and an average size of $\overline{R}_{\rm clt} = 0.78$ Mpc h^{-1}. Galaxies from regions of one square degree and centred on the galaxy clusters were downloaded from the Sloan Digital Sky Survey Data Release 6 (SDSS DR6). Investigating the photometric redshifts and cluster red-sequence of these galaxies shows that the galaxy clusters detected with the FoF algorithm are reliable out to z$\sim$0.6. Masses are estimated for the clusters using their velocity dispersions. These mass estimates are shown to be reliable by comparing them with the 2SLAQ mock halo masses. Further analysis of the simulation haloes shows that clipping out low richness groups with large radii improves the purity of catalogue from 52$\%$ to 88$\%$, while retaining a completeness of 94$\%$. Finally, the two-point correlation function of the cluster catalogue is tested. A best-fitting power law model , $\xi(r) = (r/r_0)^{\gamma}$, with parameters $r_0 = 24 \pm 4$ Mpc h^{-1} and $\gamma = -2.1 \pm 0.2$ is found, which are in agreement with other low redshift cluster samples and consistent with a ΛCDM universe.

The work presented in this section is based on work presented in Farrens et al. (2010) in MNRAS.

2.1 Introduction

Galaxy clusters are the largest bound objects in the Universe and are important structures for examining the distribution of matter and how this evolves with time. Investigating how the mass-function of galaxy clusters changes with time provides an effective method for constraining cosmological parameters (Press & Schechter 1974; Peebles 1993; Weller *et al.* 2002; Weller & Battye 2003; Haiman *et al.* 2001).

The first galaxy cluster was unknowingly detected by Charles Messier, who recorded the positions of 11 nebulae in the Virgo cluster in the 18^{th} century. Later evidence for the existence of clusters of galaxies was provided in the work of Harlow Shapley and Adelaide Ames in the 1930s (Shapley & Ames 1932). They severely undermined the idea that galaxies are randomly distributed throughout the Universe. Probably the most significant figure in the pioneering of galaxy cluster detection was George Abell (Abell 1958). Abell surveyed around three quarters of the sky using photographic plates, which meant that he had to identify the locations of galaxy overdensities by eye. To avoid including field galaxies Abell defined galaxy clusters as regions 1.5 Mpc h^{-1} in radius (the Abell Radius) containing fifty or more galaxies that are no more than two magnitudes fainter than the third brightest member of the group. Remarkably, later studies have confirmed a large number of Abell's clusters as being genuine bound structures. Due to the simplicity of this two-dimensional approach, however, serious problems can arise from *e.g.* inhomogeneities, photometric errors and projection effects. Bahcall & Soneira (1983) found an excess of power in the angular correlation function using Abell clusters and van Haarlem *et al.* (1997) show, using N-body simulations, that $\sim 1/3$ of Abell's clusters have incorrect richness estimates and $\sim 1/3$ of Abell richness class $R \geq 1$ clusters are missed. Around thirty years after Abell's work hybrid photo-digital surveys were able to improve upon the photographic plate method for detecting clusters of galaxies (Maddox *et al.* 1990). It was, however, the advent of digital CCD surveys that brought about significant advances, a good example being surveys such as the SDSS (York *et al.* 2000), 2dFGRS (Colless 1999), 6dF (Jones *et al.* 2006) and the ongoing GAMA (Driver *et al.* 2008). Large sky surveys like SDSS signify a major step forward in obtaining galaxy data, however analysing that data can be approached and interpreted in many different ways. Automated algorithms supply a means of reducing subjectivity in the analysis of large data sets. Finally, the availability of increasingly detailed simulations in recent years has enabled more powerful tests of the completeness and reliability of cluster catalogues.

In the last few decades many different algorithms have been developed to find galaxy clusters.

The *Counts in Cells* method (Couch *et al.* 1991; Lidman & Peterson 1996) looks for enhancements of galaxy surface density in a given area. *Percolation* methods group together galaxies that are separated by a distance on the sky less than a given threshold distance (Efstathiou *et al.* 1988; Davis *et al.* 1985; Dalton *et al.* 1997; Ramella *et al.* 2002). This technique was originally applied to redshift surveys by Huchra & Geller (1982). An extended version of Huchra and Geller's friend-of-friends algorithm was later developed to deal with photometric redshifts by Botzler *et al.* (2004). The changes implemented in the extended friend-of-friends algorithm were necessary to account for the large errors associated with photometric redshifts (Botzler *et al.* 2004). *Matched Filter* techniques model the spatial and luminosity distributions of galaxies in a cluster to generate a cluster likelihood map (Postman *et al.* 1996; Kim *et al.* 2002). *Voronoi Tessellation* decomposes the galaxies in a region of space into discrete points surrounded by cells facilitating the identification of overdensities (Kim *et al.* 2002; Lopes *et al.* 2004; Ramella *et al.* 2001). Another method is to look for a *Red Sequence* in the colour-magnitude relation of galaxy clusters (Gladders & Yee 2000; Koester *et al.* 2007). The *Cut and Enhance* method makes colour and colour-colour cuts to produce subsamples of galaxies in different redshift ranges (Goto *et al.* 2002). A review of various methods and techniques used for the optical detection of galaxy clusters is provided by Gal (2006).

The Percolation or friend-of-friends (FoF) algorithm is the cluster finding method of interest for the purposes of this paper. The 2dF-SDSS and QSO (2SLAQ) Luminous Red Galaxy Survey, which is subset of photometrically selected luminous red galaxies (LRGs) from SDSS (Cannon *et al.* 2006), is a good starting place for the use of the FoF algorithm because of its relatively small size and the abundance of both spectroscopic and photometric data available.

This chapter is the first of two exploring galaxy clustering in the 2SLAQ catalogue and will focus on utilising the spectroscopic data available in the catalogue. The second chapter, based on Farrens et al. (in prep), will investigate detecting clusters using photometric redshifts estimated from the SDSS photometry in the 2SLAQ catalogue and compare the results with those provided in this chapter. This comparison will demonstrate the reliability of the FoF technique to detect clusters with photometric redshifts on a consistent data set. If one is to do Cosmology with galaxy clusters large data sets are required. The fastest way of obtaining this data is through photometric surveys. Upcoming surveys like the Dark Energy Survey (DES), Euclid and Planck will obtain such photometric data for large numbers of galaxies across the sky and since it is not feasible to obtain spectra for all of these objects within the observing time frames, it is therefore necessary to develop reliable methods for detecting clusters using photometric data. Photometric cluster catalogues made from surveys such as these can cover a larger volume than spectroscopic catalogues and can thus be

used to probe the shape of the mass function at higher masses.

The following section provides some background on the 2SLAQ survey. Section 2.3 describes the Huchra and Geller Friends-of-friends method in detail and how the algorithm was implemented. Section 2.4 provides a description of the mock catalogue used to calibrate the linking parameters. Section 2.5 presents the resulting groups and clusters found using the FoF algorithm and the analysis made on these results. Finally section 2.6 shows the 2-point correlation function of the clusters. A sample of the cluster catalogue is presented at the end of this chapter.

2.2 The 2dF-SDSS and QSO (2SLAQ) Luminous Red Galaxy Survey

2SLAQ is a spectroscopic survey of around 15000 potential luminous red galaxies (LRGs) in the redshift range $0.45 \leq z \leq 0.7$ (Cannon *et al.* 2006). The target LRGs were selected photometrically from the Sloan Digital Sky Survey (York *et al.* 2000). 2SLAQ also includes a lower resolution survey of faint quasi-stellar objects (QSOs) or quasars. Observations began in March 2003 using the Two-degree Field instrument (2dF) on the 3.9m Anglo-Australian Telescope.

The 2SLAQ criteria for selecting galaxies from the Sloan Digital Sky Survey consisted in placing limits on the magnitudes, colours and star-galaxy separations of the SDSS data. The magnitude limits were imposed by requiring that

$$17.5 \leq i_{deV} - A_i < 19.8 \tag{2.1}$$

and

$$i_{\text{fibre}} < 21.2, \tag{2.2}$$

where i_{deV} is the total magnitude based on a fit of each galaxy to a de Vaucouleurs profile and A_i is the extinction in the i-band. These limits enabled the selection of bright LRGs out to $z \sim 0.8$ and eliminated objects too diffuse to produce useful spectra. The colour cuts were such that

$$0.5 < g - r < 3.0, \tag{2.3}$$

$$r - i < 2.0, \tag{2.4}$$

$$c_{\parallel} \equiv 0.7(g - r) + 1.2(r - i - 0.18) > 1.6 \tag{2.5}$$

and

$$d_{\perp} \equiv (r - i) - \frac{(g - r)}{8.0} > 0.5, \tag{2.6}$$

where $c_{\parallel}$ eliminates later-type galaxies and $d_{\perp}$ selects early-type galaxies at increasingly high redshift. Equations 2.3 and 2.4 ensure that only objects near the main locus of LRGs are selected. The star-galaxy separation criteria was

$$i_{\text{psf}} - i_{\text{model}} > 0.2 + 0.2(20.0 - i_{deV}) \tag{2.7}$$

and

$$\text{radius}_{deV(i)} > 0.2\text{”}, \tag{2.8}$$

where i_{psf} is the *i*-band magnitude of an isolated star fitted with a point spread function (PSF) model, i_{model} is the model *i*-band magnitude, which is a good proxy for the PSF, and radius_{deV} is the de Vaucouleurs radius, which is the effective radius in the model magnitude. These limits eliminate the majority of stellar contamination.

2.3 Friends-of-Friends Method

2.3.1 The Huchra & Geller Friends-of-Friends Algorithm

The Huchra & Geller friends-of-friends algorithm was developed to search for clusters of galaxies in the magnitude limited CFA1 redshift survey (Huchra & Geller 1982; Huchra *et al.* 1983; Geller & Huchra 1983). This technique is commutative and utilises only a galaxy's right ascension, declination and redshift to detect structure by finding galaxies that are separated by a distance less than some threshold, D_L, and that have a velocity difference less than some threshold, V_L. The threshold values are chosen depending on the properties of the galaxies in the catalogue. For simplicity D_L and V_L can be set as fixed values, however this may lead to some selection effects being ignored. Huchra & Geller (1982) adopt a method that compensates for the variation in the completeness of the galaxy luminosity function as a function of redshift.

2.3.2 Choice of Dynamic Friends-of-Friends Linking

The friends-of-friends algorithm used in this chapter follows the Huchra & Geller (1982) method in most respects. The only difference lies in the linking parameters, which change with the surface number density of galaxies at a given redshift. This dynamic linking length is more appropriate to this sample of galaxies as is does not assume a magnitude limit pre-selection, which would remove large numbers of galaxies usable for cluster detection and prevent any clusters at the highest redshifts from being found. The choice of this type of linking parameter has the drawback of

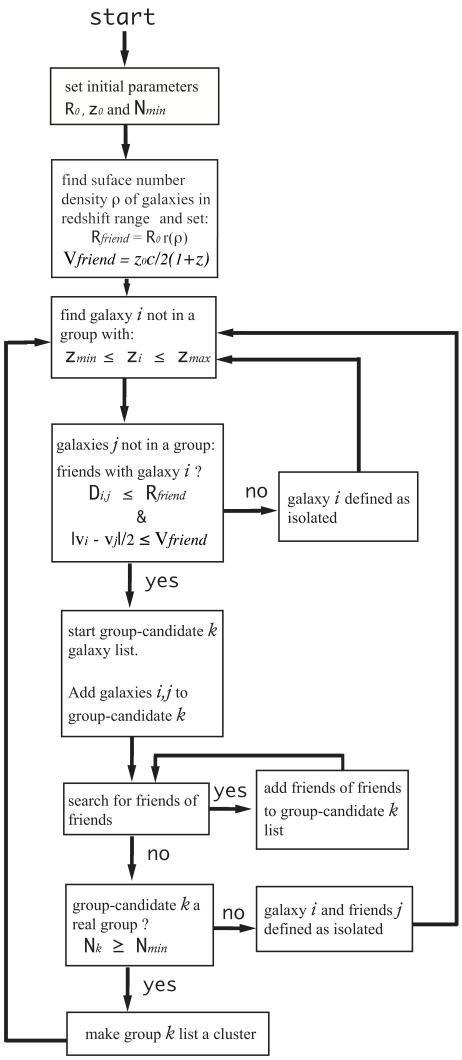


Figure 2.1: Fluxogram of the FoF algorithm with dynamic linking length. Based on Huchra & Geller (1982).

allowing for a more complex selection function that will render interpretation less straight forward. A fluxogram of the Dynamic Friends-of-Friends (hereafter DFoF) algorithm is shown in fig.2.1.

Following the steps shown in fig.2.1: First the initial values of linking length, R_0, redshift linking, z_0, and the minimum number of galaxy members needed to form a group/cluster, N_{min}, are chosen. To compensate for selection effects, R_0 is varied via equation 2.9,

$$R_{friend}(z) \propto R_0 \left(\frac{dN}{dz} \frac{dz}{dV} \frac{1}{A_{sky}} \right)^{-\frac{1}{2}}, \tag{2.9}$$

where z is redshift of each galaxy for which the variable linking length is calculated, $dN(z)/dz$ is the surface number density of galaxies in the redshift range covered by the galaxy catalogue ($z_{\rm min} \leq z \leq z_{\rm max}$), dV/dz is the differential comoving volume and A_{sky} is the fraction of sky area

covered in the catalogue relative to the total sky area. Note that the limits z_{min} and z_{max} can be modified if one does not wish to use the full redshift range of the catalogue. z_0 is converted into a velocity linking parameter, $v_{friend}(z)$, via equation 2.10,

$$v_{friend}(z) = \frac{z_0 c}{2(1+z)}, \tag{2.10}$$

where c is the speed of light. The factor of 2 in the denominator accounts for the fact that we want to be able to link galaxies that have velocities approaching along the line of sight to those that have velocities receding along the line of sight. $v_{friend}(z)$ will therefore be representative of the velocity of dispersions of the clusters being examined.

A galaxy i is selected from the catalogue that has not yet been assigned to a group and lies in the redshift range $z_{min} \leq z \leq z_{max}$. The projected distance between galaxy i and second galaxy j, D_{ij}, is calculated using

$$D_{ij} = cos^{-1}(sin(\delta_i)sin(\delta_j) + cos(\delta_i)cos(\delta_j)cos(\alpha_i - \alpha_j)) \tag{2.11}$$

where α and δ are right ascension and declination respectively.

The two galaxies are linked together (*i.e.* are friends) if they satisfy the conditions

$$D_{ij} \leq R_{friend}(z) \tag{2.12}$$

and

$$\frac{|v_i - v_j|}{2} \leq v_{friend}(z). \tag{2.13}$$

A group-candidate k is formed that includes galaxy i and its friends. A search is then made around the galaxies linked to i. This process is repeated until no further friends are found. The group-candidate k is defined as a real group if it satisfies equation

$$N_k \geq N_{min} \tag{2.14}$$

For the purposes of this paper $N_{min} \geq 3$. This is a reasonable assumption when dealing with LRGs, which are not common objects.

2.4 Mock Catalogue

To determine the optimum values of R_0 and z_0 for the DFoF code, a mock galaxy catalogue that simulates the 2SLAQ catalogue was produced. This mock catalogue contains a distribution of 7,651,076 dark matter haloes and 824,704 galaxies across an octant of the sky.

The halo catalogue was derived from the Horizon 4π simulation (Teyssier *et al.* 2009; Prunet *et al.* 2008). This is a ΛCDM dark matter N-body simulation using WMAP 3 cosmology with a 2h^{-1} Gpc periodic box on a grid of 40,963 cells. The 7×10^{10} particles were evolved using the Particle Mesh scheme of the RAMSES code (Teyssier 2002) on an adaptively refined grid (AMR) with around 1.4×10^{11} cells, reaching a formal resolution of 262,144 cells in each direction ($\sim 7\text{h}^{-1}$ kpc comoving). The simulation covers a sufficiently large volume to compute a full-sky dark matter distribution, while resolving Milky-Way size haloes with more than 100 particles and exploring small scales deeply into the non-linear regime. The dark matter distribution in the simulation was integrated in a light cone out to redshift 1, around an observer located at the centre of the simulation box. The underlying cosmology for WMAP 3 is: Ω_M = 0.24, Ω_Λ = 0.76, Ω_b = 0.042, n = 0.958, H_0 = 73 and σ_8 = 0.77.

Gravitationally bound haloes of dark matter are selected using the spherical overdensity HOP[1] method of Eisenstein & Hut (1998). HOP is based on a hybrid approach in which the local density field is first obtained by smoothing the density field with an SPH-like kernel using the n nearest neighbours. Then the particles above a given threshold are linked with their highest density neighbours until, after several "hops", they are connected to the one particle with the highest density within the region above the threshold. All particles linked to the local density maximum are identified as a group.

The haloes were then populated using the Halo Occupation Distribution (HOD), where the number of galaxies residing within each halo is drawn from a probability, $P(N|M)$, that a dark matter halo of mass M will host N galaxies.

The first moment of $P(N|M)$ is the mean number of galaxies as a function of halo mass and it is usually parameterised as a sum of a central and a satellite components (Kravtsov *et al.* 2004; Tinker 2007).

The probability that a halo contains a central galaxy is given b

$$\langle N_c|M\rangle = 0.5\left[1 + \text{erf}\left(\frac{\log_{10}(M/M_{\text{cut}})}{\sigma_{\text{cut}}}\right)\right] \tag{2.15}$$

and the number of satellite galaxies is obtained from a Poisson sampling of

$$\langle N_s|M\rangle = \left(\frac{M}{M_0}\right)^{\beta}. \tag{2.16}$$

The HOD model used is that of Blake *et al.* (2008) and includes several derived parameters that were computed at various redshifts bins between $0.4 < z < 0.7$. Table 2.1 lists the HOD parameters

[1] http://cmb.as.arizona.edu/∼eisenste/hop/hop.html

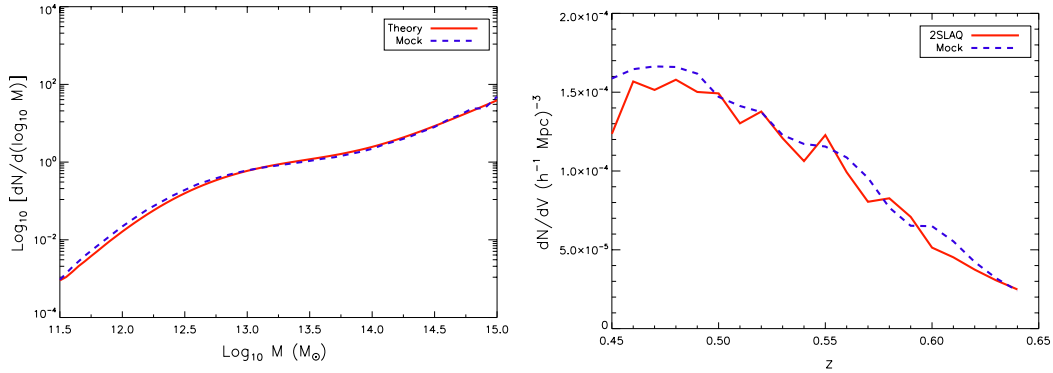


Figure 2.2: HOD (left panel) of the mock halo catalogue (blue dashed line) and the theoretical model of Blake *et al.* (2008) (red solid line). Density (right panel) as a function of redshift of the mock halo catalogue (blue dashed line) and the 2SLAQ catalogue (red solid line).

Table 2.1: The HOD parameters derived from the MegaZ-LRG sample using the methodology of Blake *et al.* (2008). The small differences between these parameters and those of Blake *et al.* (2008) reflect the difference in the assumed underlying cosmology.

Redshift slice	σ_{cut}	$\log\left(\frac{M_0}{M_\odot/h}\right)$	β	$\log\left(\frac{M_{\rm cut}}{M_\odot/h}\right)$
$0.45 < z < 0.50$	0.618	13.88	1.41	12.96
$0.50 < z < 0.55$	0.469	13.99	1.54	13.00
$0.55 < z < 0.60$	0.554	14.16	1.66	13.23
$0.60 < z < 0.65$	0.675	14.43	1.56	13.60

in each redshift bin. For the purposes of this project an evolving HOD model is constructed that smoothly interpolates between the four Blake *et al.* (2008) redshift bins as shown in fig.2.2. In the left hand panel it can be seen that the HOD model recovered from the mock catalogue matches the theoretical input. In the right panel it can be seen that the smoothly evolving HOD model, based on Blake *et al.* (2008), succeeds in reproducing the observed redshift distribution of the 2SLAQ data. This figure shows that our smoothly evolving HOD model, based on Blake *et al.* (2008), succeeds in reproducing the observed redshift distribution of the 2SLAQ data.

The radial positions of galaxies within a halo are assigned according to the NFW profile (Navarro *et al.* 1996). Specifically, it is assumed that the mass inside a given radius, properly normalised, represents the probability of containing a galaxy. Thus, integrating to the halo boundary would yield a probability of one and therefore all galaxies would be placed within this radius. The

angular position relative to the halo centre is chosen randomly for each galaxy.

In the model, all dark matter particles are assumed to be in approximately spherical virialised haloes. The velocity of a dark matter particle is the sum of two terms,

$$v = v_{\rm vir} + v_{\rm halo}, \tag{2.17}$$

the first is due to the velocity of the particle about the centre of mass of its parent halo, and the second is due to the motion of the centre of mass of the parent.

Consider the first term, $v_{\rm vir}$. We will assume that virialised haloes are isothermal spheres, so that the distribution of velocities within them is Maxwellian. This is in reasonable agreement with measurements of virial velocities within haloes in numerical simulations. If $\sigma_{\rm vir}$ denotes the rms speeds of particles within a halo, then the virial theorem requires that

$$\frac{Gm}{r} \propto \sigma_{\rm vir}^2 \propto \frac{H(z)^2}{2}\,\Delta_{\rm vir}^{1/3}(z)\left(\frac{3m}{4\pi\rho_{\rm crit}(z)}\right)^{2/3}, \tag{2.18}$$

where the final proportionality comes from the fact that all haloes have the same density whatever their mass: $m/r^3 \propto \Delta_{\rm vir}\,\rho_{\rm crit}$. This shows that $\sigma_{\rm vir} \propto m^{1/3}$: the more massive haloes are expected to be 'hotter'. At fixed mass, the constant of proportionality depends on time and cosmology, and on the exact shape of the density profile of the halo. A convenient fitting formula is provided by Bryan & Norman (1998),

$$\sigma_{\rm vir}(m,z) = 102.5\,g_\sigma\,\Delta_{\rm vir}^{1/6}(z)\left(\frac{H(z)}{H_0}\right)^{1/3}\left(\frac{m}{10^{13}M_\odot/h}\right)^{1/3}, \tag{2.19}$$

where $g_\sigma = 0.9$, and

$$\Delta_{\rm vir} = 18\pi^2 + 60x - 32x^2, \qquad \text{with} \quad x = \Omega(z) - 1 \tag{2.20}$$

and $\Omega(z) = [\Omega_m\,(1+z)^3]\,[H_0/H(z)]^2$.

It has been shown that $\sigma(M)$ is independent of local environment (Sheth & Diaferio 2001), however it may depend on position within a halo. This will mainly be due to the fact that haloes have complicated density and velocity profiles. The two-dimensional redshift-space correlation function for the mock catalogue, $\xi(\sigma,\pi)$, was calculated with the pair separation decomposed in terms of perpendicular, σ, and parallel, π, distances. Each component of separation was calculated for 20 bins equally spaced in comoving distance between $0.01 < \sigma, \pi[{\rm Mpc}h^{-1}] < 40$, creating a 400 element grid of measurements. The smooth contours are created by linear interpolation over the grid. Fig.2.3 shows the calculated 2-D correlation functions for the 2SLAQ catalogue of Ross *et al.* (2007) and the mock galaxy catalogue. It shows the expected linear squashing at large radii

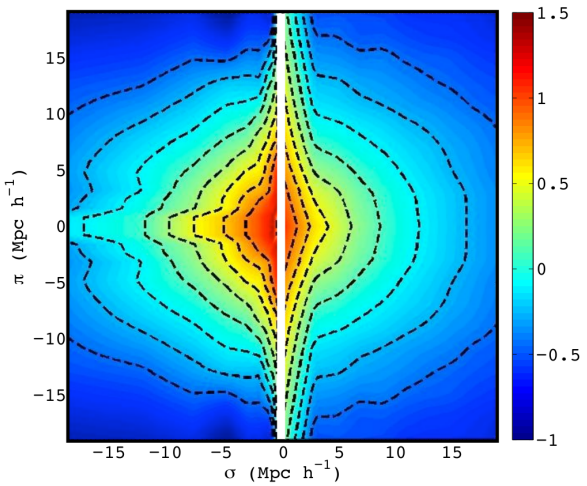


Figure 2.3: The two-dimensional redshift-space correlation function for the 2SLAQ catalogue (left half) and the mock catalogue (right half) plotted as a function of transverse, σ, and radial, π, pair separation. The colour map represents $\log(\xi)$. The 2SLAQ correlation function is that of Ross *et al.* (2007). Plot produced by Cristiano Sabiu.

and redshift distortions with 'finger of God' elongations at small scales. The colour map represents $\log(\xi)$.

Fig.2.2 and 2.3 show that the simulation reproduces the properties of 2SLAQ well enough for us to believe that it is sufficient to aid us calibrating our cluster finding parameters and therefore any limits imposed upon the values of R_0 and z_0 from the simulation will be applicable to the real data. There are, however, some caveats that should be taken into account that may cause the simulation to not be fully representative of reality. For example, the effects of things such as the way in which galaxies are added to dark matter haloes and cluster mergers at high redshifts could have influence on the results. In the following section, we discuss how the simulations can bias the results together with the analysis of the results.

2.5 Data Analysis

2.5.1 Linking Parameter Optimisation

The DFoF code was run on the 2SLAQ simulation with different values of R_0 and z_0 to find the optimum combination of the two. To gain a more physical interpretation of these parameters, hereafter R_0 and z_0 are expressed in terms of $R_{\rm friend}(z)$ and $v_{\rm friend}(z)$ at $z = 0.5$ using equations

2.9 and 2.10. The parameters were varied in the ranges

$$0 \leq R_{\text{friend}}(z = 0.5) \leq 4.2 \,(\text{Mpc } h^{-1}) \tag{2.21}$$

and

$$100 \leq v_{\text{friend}}(z = 0.5) \leq 1500 \,(\text{kms}^{-1}). \tag{2.22}$$

The choice of parameters is made based on the completeness, purity and total number of clusters in the resulting catalogues.

In order to determine the completeness and purity of the DFoF clusters relative to the simulation haloes, a membership matching code was implemented. The technique involves looking at the galaxy members assigned to each cluster by the DFoF code and matching these to the original halo member galaxies. The clusters are examined in descending order of richness to ensure that the largest clusters are the first to be matched to mock haloes. DFoF clusters may contain contributions from several mock haloes, therefore each cluster is matched to the mock halo with the highest number of shared members. Thus, if two haloes are merged into one cluster only one of the two haloes will be matched.

For the purposes of this project two matching scenarios are examined: a) a strict unique-matching regime, b) a less strict multiple-matching regime. In the first case each cluster is uniquely matched to one halo and any cluster that corresponds to a halo that has already been matched, which will be of equal or lesser richness, will be ignored. In the latter case multiple clusters are allowed to match to the same halo, which increases the completeness. The completeness and purity are defined via

$$\text{Completeness} = \frac{N_{\text{matches}}}{N_{\text{haloes}}} \tag{2.23}$$

and

$$\text{Purity} = \frac{N_{\text{matches}}}{N_{\text{clusters}}}. \tag{2.24}$$

Where N_{matches} is the total number of unique or multiple matches, N_{clusters} is the total number of DFoF clusters found and N_{haloes} is the total number of mock haloes. N_{haloes} is calculated ignoring all mock haloes with less than 3 galaxy members as this is the detection limit of the DFoF code.

Fig.2.4 shows the total number of clusters found in the 2SLAQ mock (top panel), unique completeness (middle left panel), unique purity (middle right panel), non-unique completeness (bottom left panel) and non-unique purity (bottom right panel) as a function of $R_{\text{friend}}(z = 0.5)$

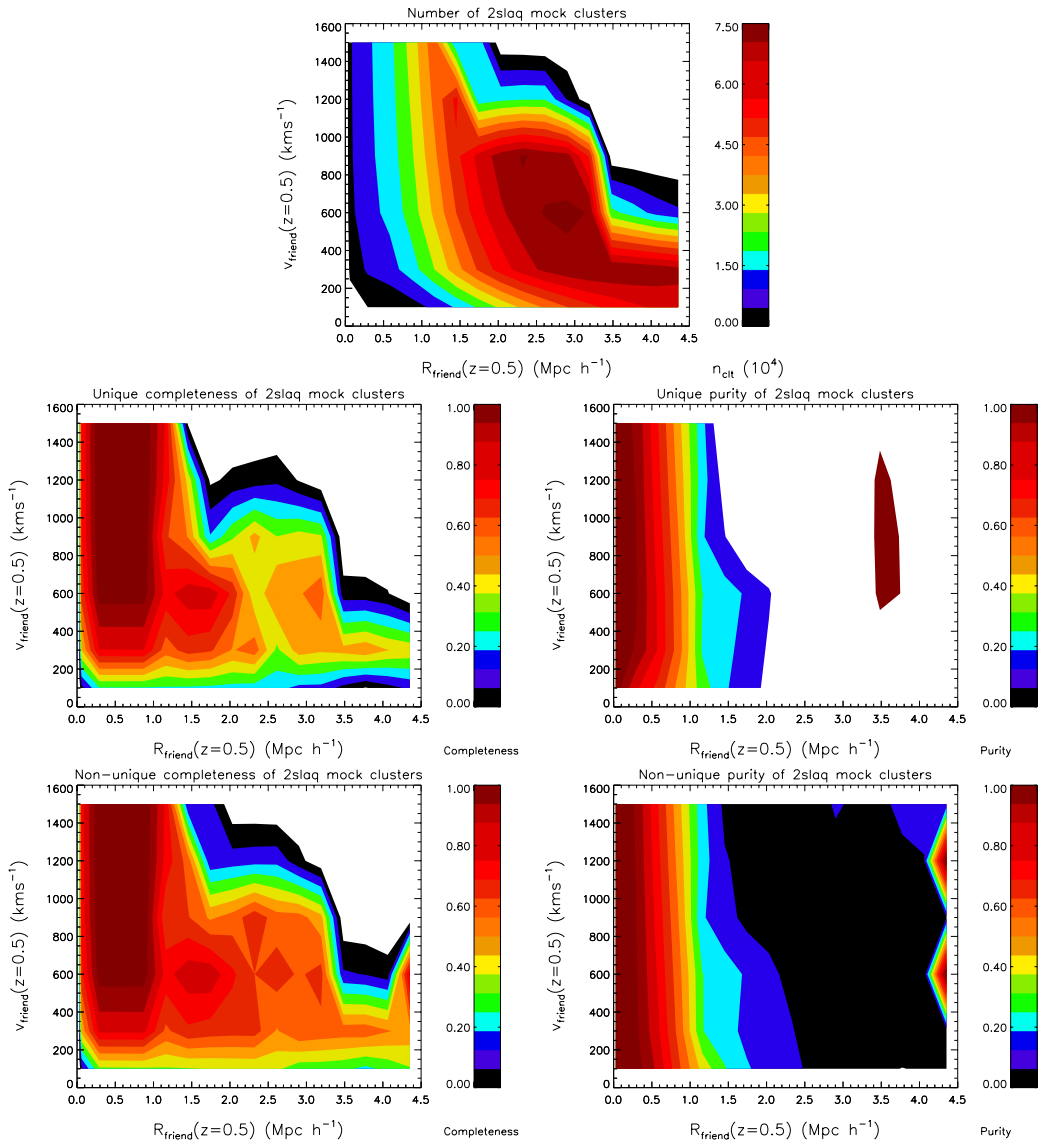


Figure 2.4: Number of clusters detected in the mock halo catalogue as a function of $\mathrm{R}_{\mathrm{friend}}(z = 0.5)$ and $v_{\mathrm{friend}}(z = 0.5)$ (top panel). Unique completeness (middle left panel) and purity (middle right panel) of the DFoF clusters relative to mock haloes as a function of $R_{\mathrm{friend}}(z = 0.5)$ and $v_{\mathrm{friend}}(z = 0.5)$. Non-unique completeness (bottom left panel) and purity (bottom right panel) of the DFoF clusters relative to mock haloes as a function of $R_{\mathrm{friend}}(z = 0.5)$ and $v_{\mathrm{friend}}(z = 0.5)$.

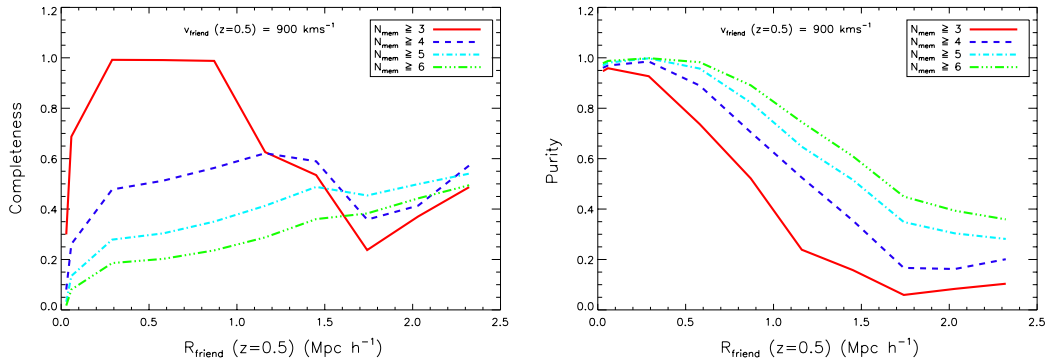


Figure 2.5: Completeness (left panel) and purity (right panel) as a function of $R_{\text{friend}}(z = 0.5)$ for a fixed value of $v_{\text{friend}}(z = 0.5) = 900$ kms^{-1}. The red solid line shows clusters with $N_{\text{mem}} \geq 3$, the blue dashed line shows clusters with $N_{\text{mem}} \geq 4$, the light blue dot-dashed line shows clusters $N_{\text{mem}} \geq 5$ and the green triple dot-dashed line shows clusters with $N_{\text{mem}} \geq 6$.

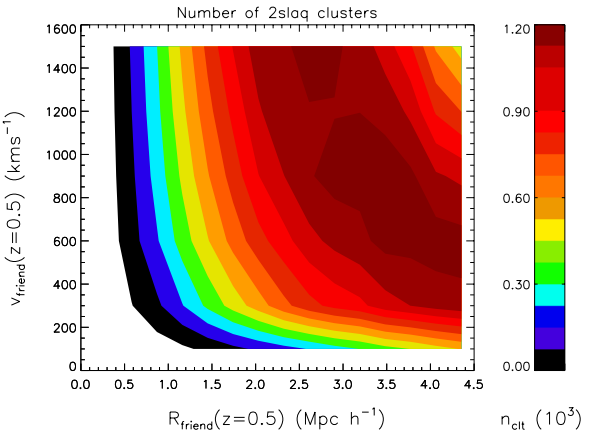


Figure 2.6: Number of clusters detected in the 2SLAQ catalogue as a function of $R_{\text{friend}}(z = 0.5)$ and $v_{\text{friend}}(z = 0.5)$.

and $v_{\text{friend}}(z = 0.5)$. This figure clearly shows that the DFoF results are primarily dependent on the choice of $R_{\text{friend}}(z = 0.5)$. Looking at the overall contour shape in the top panel, one can see that the number of clusters found peaks around $R_{\text{friend}}(z = 0.5) = 2.8$ Mpc h^{-1} and $v_{\text{friend}}(z = 0.5) = 900$ kms^{-1} after which clusters are merged together creating highly unphysical structures. This sets an upper limit on both of the parameters, although one would intuitively expect the value of $R_{\text{friend}}(z = 0.5)$ to be much lower. For $v_{\text{friend}}(z = 0.5) > 400$ kms^{-1}, the catalogue is fully complete in the range $0.35 < R_{\text{friend}}(z = 0.5) < 0.87$ Mpc h^{-1} and fully pure out to $R_{\text{friend}}(z = 0.5) = 0.28$ Mpc h^{-1} for both the unique and non-unique regimes.

Since $R_{\text{friend}}(z = 0.5)$ has a greater effect on the completeness and purity, $v_{\text{friend}}(z = 0.5)$ can

be fixed to the limit imposed by the top panel of fig.2.4 and examine how the completeness and purity vary with just $R_{\text{friend}}(z = 0.5)$. Fig.2.5 shows the variations in completeness and purity as a function of $R_{\text{friend}}(z = 0.5)$ for different richness cuts using a fixed value of $v_{\text{friend}}(z = 0.5) =$ 900 kms^{-1}. The richness cuts show what the completeness and purity would look like if groups are removed that have fewer members than some given threshold. It should be noted that the value of N_{haloes} is unchanged and therefore haloes which have fewer members than the richness cut threshold are no longer matched. This figure indicates that the richer clusters are more pure, but less complete, as one would expect. Because it is not possible to produce a catalogue that is 100% complete and 100% pure, it is necessary to choose one or the other. If a catalogue is chosen that is $\sim 100\%$ complete, fig.2.5 shows it can be cut by richness to improve the purity.

Finally the DFoF code was run on the real 2SLAQ catalogue to investigate the number of clusters found for a given set of linking parameters. Fig.2.4 shows the total number of clusters found in the 2SLAQ catalogue as a function of $R_{\text{friend}}(z = 0.5)$ and $v_{\text{friend}}(z = 0.5)$. Both this figure and the top panel of fig.2.4 have the largest N_{clt} contours roughly in the range $2.5 \leq R_{friend}(z = 0.5) \leq 4.0$ Mpc h^{-1}.

Therefore in order to obtain a fully complete catalogue with high purity and the largest number of clusters possible, based on the results in fig.2.4, 2.5 and 2.6, values of $R_{\text{friend}}(z = 0.5) = 0.87$ Mpc h^{-1} and $v_{\text{friend}}(z = 0.5) = 900$ kms^{-1} were chosen to find groups and clusters in the 2SLAQ catalogue. These values correspond to a 2SLAQ mock cluster catalogue that is 98% complete and 52% pure.

2.5.2 Basic Results

Running the DFoF algorithm using a linking length of $R_{\text{friend}}(z = 0.5) = 0.87$ Mpc h^{-1} and a velocity linking parameter of $v_{\text{friend}}(z = 0.5) = 900$ kms^{-1} on the 13,133 galaxies in the 2SLAQ sample produced a total of 313 groups and clusters containing 1,152 member galaxies. Fig.2.7 shows the angular position of the centres these groups and clusters and their distribution with respect to redshift. The cluster centres and redshifts are taken as the average values of each of the galaxy members.

Fig.2.8 shows the distribution of cluster velocity dispersion, σ_v, as a function of redshift. Fig.2.9 shows the distribution of cluster size, R_{clt}, as a function of redshift. Where R_{clt} is defined as the projected distance from the cluster centre to the farthest galaxy member. The 'x's indicate clusters with three members, the blue circles indicate clusters with between four and six members and the red square indicate clusters with seven or more members. These plots illustrate that on average the

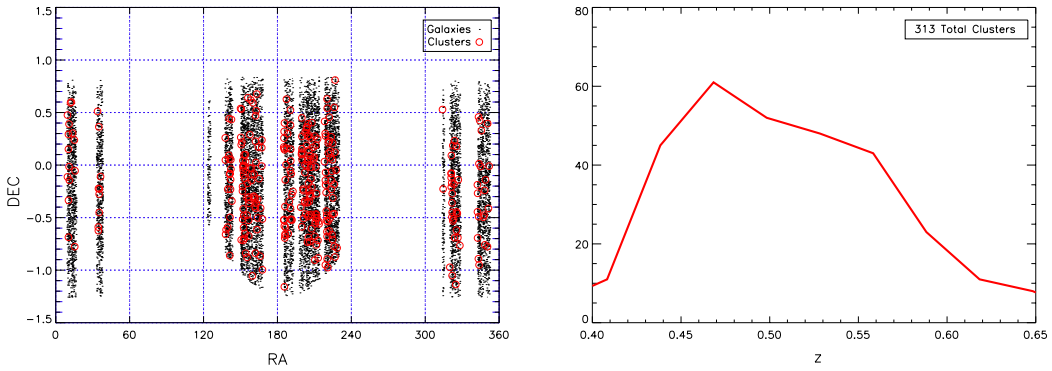


Figure 2.7: Distribution of clusters in RA and Dec (left panel). Histogram of clusters as function of redshift (right panel).

richest clusters have the largest sizes and velocity dispersions as expected.

2.5.3 Background Subtraction

The DFoF groups and clusters were examined by looking for peaks in the photometric redshift distribution and colour-magnitude relation of SDSS galaxies around the cluster centres. Using an SQL query regions of one square degree and centred on the FoF cluster candidates were downloaded from the SDSS DR6 catalogue. These squares were downloaded for each of the 313 DFoF groups and clusters and include all SDSS galaxies with photometric redshifts. The properties downloaded for each galaxy were: id, ra, dec, u-magnitude, g-magnitude, r-magnitude, i-magnitude, z-magnitude, model i-magnitude, photometric redshift, photometric redshift error and star likelihood. In each of these squares all galaxies within a 1 Mpc h^{-1} radius of the centre were taken as cluster members with background, regardless of redshift, and all galaxies between 3 and 7 Mpc h^{-1} were used as field galaxies. Galaxies between 1 and 3 Mpc h^{-1} were ignored. These limits were chosen because, on average, one would not expect galaxies father than 3 Mpc h^{-1} from a real cluster centre to be genuine members. Using the DFoF cluster redshifts, the data from the corresponding SDSS squares were stacked in redshift slices of $\Delta z = 0.05$ in the range $0.4 \leq z \leq 0.7$. The total number of groups and clusters with three or more and four or more members in each redshift slice are listed in Table-2.2.

In each cluster redshift bin the stacked SDSS galaxies were binned by photometric redshift. The number of SDSS galaxies assigned to the field was then subtracted from the number of SDSS galaxies assigned as cluster members plus background in each redshift bin. This subtraction was done taking into account the difference in area between the two regions. Fig.2.10 shows the number

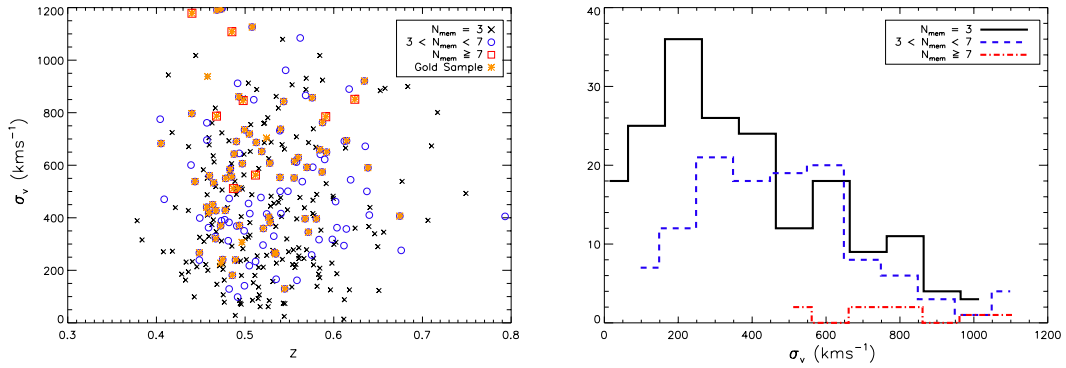


Figure 2.8: Distribution of cluster velocity dispersion, σ_v, as a function of redshift (left panel). The 'x's indicate clusters with less than four members, the blue circles indicate clusters with between four and seven members and the red squares indicate clusters with more than seven members. The gold asterisks highlight the groups and clusters that form part of the gold sample (see §2.5.5). Histogram of galaxy members as a function of cluster velocity dispersion for the 3 choices of richness as in the left hand plot (right panel).

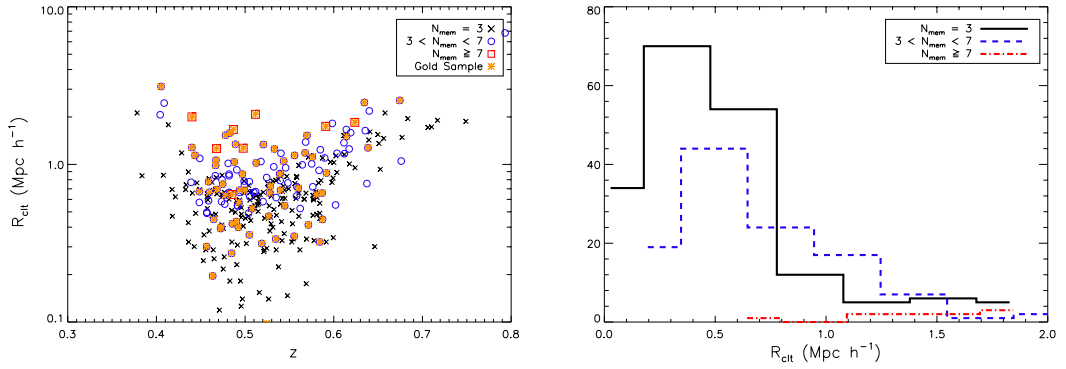


Figure 2.9: Distribution of cluster size, $R_{\rm clt}$, as a function of redshift (left panel). The 'x's indicate clusters with less than four members, the blue circles indicate clusters with between four and seven members and the red squares indicate clusters with more than seven members. The gold asterisks highlight the groups and clusters that form part of the gold sample (see §2.5.5). Histogram of galaxy members as a function of cluster size for the 3 choices of richness as in the left hand plot (right panel).

Table 2.2: Redshift Slices

z Range	N_{clt} with $N_{mem} \geq 3$	N_{clt} with $N_{mem} \geq 4$
$0.40 \leq z \leq 0.45$	26	10
$0.45 \leq z \leq 0.50$	99	46
$0.50 \leq z \leq 0.55$	85	30
$0.55 \leq z \leq 0.60$	62	24
$0.60 \leq z \leq 0.65$	26	14
$0.65 \leq z \leq 0.70$	8	2

of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters. The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies. In these plots the background subtraction should average out to zero if there is no overdensity in the field. A sharp peak can clearly been seen in each of the plots at the cluster redshift range. This trend is reliable up to $z \sim 0.6$ after which the peaks cannot clearly be distinguished. This a strong indication that the DFoF cluster candidates are genuine structures. The discrepancies seen after $z \sim 0.6$ are probably due to the relatively low number of clusters at these redshifts plus a result of poor photometric redshift estimates for more distant objects in the SDSS catalogue.

To look for the colour-magnitude relation of the groups and clusters, i-magnitude vs $(g-i)$-colour maps were made for both field galaxies and cluster+field galaxies in each SDSS square. The C-M maps were then stacked by cluster redshift in redshift slices as seen in Table-2.2. Finally the stacked field galaxy maps were subtracted from the stacked cluster galaxy maps taking into account the relative areas. Fig.2.11 shows the background subtracted C-M diagrams for SDSS galaxies. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence. The cluster red sequence is an observational property whereby cluster galaxies are more red than field galaxies at the same redshift. As with the plots in fig.2.10 this trend is visible out to $z \sim 0.6$ and similarly indicates that the structures are genuine.

2.5.4 Mass estimates

In order to calculate masses for the sample of groups and clusters, mass estimates calculated using equation 2.19 are compared with the true masses of the 2SLAQ simulation haloes. Masses were calculated for the DFoF catalogue produced from the 2SLAQ mock with linking parameters

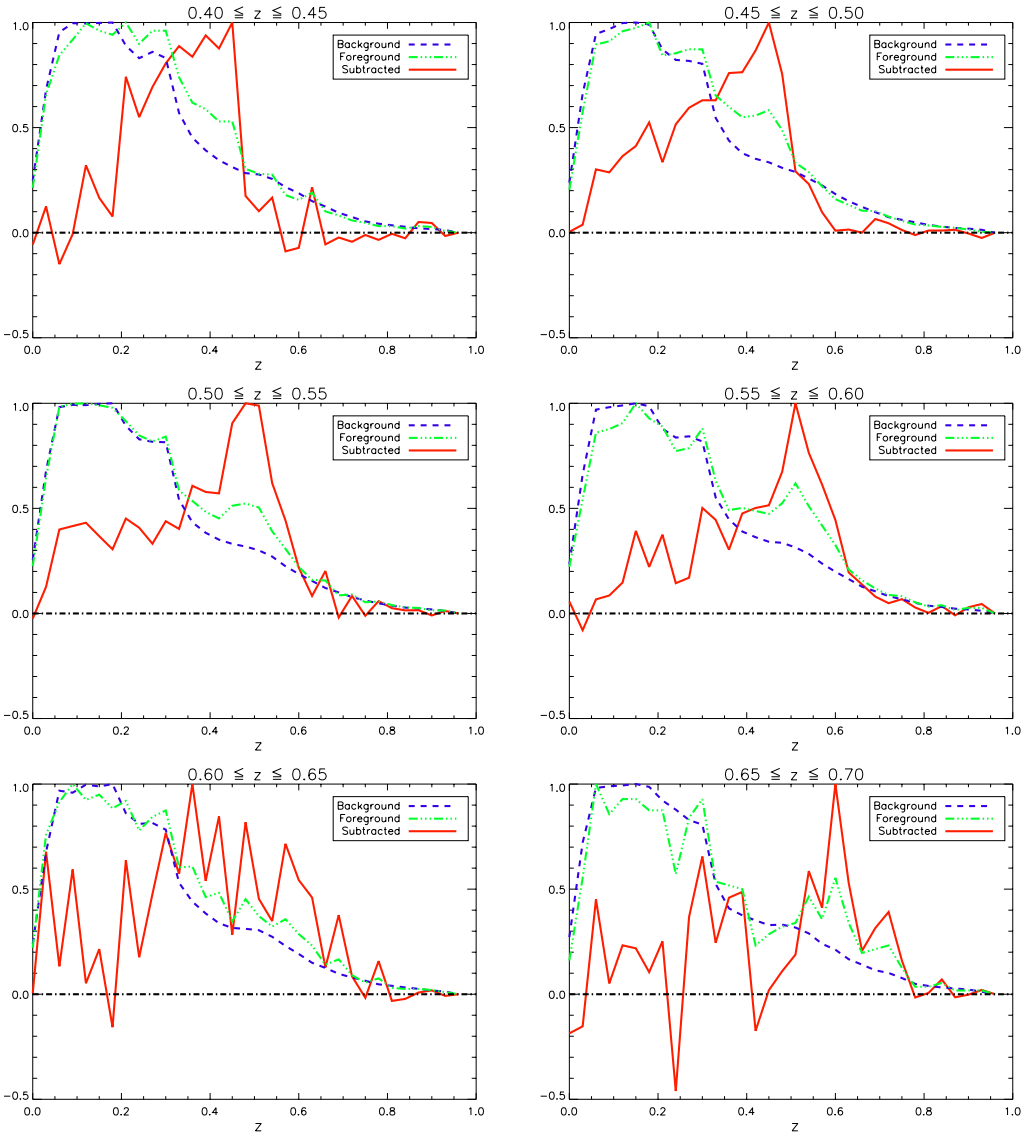


Figure 2.10: Number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters. The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies.

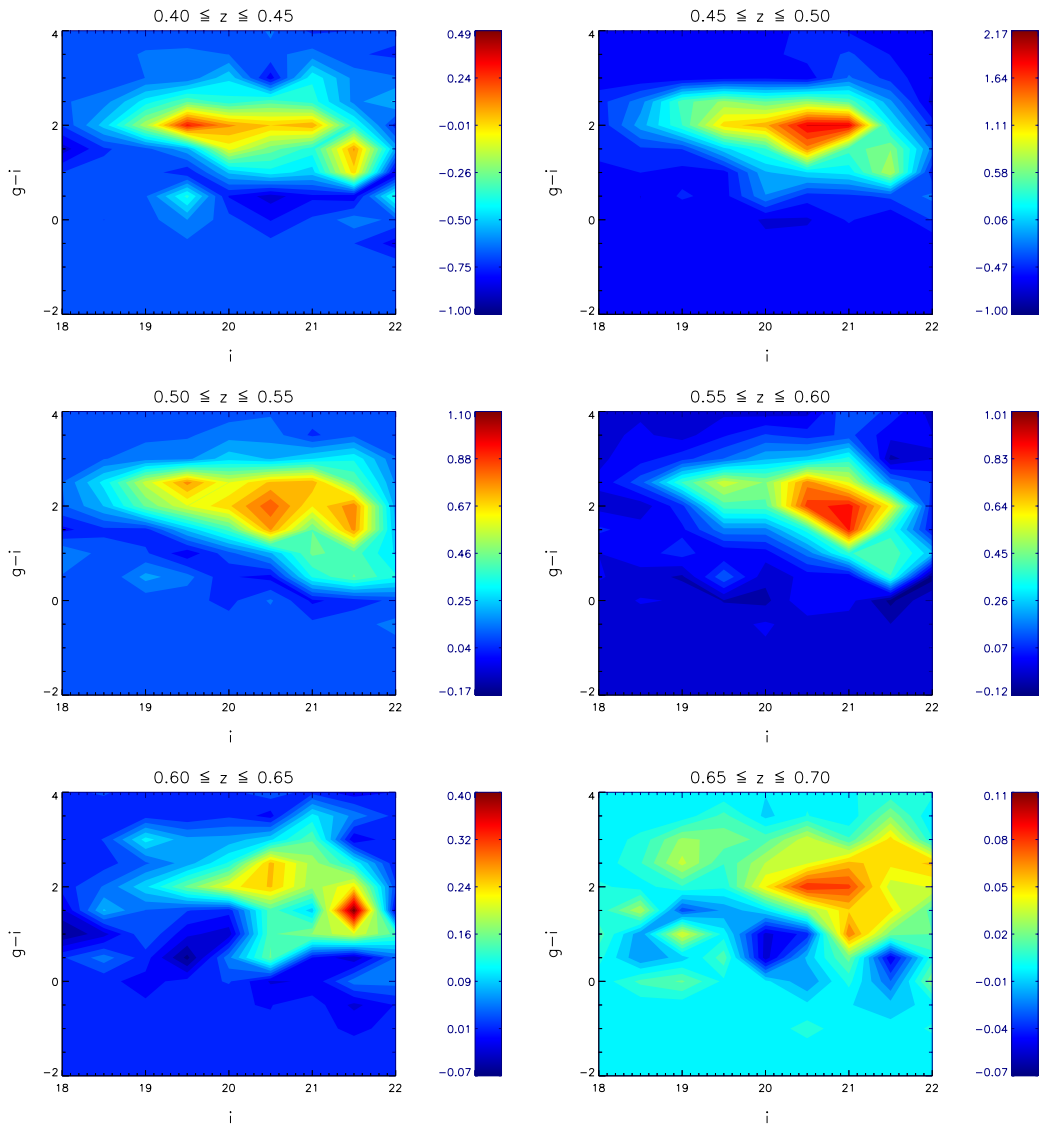


Figure 2.11: Colour-Magnitude diagrams for background subtracted SDSS galaxies. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence.

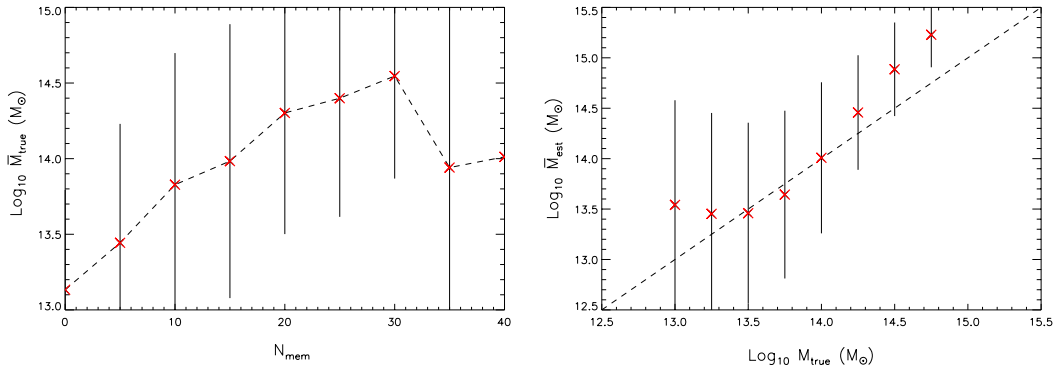


Figure 2.12: Average true cluster mass, $\overline{M}_{\rm true}$, as a function of cluster richness, $N_{\rm mem}$ (left panel). Average estimated mass, $\overline{M}_{\rm est}$, as a function of true mass (right panel).

$R_{\rm friend}(z = 0.5) = 0.87$ Mpc h^{-1} and $v_{\rm friend}(z = 0.5) = 900$ kms^{-1} using the cluster velocity dispersions (hereafter $M_{\rm est}$). The membership matching code described in section 5.1 was then implemented to assign to each cluster a mock halo mass (hereafter $M_{\rm true}$).

Fig.2.12 shows the average true cluster mass, $\overline{M}_{true}$, as a function of cluster richness, N_{mem}, and the average estimated mass, $\overline{M}_{est}$, as a function of true mass. The error bars show the standard deviation from the mean values. The left panel of this plot shows the expected trend between cluster richness and mass. The right panel compares the matched halo masses to those calculated from the cluster velocity dispersions according to equation 2.19. The deviations from the $x = y$ trend are principally the result of contaminating galaxies in the richness estimates and the small number of high mass clusters detected. There is, however, a degree of overestimation in the cluster masses from the velocity dispersions even when a mock halo has been detected perfectly.

The range of cluster masses is a good match to known masses of massive clusters. Therefore, from this analysis one can assume that mass estimates made for the real 2SLAQ cluster velocity dispersions will be approximately representative of their true physical masses with some amount of overestimation.

2.5.5 Clipping

In section 5.1 a catalogue was produced that optimised the completeness, while maintaining the highest possible purity. This choice resulted in a catalogue that was 98% complete and 52 % pure. Fig. 2.5 shows that clipping groups and clusters with the lowest richnesses will improve the purity, but will reduce the completeness. This plot also shows that the purity is much higher for lower values of $R_{\rm friend}(z = 0.5)$, which would correspond to structures that have smaller radial sizes.

Therefore it appears that the purity is mainly affected by the size and richness of the groups and clusters.

In an attempt to improve the purity while preserving the completeness, several new catalogues were made by clipping out groups and clusters with radii larger than some threshold for a given richness from the original 2SLAQ mock cluster catalogue. Table 2.3 shows the total number of groups and clusters with 3 and 4 members for various radial cuts. R_{clip} is the radial distance threshold in Mpc h^{-1} above which all groups and clusters will be removed, N_{tot} is the total number of groups and clusters for a given R_{clip}, N_{true} is the number of groups and clusters that are matched to 2SLAQ mock haloes for a given R_{clip} and the ratio N_{true}/N_{tot} gives a measure for the purity for a given R_{clip}. As can been seen in the table when no cuts are made groups of 3 and 4 members are only 38% and 57% pure respectively. Above richness of 4 all groups and clusters are above 80% pure. Therefore, the majority of the contamination in the cluster catalogue arises from these small groups.

Taking only groups and clusters with sizes less than 0.11 Mpc h^{-1} for N_{mem} = 3 and less than 0.49 Mpc h^{-1} for N_{mem} = 4 produces a new catalogue which is 94% complete and 88% pure. These values were chosen because they provide the best improvement to the purity with minimal change to the completeness. Thus, by clipping out clusters with a low richness (*i.e.* few members), but relatively large sizes, the purity can be improved by around 36%, while reducing the completeness by only 4%.

Applying this same clipping procedure to the 'real' 2SLAQ groups and clusters, the catalogue can be separated into 'gold' and 'silver' samples. Where the gold clusters are those that pass the clipping procedure and therefore are the most likely to be genuine. The silver clusters fail the clipping procedure and may still be genuine, but the probability is lower. Out of the 313 total 2SLAQ groups and clusters, 70 are gold and the remaining 243 are silver.

It should be noted that, although the mock appears to be a good representation of the 2SLAQ galaxies, it is difficult to interpret how well this clipping procedure will translate to the real 2SLAQ catalogue. Therefore the silver sample clusters are not removed, rather they are assigned a lower likelihood of being genuine than the gold sample clusters. The final 2SLAQ cluster catalogue contains an additional column that identifies each cluster as belonging to the gold or silver samples with 'G' or 'S' respectively. In fig.2.8 and 2.9 the groups and clusters that form part of the gold sample are highlighted with gold asterisks.

Table 2.3: Radial Clipping: R_{clip} is the radial distance threshold in Mpc h^{-1} above which all groups and clusters will be removed, N_{tot} is the total number of groups and clusters for a given R_{clip}, N_{true} is the number of groups and clusters that are matched to 2SLAQ mock haloes for a given R_{clip} and the ratio N_{true}/N_{tot} gives a measure for the purity for a given R_{clip}.

N_{mem} = 3				N_{mem} = 4			
R_{clip} (Mpc h^{-1})	N_{tot}	N_{true}	N_{true}/N_{tot}	R_{clip} (Mpc h^{-1})	N_{tot}	N_{true}	N_{true}/N_{tot}
-	17382	6741	0.388	-	5927	3381	0.571
0.60	13998	6739	0.481	0.70	4561	3306	0.725
0.49	11359	6739	0.593	0.60	3987	3131	0.785
0.39	9367	6739	0.719	**0.49**	**3362**	**2862**	**0.851**
0.28	7998	6739	0.842	0.39	2966	2697	0.909
0.21	7399	6738	0.911	0.32	2728	2582	0.946
0.14	6993	6708	0.959	0.25	2577	2505	0.972
0.11	**6723**	**6563**	**0.976**	0.18	2433	2415	0.992
0.07	5789	5727	0.989	0.11	2116	2108	0.996

2.5.6 Cluster Images

Fig. 2.13 shows SDSS DR7 optical *g, r and i*-band colour images of the regions around the DFoF detected clusters CL_008, CL_204, CL_122, CL_024 and CL_038. The yellow circles highlight the positions of the cluster member galaxies, which are labeled with their individual redshifts. These clusters were chosen to sample serval different redshift bins as listed in Table 2.2.

These images show that the galaxies are distributed in a small area on the sky and are close in redshift space. An important point to notice is that the distribution of cluster LRGs in each image varies from spherical to elongated filamentary structures. This is an advantage of the FoF method, which makes no prior assumptions about the shape of the groups and clusters. CL_204 (top right panel) in particular shows a rather filamentary distribution of cluster LRGs. The full cluster will be comprised of many galaxy most of which will not be LRGs and therefore the true shape could well be more spherical. Other techniques that have prerequisites on the shape of clusters would not be able to detect this particular object using only the LRGs.

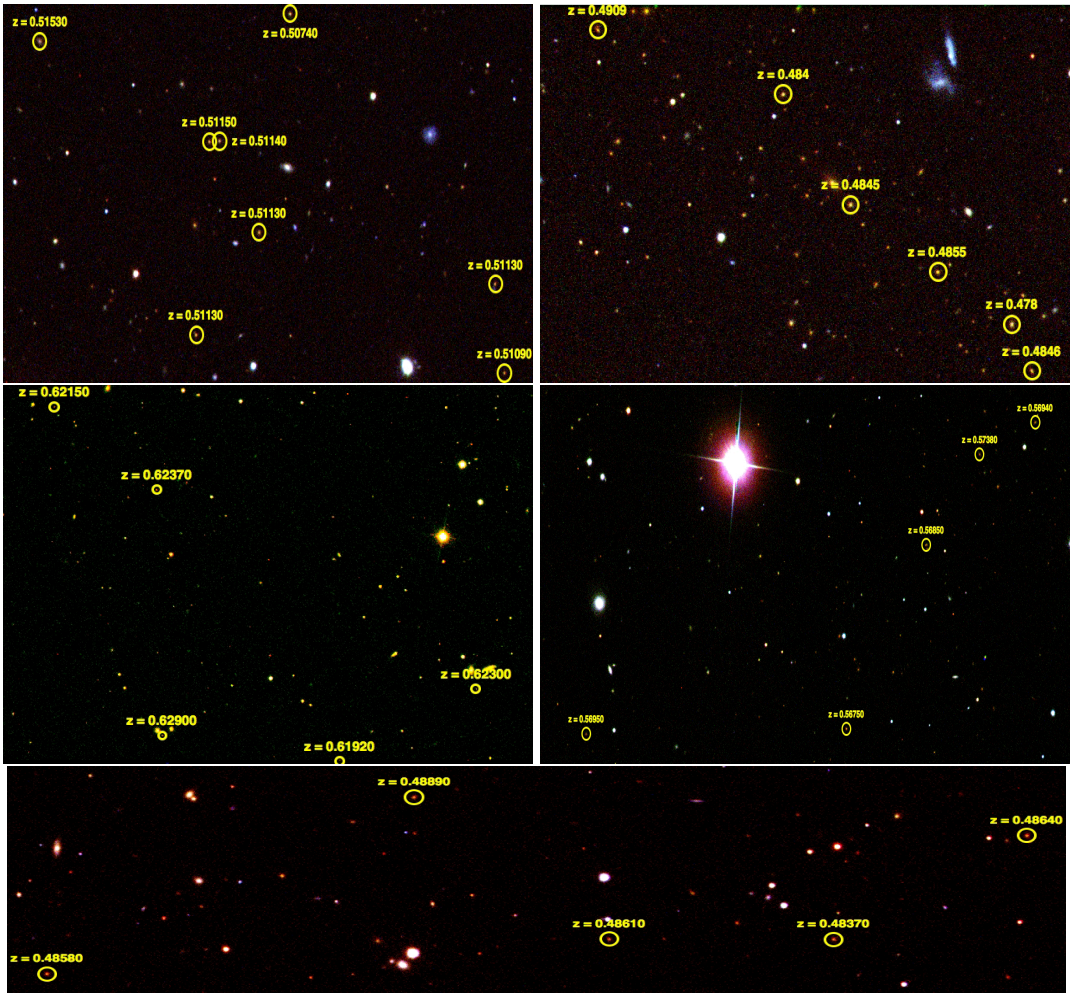


Figure 2.13: SDSS *g, r and i*-band colour images of field around CL_008 (top left panel), CL_204 (top right panel), CL_122 (middle left panel), CL_024 (middle right panel) and CL_038 (bottom panel). Yellow circles indicate the location of the cluster member galaxies, which are labeled with their individual redshifts.

2.6 Correlation Function

The spatial distribution of galaxies and clusters of galaxies contains a wealth of information regarding the underlying cosmological model. The most widely used method in the literature for condensing this information is to measure the autocorrelation function of the positional data. For our purposes we measure the correlation function simply to compare the statistical spatial distribution of our cluster sample to other derived cluster samples.

2.6.1 2SLAQ Clusters

We calculate the two-point correlation function for the 313 2SLAQ groups & clusters obtained previously. Using the 2SLAQ spectroscopic galaxy catalogue we create a random catalogue, which replicates the angular completeness on the sky, this process is handled straightforwardly in the Healpix[2] package of software. The radial distribution of the random catalogue is obtained from a smooth spline fit to the galaxy redshift distribution, $n(z)$. The smoothing ensures the exclusion of large scale structure voids and filaments.

The correlation function is calculated with the Landy & Szalay estimator (Landy & Szalay 1993),

$$\xi = \frac{DD - 2DR + RR}{RR}, \tag{2.25}$$

where DD, DR and RR are the number of pairs of points in the data, D, and random, R, catalogues. The number of pairs are calculated in 14 bins, equally separated in log space, from $r = 5 - 90$ Mpc h^{-1}.

The correlation function is usually represented as a power law, $\xi(r) = (r/r_0)^{\gamma}$, thus making a comparison with other works quite straight forward. In fig.2.14 we plot the two-point correlation function of the 2SLAQ clusters (squares with errorbars) and a best-fit power-law slope (red solid line). Uncertainties in the measurement of $\xi(r)$ are due from Poisson noise (or shot noise), which arises from a finite number of clusters that do not trace the underlying field, and cosmic variance, which is a statistical uncertainty introduced by the fact that we can only observe a fraction of the Universe at any given time. Cosmic variance is most significant at large scales as highlighted by the errorbars in fig.2.14. The errorbars are estimated using the jackknife method, which involves dividing the survey into N sections with equal area or volume. The variance and mean are estimated from N measures of our statistic. Each measurement is performed on the survey with region i removed, where $i = 1, ..., N$.

The jackknife estimate of the variance is given by

$$\sigma^2_{\xi}(r_i) = \frac{N_{jk} - 1}{N_{jk}} \sum_{j=1}^{N_{jk}} [\xi_j(r_i) - \overline{\xi}(r_i)]^2, \tag{2.26}$$

where N_{jk} is the number of Jackknife samples used and r_i represents a single bin in our statistic, ξ (Lupton 1993). In this analysis we set $N_{jk} = 20$ samples.

The power law model is best fit with the parameters $r_0 = 24 \pm 4$ Mpc h^{-1} and $\gamma = -2.1 \pm 0.2$, and the value of the reduced chi-squared is $\chi^2_{red} = 0.94$.

[2]This software is available from http://healpix.jpl.nasa.gov/

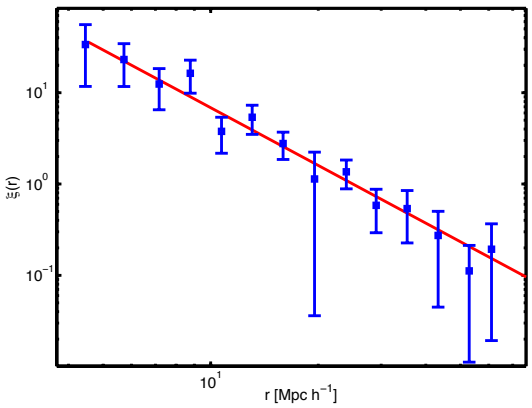


Figure 2.14: Two-point correlation function for the groups and clusters found in the 2SLAQ catalogue using the DFoF code (blue squares with error bars) and best-fitting power law slope (red solid line). $\xi_{(R)}$ was measured using the estimator of Landy & Szalay (1993). Plot produced by Cristiano Sabiu.

Comparing the r_0 value obtained with those found using other low redshift cluster samples shows good agreement, *e*.g $r_0 = 26 \pm 4.5$ Mpc h^{-1} (Borgani *et al.* 1999), $19.4 \leq r_0 \leq 23.3$ Mpc h^{-1} (Miller *et al.* 1999), 18.8 ± 0.9 Mpc h^{-1} (Collins *et al.* 2000). See Nichol (2001) for a good review of r_0 values for various cluster samples. The clustering length found here is also consistent with that expected for ΛCDM, $22 \leq r_0 \leq 27$ Mpc h^{-1} (Colberg *et al.* 2000).

2.6.2 2SLAQ Mock

The correlation function of the 2SLAQ groups and clusters is also compared with that of the mock halo catalogue presented in §2.4. Fig.2.15 shows the two-point correlation function for the haloes in the 2SLAQ mock galaxy catalogue (black squares with error bars) and best-fitting power law slope (red solid line). The correlation of the mock haloes is best fit with parameters $r_0 = 23.05 \pm 0.72$ Mpc h^{-1} and $\gamma = -1.95 \pm 0.05$, which are perfectly consistent with the parameters found for the 2SLAQ groups and clusters. The errorbars are measured using the jackknife method as before.

Although the errors are much larger for the real 2SLAQ data, this analysis indicates that the catalogue of groups and clusters found within 2SLAQ using the DFoF code shows the correct level of clustering with respect to the mock halo catalogue.

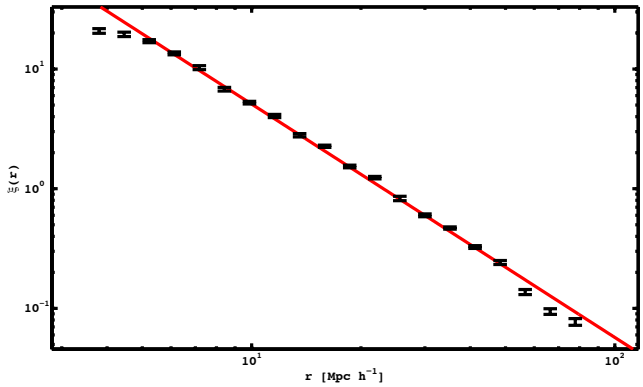


Figure 2.15: Two-point correlation function for the haloes in the 2SLAQ mock galaxy catalogue (black squares with error bars) and best-fitting power law slope (red solid line). $\xi_{(R)}$ was measured using the estimator of Landy & Szalay (1993). Plot produced by Cristiano Sabiu.

2.6.3 2SLAQ LRGs

The correlation length of the 2SLAQ LRG sample was calculated by Ross *et al.* (2007), $r_0 = 7.45 \pm 0.35$ Mpc h^{-1} (see fig.2.16), and by Sawangwit *et al.* (2009), $r_0 = 7.5 \pm 0.04$ Mpc h^{-1}. Typical galaxies have correlation lengths of $r_0 \sim 5$ Mpc h^{-1}, which means that LRG are strongly clustered objects. The correlation length of the 2SLAQ clusters is around five times greater than typical galaxies and around three times that of the 2SLAQ LRGs. This is to be expected as LRGs are more biased at all scales than typical galaxies, and at linear scales galaxy clusters are more biased than LRGs and typical galaxies (Percival *et al.* 2007; Fedeli *et al.* 2011).

2.7 Conclusions

We have written an optical cluster finding algorithm based on that of Huchra & Geller (1982). Our dynamic friends-of-friends (DFoF) code uses a linking length that compensates for selection effects by changing size depending on the surface number density of galaxies at a given redshift.

We produced a mock catalogue, which is representative of the 2SLAQ catalogue, using the Horizon 4π simulation and the HOD prescription of Blake *et al.* (2008) in order to determine the linking parameters, $R_{\text{friend}}(z)$ and $v_{\text{friend}}(z)$, in the DFoF code. The code was run with various combinations of the two parameters producing a set of distinct catalogues. We wrote a membership matching code in order to determine the completeness and purity of the resulting catalogues relative to the original mock haloes. Based on this analysis values of $R_{\text{friend}}(z = 0.5) = 0.87$ Mpc h^{-1} and $v_{\text{friend}}(z = 0.5) = 900$ kms^{-1} are chosen, which correspond to a catalogue that is 98% complete

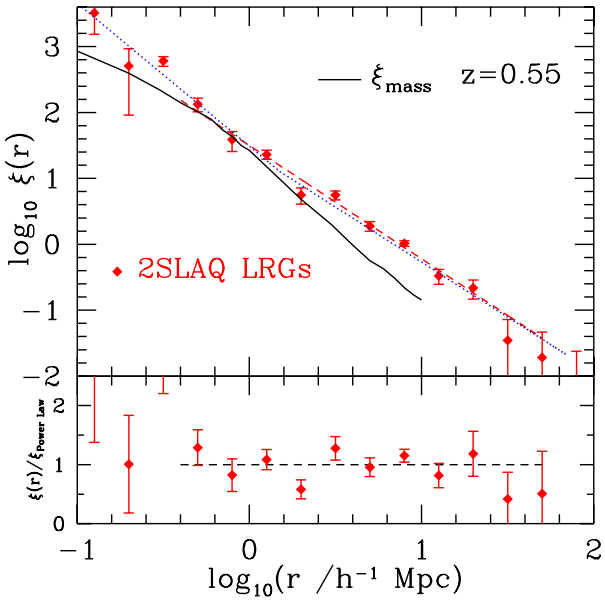


Figure 2.16: The real-space 2-point correlation function for the 2SLAQ LRG Survey (filled, red, diamonds) for the Λ cosmology. The best-fit single power-law with $r_0 = 7.45 \pm 0.35$ and $\gamma = 1.72 \pm 0.06$ is given by the dashed (red) line. The double power-law fit reported for the angular correlation, $w(\theta)$, is shown by the dotted (blue) line. The solid (black) line is a theoretical prediction for the $\xi_{\rm mass}(z = 0.55)$ using the simulations from (Colín *et al.* 1999). These models have $(\Omega_m, \Omega_\Lambda) = (0.3, 0.7)$, $h = 0.7$ and a $\sigma_8 = 1.0$. The lower panel shows the 2SLAQ LRG $\xi(r)$ measurements (assuming a Λ cosmology) divided by this best-fitting power law with the dashed line covering $0.4 < \sigma < 50$ Mpch^{-1} (Ross *et al.* 2007).

and 52% pure with the largest number of groups and clusters possible. Running the DFoF code with these values, a catalogue of 313 groups and clusters containing 1152 member galaxies was produced. The galaxy groups and clusters have an average velocity dispersion of $\overline{\sigma}_v = 467.97$ kms^{-1} and an average size of $\overline{R}_{\rm clt} = 0.78$ Mpc h^{-1}.

We tested the validity of the catalogue by obtaining SDSS galaxies in a 1deg^2 region around each cluster centre. We then subtracted the background signal from all the SDSS galaxies within 1 Mpc h^{-1} of the cluster centre and stacked the results in bins according to cluster redshift. Each of these bins was examined in SDSS photometric redshift space and colour-magnitude space. This analysis shows that the groups and clusters are reliable out to $z \sim 0.6$. The discrepancies beyond this range may be owing to the small number of groups and clusters detected at higher redshifts. Also, there may be some contamination effects in the histograms from the photometric redshift errors in the SDSS data.

We produce mass estimates for the catalogue using the cluster velocity dispersions according to

equation 2.19. We tested the reliability of this approach by comparing mass estimates for clusters found in the 2SLAQ mock with the true masses of the mock haloes to which they are matched. We find that the mass estimates of the mock catalogue groups and clusters are for the most part a good to fit to the true halo masses. The deviations seen are most likely the result of contaminating galaxies in the richness estimates and the small number of high mass clusters detected. In general one can see that the range of cluster masses is a good match to known masses of massive clusters.

We analysed optical SDSS *g, r and i*-band colour images of a selection of the clusters, which span different redshifts. We observe that the galaxy members are distributed in a small region of space, on the sky and in redshift. We also see an overdensity of red galaxies around the cluster centres. This is strong evidence that the clusters are genuine. The distribution of LRGs in the clusters varies from spherical to elongated filamentary structures. This highlights an advantage of the percolation method in that it makes no prior assumptions about the cluster shapes, which allows us to detect some structures that other methods may not.

We test different clipping procedures on the 2SLAQ mock group and cluster sizes. This analysis indicated that the majority of the contamination in the cluster catalogue was attributed to groups with 3 or 4 members. We found that by clipping out clusters with $R_{clt} < 0.11$ Mpc h^{-1} for N_{mem} = 3 and $R_{clt} < 0.49$ Mpc h^{-1} for N_{mem} = 4 improves the purity from 52% to 88%, while reducing the completeness by only 4%, in the 2SLAQ mock catalogue. By applying this procedure to the real 2SLAQ group and cluster catalogue, the clusters were separated into 'gold' and 'silver' samples. Where the gold samples consists of clusters that passed the clipping procedure and are therefore the most reliable and the silver sample consists of clusters that failed the clipping procedure and may still be genuine, however less reliable. Out of the 313 total 2SLAQ groups and clusters, 70 are gold and the remaining 243 are silver.

Finally, the two-point correlation function of the cluster catalogue was tested. A best-fitting power law model, $\xi(r) = (r/r_0)^{\gamma}$, was found with parameters $r_0 = 24 \pm 4$ Mpc h^{-1} and $\gamma = -2.1 \pm 0.2$. The value of the reduced chi-squared is $\chi^2_{red} = 0.94$. These values are in good agreement with those of the mock galaxy catalogue and with those found in the literature (Nichol 2001).

Future surveys such as the Dark Energy Survey (DES), Euclid and Planck will images millions of galaxies across the whole sky. An abundance of photometric data will be obtained for each of these objects, however it will not be possible to obtain spectroscopic data for all of them. Therefore, it is important to develop reliable cluster finding techniques that utilise the photometric data that will be available. Photometric redshifts, for example, provide a useful way to probe the properties of

Table 2.4: Cluster Catalogue Sample

N$\bar{o}$	ID	N_{mem}	RA	Dec	z	σ_v (kms^{-1})	R_{clt} (Mpc h^{-1})	$\log_{10}$ M_{est} ($M_\odot$)	Sample
1	CL_0008	011	152.85693	-0.00323	0.51165	0562.51	02.08	14.3051	G
2	CL_0031	010	210.59379	-0.55691	0.44007	1178.60	02.00	15.2916	G
3	CL_0060	009	035.42705	-0.22677	0.59094	0784.58	01.74	14.7136	G
4	CL_0140	008	342.39005	-0.44521	0.49785	0846.21	01.27	14.8415	G
5	CL_0204	007	158.86053	-0.02893	0.48510	1108.13	00.65	15.1969	G
6	CL_0038	007	191.56150	-0.51722	0.48679	0511.13	01.66	14.1882	G
7	CL_0051	007	190.76950	+0.03930	0.46779	0786.90	01.26	14.7564	G
8	CL_0122	007	161.24470	-0.30396	0.62371	0851.02	01.85	14.8094	G
9	CL_0064	006	202.19576	-0.35692	0.52837	0381.83	00.73	13.7950	G
10	CL_0024	006	343.37305	-0.49057	0.56982	0591.89	01.52	14.3531	G

galaxies along the line of sight when spectroscopic data is not present. All of the results presented in the chapter will be compared to those found in chapter 3, which will examine the clustering in the 2SLAQ catalogue using photometric redshifts. This comparison will test the reliability of the DFoF code to detect structures using photometric data.

Cluster Catalogue

The catalogue contains 313 clusters with the following properties: identifier, number of galaxy members (or richness), right ascension, declination, redshift, velocity dispersion (in kms^{-1}), radial size (in Mpc h^{-1}), estimated mass (in $\log_{10}$ $M_\odot$) and sample to which it belongs (G: gold ; S: silver). A sample of the first 10 clusters is shown in table-2.4.

CHAPTER 3

OPTICAL CLUSTER DETECTION II: PHOTOMETRIC FRIENDS-OF-FRIENDS

Abstract

A slightly modified version of the Botzler *et al.* (2004) photometric redshift friends of friends algorithm, which uses a dynamic linking length that changes with the surface number density of galaxies is presented. The code is applied to 2dF-SDSS and QSO (2SLAQ) luminous red galaxy survey using photometric redshifts obtained using the neural network code ANNz (Collister & Lahav 2004). A cluster catalogue is produced that matches the spectroscopic catalogue obtained in chapter 2 to 38% in completeness without any cuts (*i.e.* including groups and clusters with $N_{\rm mem} \geq 3$) and up to 80% if matching is only considered for 2SLAQ spectroscopic clusters with $N_{\rm mem} \geq 6$. The code is also run on the Mega$z$-LRG DR7 catalogue with linking parameters chosen based on the physical properties of the clusters found and the completeness of the Mega$z$-LRG clusters found within the 2SLAQ area to the 2SLAQ spectroscopic clusters. Two catalogues are chosen for further analysis as no obvious agreement can be found between these criteria. One catalogue that has physical properties that would be expected of clusters and one catalogue that has a high completeness relative to 2SLAQ, but includes some unphysical structures due to merging in photometric redshift space. Galaxies from regions of one square degree and centred on the galaxy clusters were downloaded from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7). Investigating the photometric redshifts and cluster red-sequence of these galaxies shows that the galaxy clusters detected with the FoF algorithm are reliable out to $z \sim 0.6$ for the 2SLAQ

photometric catalogue, consistent with the spectroscopic results, and out to $z \sim 0.65$ for the Mega$z$-LRG catalogues. These results appear to favour the more complete Mega$z$-LRG cluster catalogue with less purity and some merging of structures. The radial profiles of all three cluster catalogues is examined, also indicating a better agreement between 2SLAQ and the more complete Mega$z$-LRG catalogue. A best-fitting power law model , $\xi(r) = (r/r_0)^{\gamma}$, with parameters $r_0 = 9.55 \pm 4.23$ Mpc h^{-1} and $\gamma = -1.115 \pm 0.29$ is found. The clustering length is less than that of the 2SLAQ spectroscopic clusters due to the inclusion of more false detections in the photometric 2SLAQ cluster catalogue.

3.1 Introduction

Groups and clusters of galaxies are the largest gravitationally bound objects that one can observe in the Universe. Studying the mass function of these groups and clusters enables one to constrain cosmological parameters and to probe dark matter (Press & Schechter 1974).

At present there are many different techniques designed to detect clusters of galaxies at optical wavelengths (*e.g.* Couch *et al.* 1991; Lidman & Peterson 1996; Efstathiou *et al.* 1988; Davis *et al.* 1985; Dalton *et al.* 1997; Ramella *et al.* 2002; Kim *et al.* 2002; Lopes *et al.* 2004; Ramella *et al.* 2001). In the previous chapter (§4) the percolation method of Huchra & Geller (1982), which detects structure in spectroscopic surveys, was introduced. This method has been shown to be reliable (Huchra *et al.* 1983; Geller & Huchra 1983), however due to the time it takes to obtain decent quality spectra for a large number of objects many of the current and upcoming surveys, such as the Sloan Digital Sky Survey (SDSS, York *et al.* 2000), DES (The Dark Energy Survey Collaboration 2005) and EUCLID (Peacock 2008), are focused on providing good quality photometric data for large portions of the sky.

One way of using the photometric data to obtain clustering information is by estimating photometric redshifts for each of the objects. There are many techniques available for photometric redshift estimation most of which use either template fitting or a training set (*e.g.* Koo 1985; Loh & Spillar 1986; Connolly *et al.* 1995; Bolzonella *et al.* 2000; Collister & Lahav 2004), however the errors associated with these photometric redshifts can be around two orders of magnitude larger than spectroscopic redshifts. Botzler *et al.* (2004) showed that the percolation method of Huchra & Geller (1982) could be expanded for use with photometric redshifts and take into account the larger errors associated with them.

In this chapter the results from chapter 2 are used to calibrate photometric version of the

friends-of-friends algorithm for use on the Mega$z$-LRG catalogue. Mega$z$-LRG is a large subset of luminous red galaxies from the SDSS with photometric redshifts (Collister *et al.* 2007). Mock galaxy catalogues are produced with the intention of constraining the friends-of-friends linking parameters, however it is shown that more realistic mocks would be required to get reliable constraints.

The following section provides further information of the data used for calibration and analysis. Section 3.3 describes the Botzler *et al.* (2004) friends-of-friends method in detail along with modifications that were made. Section 3.4 describes the mock galaxy catalogue produced to constrain the linking parameters. Section 3.5 presents the catalogues produced and their analysis. Finally, section 3.6 shows the results of examining the cluster two-point correlation functions for the 2SLAQ photometric catalogue.

3.2 Data

3.2.1 2SLAQ

The 2dF-SDSS and QSO (2SLAQ) Luminous Red Galaxy Survey is a spectroscopic survey of around 15000 potential luminous red galaxies (LRGs) in the redshift range $0.45 \leq z \leq 0.7$ (Cannon *et al.* 2006). A more detailed description of this catalogue is provided in §2.2. One of the principal benefits of this sample is the availability of both spectroscopic and photometric data for each of the galaxies.

3.2.1.1 Photometric Redshifts

The ANNz program (Collister & Lahav 2004), which is described in detail in §1.4.4.2, was employed to estimate photometric redshifts for the 2SLAQ catalogue. The 2SLAQ spectroscopic data was used to train the neural network to model the dependency between the 2SLAQ magnitudes and photometric redshifts (it should be noted that the 2SLAQ quasar data was not used). This dependency was obtained by using around two thirds of the 2SLAQ catalogue as a mock target sample or evaluation set and using the remaining third as a training set. An architecture of 5 input nodes, 1 output node and 2 hidden layers each with 10 nodes was set up ($N_{\rm in}$:N_1:N_2:$N_{\rm out}$ $\rightarrow$ 5:10:10:1). The 5 inputs from the mocks are the SDSS *u, g, r, i* and z magnitudes. Ten nodes were chosen for the hidden layers to ensure there were twice as many nodes as there were inputs.The accuracy of the photometric redshifts acquired from the evaluation set is characterised by the bias,

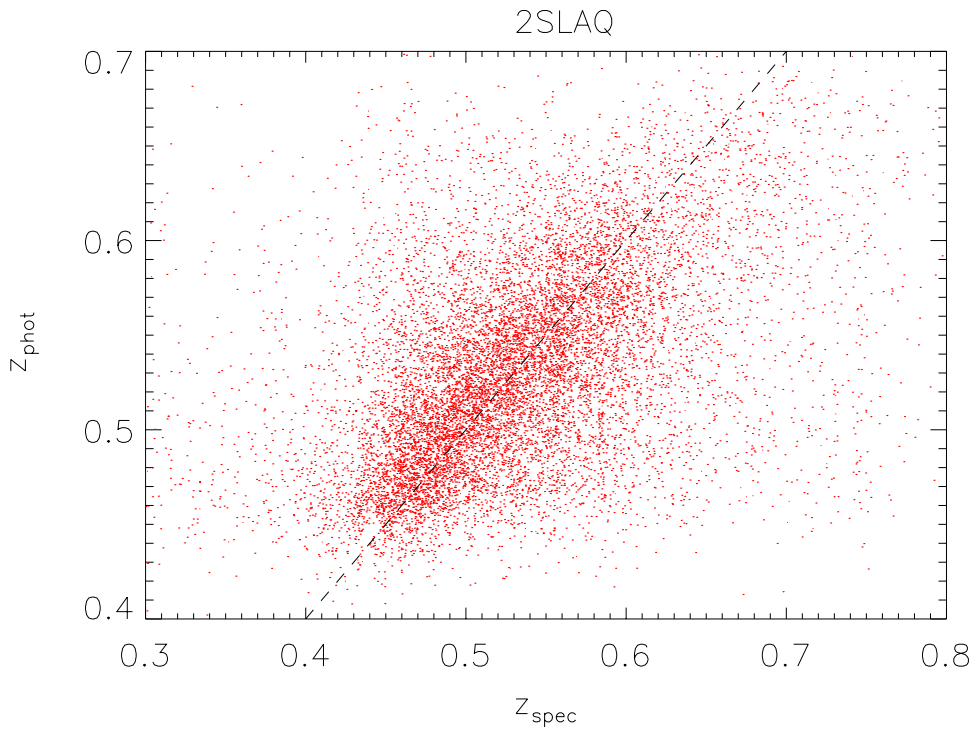


Figure 3.1: 2SLAQ spectroscopic redshifts versus photometric redshifts obtained form ANNz.

δz, and the dispersion, σ_z^2, via

$$\delta z \equiv z_{\rm phot} - z_{\rm spec} \tag{3.1}$$

and

$$\sigma_z^2 = \langle (\delta z)^2 \rangle - \langle \delta z \rangle^2, \tag{3.2}$$

where $z_{\rm phot}$ is the photometric redshift and $z_{\rm spec}$ is the spectroscopic redshift. Fig.3.1 shows the 2SLAQ photometric redshifts as a function of the spectroscopic redshifts.

3.2.2 Mega$z$-LRG

Mega$z$-LRG is a photometric redshift catalogue of over a million SDSS luminous red galaxies (LRGs) in the redshift range 0.4 $\leq$ z $\leq$ 0.7 with limiting magnitude $i < 19.8$ (Collister *et al.* 2007). The LRGs used to make the MegaZ-LRG catalogue were taken from the Sloan Digital Sky Survey data release 7 (SDSS DR7) five-band (u, g, r, i, z; Fukugita *et al.* 1996; Smith *et al.* 2002) imaging data. LRGs are the most massive and most intrinsically luminous galaxies in the Universe. The prominent 4000 Å break in the LRG spectra facilitates the measurement of photometric redshifts for these galaxies (Eisenstein *et al.* 2003).

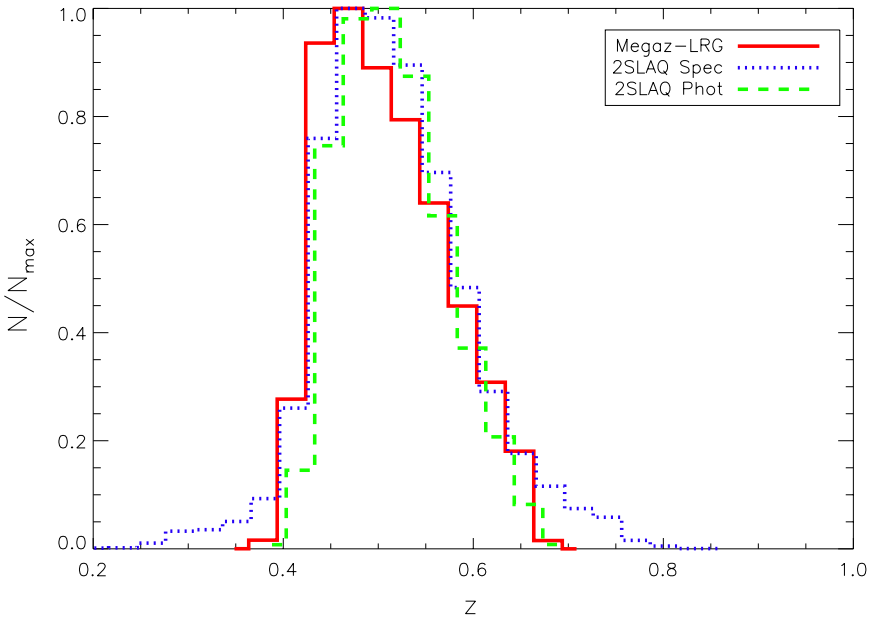


Figure 3.2: Histogram of Mega$z$-LRG DR7 photometric redshifts (red solid line), 2SLAQ spectroscopic redshifts (blue dotted line) and 2SLAQ photometric redshifts (green dashed line). All three lines are normalised to $N/N_{\rm max}$.

3.2.2.1 Photometric Redshifts

Mega$z$-LRG photometric redshifts were estimated using ANNz (Collister & Lahav 2004). The full 2SLAQ catalogue (Cannon *et al.* 2006) was used as the training set. Fig.3.2 shows the $N(z)$ distribution for the Mega$z$-LRG catalogue photometric redshifts.

3.2.2.2 Star-Galaxy Separation

In imaging data it is important to be able to discriminate between stars and galaxies, particularly for extragalactic studies. In general it is assumed that galaxies are more extended sources than stars in images. Several codes have been developed to make the distinction between stars and galaxies using this assumption. Bertin & Arnouts (1996) have shown that neural networks can successfully be applied to this problem.

The 2SLAQ evaluation set contains $\sim 5\%$ stellar contamination. Collister *et al.* (2007) use the 2SLAQ spectroscopy to train the ANNz code to distinguish between stars and galaxies. This reduces the stellar contamination to $\sim 2\%$ in the evaluation set and from this the full Mega$z$-LRG catalogue contains a final column which lists the probability that each object is a galaxy and not star.

3.2.2.3 Catalogue

The complete Mega$z$-LRG DR5 catalogue lists 1,468,028 objects each with ID, right ascension, declination, *u* magnitude, *g* magnitude, *r* magnitude, *i* magnitude, z magnitude, model-*i* magnitude, photometric redshift, error on the photometric redshift, the probability that the object is a galaxy and the error on this probability.

3.3 Photometric Friends-of-Friends Method

The friends-of-friends algorithm discussed in this chapter expands upon that described in the previous chapter for use with photometric redshifts. This method essentially follows that of Botzler *et al.* (2004), the principal difference being, as in the previous chapter, the use of dynamic linking parameters that change with the surface number density of galaxies at a given redshift. This approach does not require any magnitude limit on the galaxy sample, which makes more galaxies available for cluster detection. A fluxogram of the photometric Dynamic Friends-of-Friends (hereafter DFoF) algorithm is shown in fig.3.3.

Following the steps shown in fig.3.3: First the initial values of linking length, R_0, the photometric redshift linking parameter, K, and the minimum number of galaxy members needed to form a group/cluster, N_{min}, are chosen. To compensate for selection effects, R_0 is varied via

$$R_{\text{friend}}(z) \propto R_0 \left(\frac{dN}{dz} \frac{dz}{dV} \frac{1}{A_{\text{sky}}} \right)^{-\frac{1}{2}}, \tag{3.3}$$

where dN/dz is the surface number density of galaxies at each redshift in the range covered by the galaxy catalogue, dV/dz is the differential comoving volume and A_{sky} is the fraction of sky area covered in the catalogue relative to the total sky area. This dynamic linking parameter is required because the fixed parameters used by Botzler *et al.* (2004) require that the sample be volume limited. Due to the colour cuts included in 2SLAQ, volume limiting the catalogue removes a large number of objects that would contribute to the clustering.

The entire catalogue is divided into redshift slices with size, δz_{ini}, of the order of the typical photometric redshift error of the galaxies, δz_i, in the range $z_{\text{min}} \leq z \leq z_{\text{max}}$. For the work presented in this chapter $\delta z_{ini} = 0.01$. Percolation is performed on each redshift slice independently.

A galaxy i is selected from the catalogue that has not yet been assigned to a group and lies in the redshift range $z_{\text{min}} \leq z \leq z_{\text{max}}$. The projected distance between galaxy i and second galaxy j, D_{ij}, is calculated using

$$D_{ij} = \cos^{-1}(\sin(\delta_i)\sin(\delta_j) + \cos(\delta_i)\cos(\delta_j)\cos(\alpha_i - \alpha_j)), \tag{3.4}$$

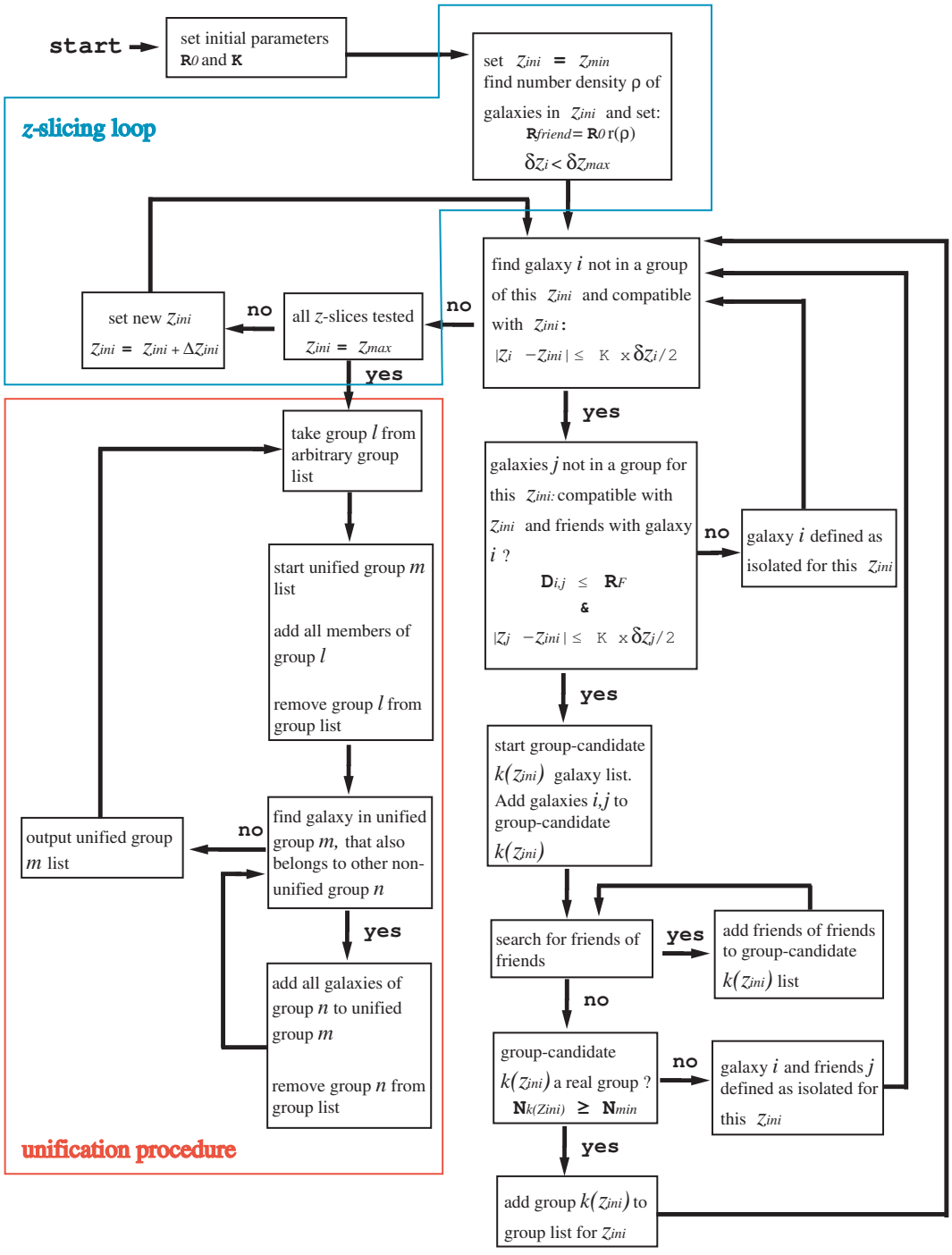


Figure 3.3: Fluxogram of the photometric FoF algorithm with dynamic linking length. The blue and red borders show the additional redshift slicing and merging procedures respectively. Based on Botzler *et al.* (2004).

where α and δ are right ascension and declination respectively.

The two galaxies are linked together (*i.e.* are friends) if they satisfy the conditions

$$D_{ij} \leq R_{\text{friend}}(z) \tag{3.5}$$

and

$$|z_i - z_{ini}| \leq K \times \frac{\delta z_i}{2}, \tag{3.6}$$

where z_i is the photometric redshift of the galaxy and z_{ini} is the redshift of the slice. Equation 3.6 is important because it shows that galaxies can exist in multiple redshift slices and therefore can form part of multiple clusters. This is the principle equation that makes this version of the FoF code applicable to photometric redshifts.

A protogroup-candidate k is formed that includes galaxy i and its friends. A search is then made around the galaxies linked to i. This process is repeated until no further friends are found. The protogroup-candidate k is defined as a real protogroup if it satisfies

$$N_k \geq N_{\text{min}}. \tag{3.7}$$

For the purposes of this work $N_{\text{min}} > 3$. This is a reasonable assumption when dealing with LRGs, which are not common objects.

Once this process have been performed on all redshift slices (*i.e.* $z_{ini} = z_{\text{max}}$) the final merging procedure begins in which a search is carried-out for protogroups with galaxies in common. If a common galaxy is found between two protogroups a final group candidate n is formed, which contains all unique group members from the protogroups.

3.4 Mock Catalogue

As in the previous chapter, a simulated galaxy catalogue was produced to determine the optimal choice of linking parameters, R_0 and K. The halo catalogue was derived from the Horizon 4π simulation (Teyssier *et al.* 2009; Prunet *et al.* 2008). A full description of how the spectroscopic mock catalogue was produced is provided in §2.4.

To make the spectroscopic mock applicable to the 2SLAQ photometric data, photometric redshifts were added to the mock galaxies using

$$z_{\text{phot}} = z_{\text{spec}} + b(z_{\text{spec}}) + erf^{-1}(x)\sigma\sqrt{2} + \mu, \tag{3.8}$$

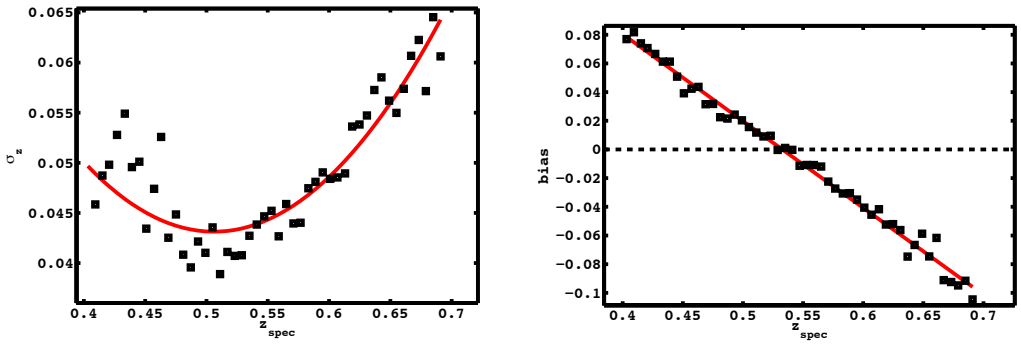


Figure 3.4: Photometric redshift variance of 2SLAQ data in spectroscopic redshift bins (left panel). Bias between photometric versus spectroscopic redshifts and $x = y$ (right panel). Plots produced by Cristiano Sabiu.

where μ and σ are the mean and the variance of the photometric redshifts in a given bin respectively, b is the bias given by $\mu_{\text{phot}} - \mu_{\text{spec}}$, and x is a random number between 0 and 1. $erf^{-1}(x)$ is the inverse error function given by

$$erf^{-1}(x) = \sum_{k=0}^{\infty} \frac{c_k}{2k+1} \left(\frac{\sqrt{\pi}}{2} x \right)^{2k+1}. \tag{3.9}$$

σ was measured by examining the variance of 2SLAQ photometric redshifts, which were obtained with ANNz, in bins of spectroscopic redshift as shown in the left panel of fig.3.4. The bias was obtained from the difference between the mean 2SLAQ photometric and spectroscopic redshifts as seen in the right panel of fig.3.4.

Since the photometric redshifts scatter the positions of the mock galaxies along the line of sight, the number density in the range $0.4 \leq z \leq 0.65$, which is the range of interest for comparison with 2SLAQ, will be altered. To compensate for this effect, galaxies are removed at random in redshift bins where the number density of the mock does not match that of 2SLAQ. This was done in two different ways, producing two different versions of the mock. In the first version, photometric redshifts are added to the galaxies and then objects are removed to maintain the the correct number density. The final version 1 photometric mock catalogue contains 388,029 galaxies. For the second version of the mock, galaxies were removed first and then photometric redshifts were added to the remaining objects. The final version 2 photometric mock catalogue contains 419,452 galaxies. Both mocks contain around half as many objects as the spectroscopic mock catalogue, however the number density in the range $0.4 \leq z \leq 0.65$ is the same as can be seen in fig.3.5. Fig.3.6 shows the distribution of the mock spectroscopic redshifts versus photometric redshifts. The mocks match the

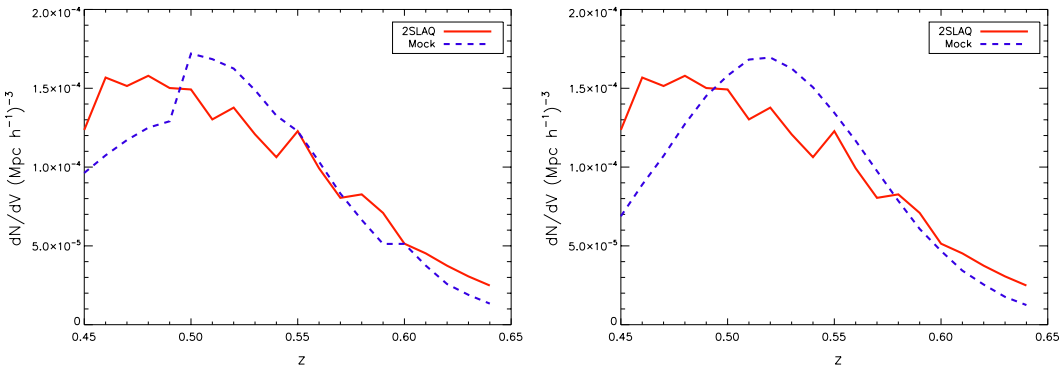


Figure 3.5: Number density as a function of redshift of the photometric mock halo catalogue (blue dashed line) and the 2SLAQ catalogue (red solid line). Mock version 1 (left panel), mock version 2 (right panel).

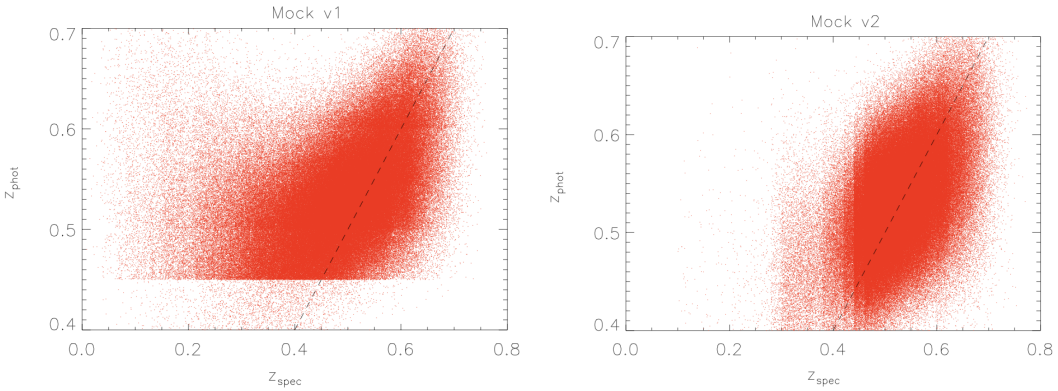


Figure 3.6: Spectroscopic redshifts versus photometric redshifts for 2SLAQ photometric mock catalogue. Mock version 1 (left panel), mock version 2 (right panel).

2SLAQ photometric redshifts in the range $0.4 \leq z \leq 0.65$, however the fit is better in the second version of the mock (right panel).

Both versions of the mock were used in the data analysis to determine what effect, if any, would the way in which galaxies were removed have on the results.

3.5 Data Analysis

3.5.1 Linking Parameter Optimisation

The method for determining the best combination of linking parameters for the photometric DFoF essentially follows that applied in §2.5, however several important differences should be noted.

In the previous chapter, the linking parameters were constrained by comparing cluster cat-

alogues, obtained by running the DFoF code with different combinations of parameters, with simulated haloes. The previous section described how photometric redshifts were included in this same mock catalogue, producing two different versions of a photometric mock catalogue. It should be noted that these mocks mimic the number density of 2SLAQ and cannot be directly applied to Mega$z$-LRG. The reason being that Mega$z$-LRG has a much higher number density of objects (1,939,760 galaxies within $\sim$8000 deg^2) than 2SLAQ (13,133 galaxies within $\sim$145 deg^2) and therefore the linking parameters for each catalogue must be constrained independently. It would be necessary to produce a mock that mimics the Mega$z$-LRG number density to preform the same analysis, however for the purposes of this project this approach was not pursued. The reason for this is that the mock would need to be more realistic than that produced in §3.4 to provide useful information about the completeness and purity of the linking parameters, which would require a lot more time than can currently be applied to this project. This point is explained in greater detail in the following subsections.

Another important difference between this chapter and the previous one is that the spectroscopic 2SLAQ cluster catalogue presented in chapter 2 serves as a sort of 'truth table' for a cluster catalogue obtained using the 2SLAQ photometric redshifts. This provides an additional method for constraining the linking parameters. In addition, since all of the 2SLAQ galaxies are included in the Mega$z$-LRG catalogue, the spectroscopic 2SLAQ cluster catalogue also provides a way of determining the best set of linking parameters for Mega$z$-LRG. As the right combination of parameters should reproduce the majority of the same structures.

Finally, as no mock analysis is performed for Mega$z$-LRG, the linking parameters for this catalogue can further be constrained simply by examining the physical properties of the resulting clusters from a given combination of R_0 and K.

3.5.1.1 2SLAQ

The DFoF code was run on the the two versions of the photometric mocks varying R_0 and K in order to determine the combination of the two parameters that best reproduce the simulated haloes. DFoF was additionally run with a range of R_0 and K values on the 2SLAQ photometric data in an effort to find a set of parameters that reproduce the 2SLAQ spectroscopic cluster catalogue. To maintain a more physical interpretation of the friend radius, R_0 is hereafter expressed in terms of $R_{\rm friend}(0.5)$ according to equation 3.3.

In order to determine the completeness and purity of the DFoF clusters relative to the simulation haloes, a membership matching code was implemented. The technique involves looking at the

galaxy members assigned to each cluster by the DFoF code and matching these to the original halo member galaxies. The clusters are examined in descending order of richness to ensure that the largest clusters are the first to be matched to mock haloes. DFoF clusters may contain contributions from several mock haloes, therefore each cluster is matched to the mock halo with the highest number of shared members. Thus, if two haloes are merged into one cluster only one of the two haloes will be matched. Each cluster is uniquely matched to one halo and any cluster that corresponds to a halo that has already been matched, which will be of equal or lesser richness, will be ignored. The completeness and purity are defined via

$$\text{Completeness} = \frac{N_{\text{matches}}}{N_{\text{haloes}}} \tag{3.10}$$

and

$$\text{Purity} = \frac{N_{\text{matches}}}{N_{\text{clusters}}}, \tag{3.11}$$

where N_{matches} is the total number of unique matches, N_{clusters} is the total number of DFoF clusters found and N_{haloes} is the total number of mock haloes. N_{haloes} is calculated ignoring all mock haloes with less than 3 galaxy members as this is the detection limit of the DFoF code. It should be noted that both the completeness and the purity are functions of the richness of the objects in the 'true' halo catalogue and the DFoF cluster catalogue respectively. In this section both the 2SLAQ spectroscopic catalogue and the mocks are considered to be true halo catalogues.

Fig.3.7 shows the number of clusters detected in the 2SLAQ photometric redshift catalogue (top left panel), and the unique completeness (top right panel) and purity (middle left panel) of the photometric clusters relative to 2SLAQ spectroscopic groups and clusters all as a function of $\text{R}_{\text{friend}}(\text{z} = 0.5)$ and K. This figure also shows the unique completeness for groups and clusters matched to 2SLAQ spectroscopic groups and clusters with $N_{\text{mem}} \geq 6$ (middle right panel) and the unique completeness (bottom left panel) and purity (bottom right panel) relative to the 2SLAQ gold sample groups and clusters (see §2.5.5). This figure indicates that the maximum number of clusters is found around $R_{\text{friend}}(0.5) = 1.8\ h^{-1}$Mpc. All three completeness plots show the same distribution in which the completeness peaks around $R_{\text{friend}}(0.5) = 0.8\ h^{-1}$Mpc then dips and peaks again around $R_{\text{friend}}(0.5) = 1.2\ h^{-1}$Mpc. This indicates that beyond the initial peak structures are merging together lowering the completeness before smaller structures are formed increasing it again. The first peak corresponds to a cluster catalogue that is 38% complete and 25% pure relative to the 2SLAQ spectroscopic cluster catalogue (this includes all objects with $N_{\text{mem}} \geq 3$). The completeness can be improved to around 80% if matching is only made with

2SLAQ groups and clusters with six or more members or to around 70% if the groups and clusters are compared to the 2SLAQ gold sample clusters. Both purity plots also show a similar distribution in which the purity peaks around $R_{\text{friend}}(0.5) = 0.2\ h^{-1}$Mpc for K values between 1.4 and 1.8 and the drops before increasing again around $R_{\text{friend}}(0.5) = 3.5\ h^{-1}$Mpc. This later increase in purity is the result of mass merging of the clusters that eventually reaches a point at which the cluster catalogue contains only a few clusters that include all the galaxies in the galaxy catalogue.

Fig.3.8 shows the number of clusters detected in both versions of the mock photometric redshift catalogue (top panels), and the unique completeness (middle panels) and purity (bottom panels) of the DFoF clusters relative to mock haloes all as a function of $R_{\text{friend}}(z = 0.5)$ and K. Some parameter combinations were not run for the first version of the mock because the clusters merged into highly unphysical structures, this is the cause of the sharp borders seen in the left panels. This figure clearly shows that the completeness is significantly affected by the way in which the mock is produced as the levels of completeness differ quite significantly from one mock to another. Additionally, there seems to be little correlation with the results from the 2SLAQ photoz matching. The first version of the mock peaks in completeness at $R_{\text{friend}}(0.5) = 0.15\ h^{-1}$Mpc and $K = 1.8$ and the second version of the mock peaks initially at $R_{\text{friend}}(0.5) = 0.25\ h^{-1}$Mpc and $K = 1.2$ and then again at $R_{\text{friend}}(0.5) = 0.8\ h^{-1}$Mpc and $K = 1.5$. The first two sets of parameters when run on the real 2SLAQ photometric data produce zero clusters. The last set of parameters ($R_{\text{friend}}(0.5) = 0.8\ h^{-1}$Mpc and $K = 1.5$) coincides with the 2SLAQ photoz matching results ($R_{\text{friend}}(0.5) = 0.8\ h^{-1}$Mpc and $K = 1.8$), however perhaps only coincidentally.

These results indicate that the mocks are too basic to reveal the true levels of completeness and purity for the 2SLAQ photometric data. A decent degree of correlation would be expected between the mocks and the 2SLAQ photometric catalogue since the number densities of both catalogues are roughly the same. This can partially be attributed to the simplistic manner in which the photometric redshifts were assigned to the galaxies, however the way in which objects are removed from the photometric mocks appears to have a significant effect on the clustering. A more sophisticated mock would be required to improve the results from this analysis. Ideally this mock would contain realistic colours for each of the galaxies in the catalogue, which would allow photometric redshifts to estimated directly. Also the way in which the photometric mock is diluted to match the number density of the spectroscopic mock would have to be done in way that does not significantly effect the clustering. For purposes of this project it was decided to leave a more detailed mock analysis for a later study and therefore no mocks were created to simulate the Mega$z$-LRG catalogue in this chapter.

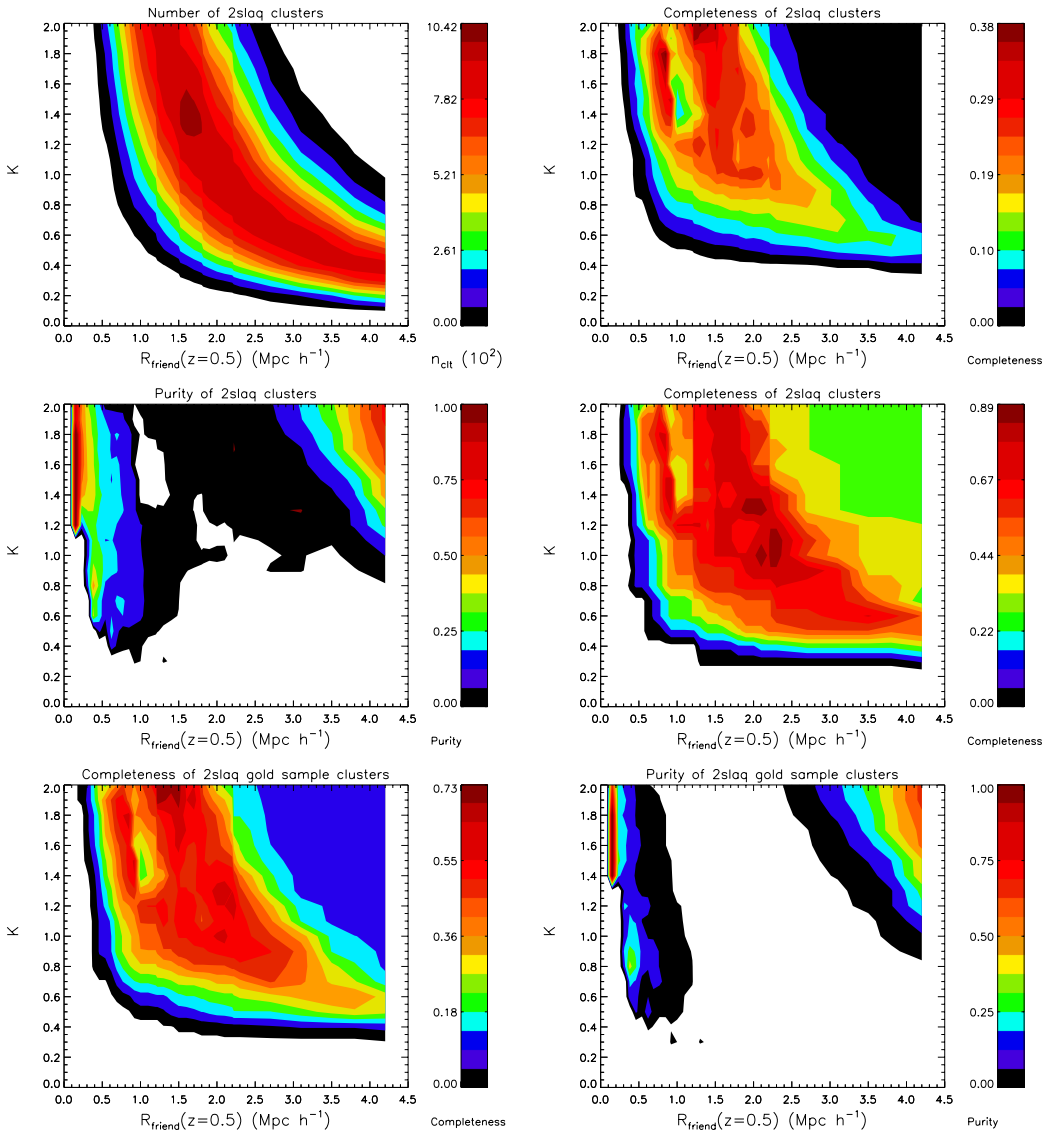


Figure 3.7: Number of clusters detected in the 2SLAQ photometric redshift catalogue as a function of $R_{friend}(z = 0.5)$ and K (top left panel). Unique completeness (top right panel) and purity (middle left panel) of the photometric clusters relative to 2SLAQ spectroscopic groups and clusters as a function of $R_{friend}(z = 0.5)$ and K. Unique completeness for groups and clusters with $N_{mem} \geq 6$ (middle right panel) and completeness (bottom left panel) and purity (bottom right panel) of 2SLAQ photometric redshift catalogue relative to 2SLAQ spectroscopic gold sample groups and clusters.

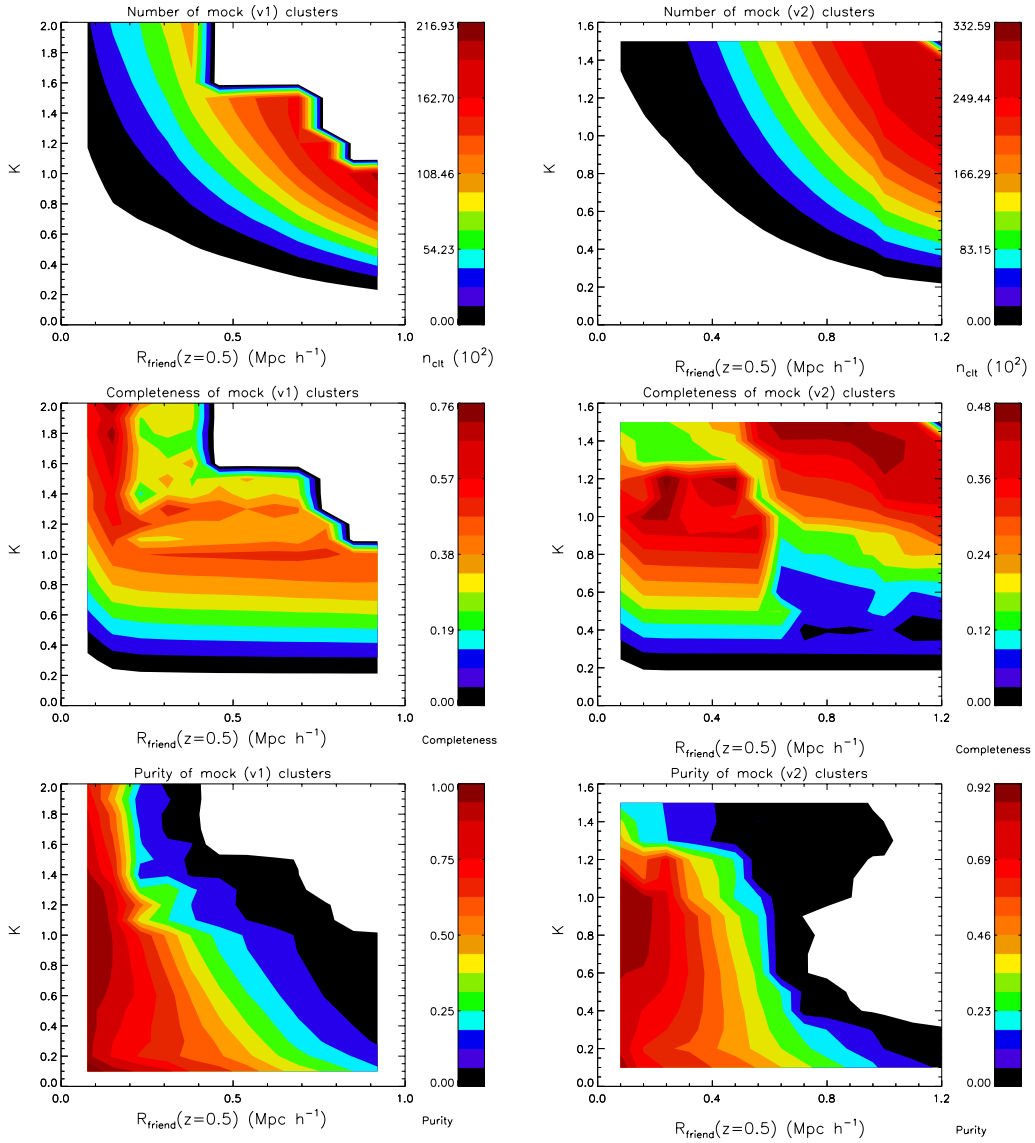


Figure 3.8: Number of clusters (top panels) detected in the mock photometric redshift catalogues (v1 mock left panels, v2 mock right panels) as a function of $R_{friend}(z = 0.5)$ and K. Unique completeness (middle panels) and purity (bottom panels) of the FoF clusters relative to mock haloes as a function of $R_{friend}(z = 0.5)$ and K.

Using only the results from the 2SLAQ photometric cluster analysis, the parameter set $R_{\text{friend}}(0.5) = 0.8\ h^{-1}$Mpc and $K = 1.8$ was chosen for finding groups and clusters in the 2SLAQ photometric catalogue as the resulting catalogue contains the most realistic looking structures out of the two peaks in fig.3.7.

3.5.1.2 Mega$z$-LRG

As the mock results for the 2SLAQ photometric catalogue were unreliable, for the Mega$z$-LRG catalogue no mock analysis is performed. To determine the best choice of FoF parameters for Mega$z$-LRG an analysis is made of the resulting physical cluster properties from a given combination of $R_{\text{friend}}(z = 0.5)$ and K.

Fig.3.9 shows the number of clusters detected in the Mega$z$-LRG DR7 catalogue, the richness of the largest cluster, the size of the largest cluster and the average size of all clusters as a function of $R_{\text{friend}}(z = 0.5)$ and K. A threshold of 100 LRG members is chosen to limit the production of unphysical Mega$z$-LRG clusters, which is indicated with a dashed line in the top right panel of the figure. The maximum physical size of the largest cluster is limited to a few hundred h^{-1}Mpcs, which is indicated with a dot-dashed line in the bottom left panel of the figure. This is well above the expected size of the largest cluster, but this soft limit allows for some merging of structures in the cluster catalogue. It should also be noted that this maximum value is not representative of the average size of structures in the cluster catalogue. Finally, the average size of all clusters in the catalogue is limited to around 1 h^{-1}Mpc, which is indicated with a dotted line in the bottom left panel of the figure.

To gain some understanding of the levels of completeness and purity for a given set of $R_{\text{friend}}(z = 0.5)$ and K in Mega$z$-LRG, a cut version of the catalogue was produced, which contains only Mega$z$-LRG galaxies within the 2SLAQ area as shown in fig.3.10. The groups and clusters found within this area can then be compared with the 2SLAQ spectroscopic cluster catalogue (as all 2SLAQ galaxies are contained within Mega$z$-LRG) in order to get an idea of the completeness and purity as a function of $R_{\text{friend}}(z = 0.5)$ and K. The combination of parameters for Mega$z$-LRG will differ from those of 2SLAQ photoz, as the Mega$z$-LRG catalogue has a much higher number density of objects. 2SLAQ contains around 13,000 objects within 145 deg^2, while Mega$z$-LRG DR7 contains around 2,000,000 objects within an area of around 8000 deg^2. This difference arrises because the 2SLAQ catalogue was produced by selecting LRGs at random from this area and it is therefore not complete.

Fig.3.11 shows the number of Mega$z$-LRG clusters detected in the 2SLAQ area (top left panel),

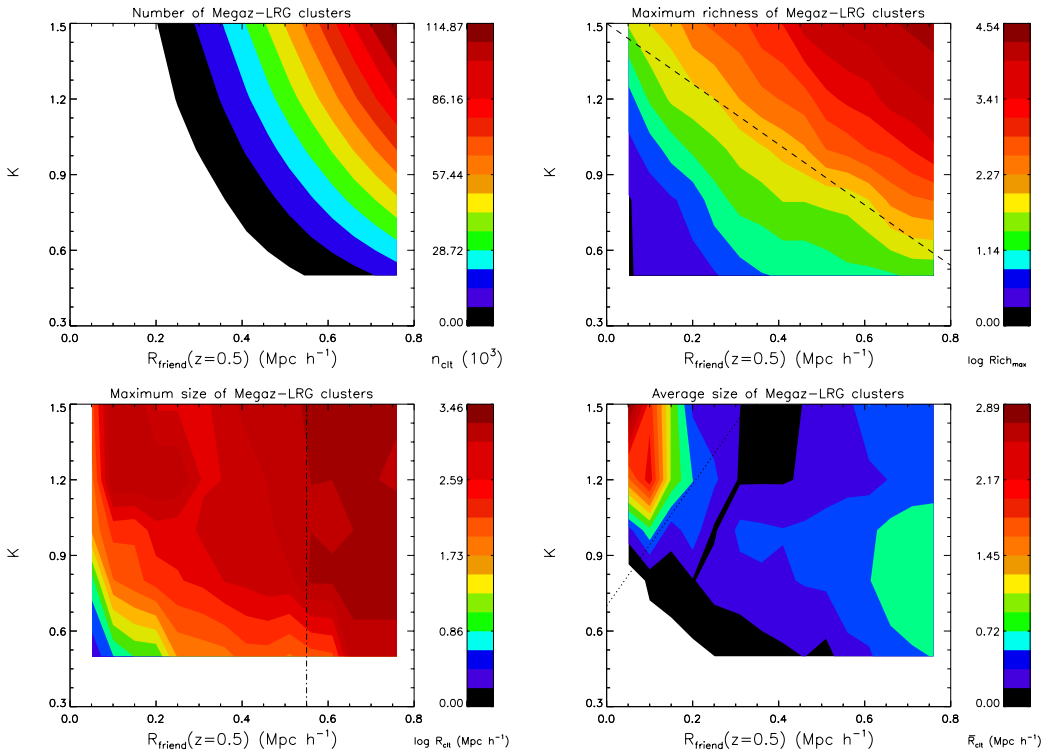


Figure 3.9: Number of clusters detected in the Mega$z$-LRG catalogue (top left panel), the richness of the largest cluster (top right panel), the size of the largest cluster (bottom left panel) and the average size of all clusters as a function of $R_{friend}(z = 0.5)$ and K (bottom right panel).

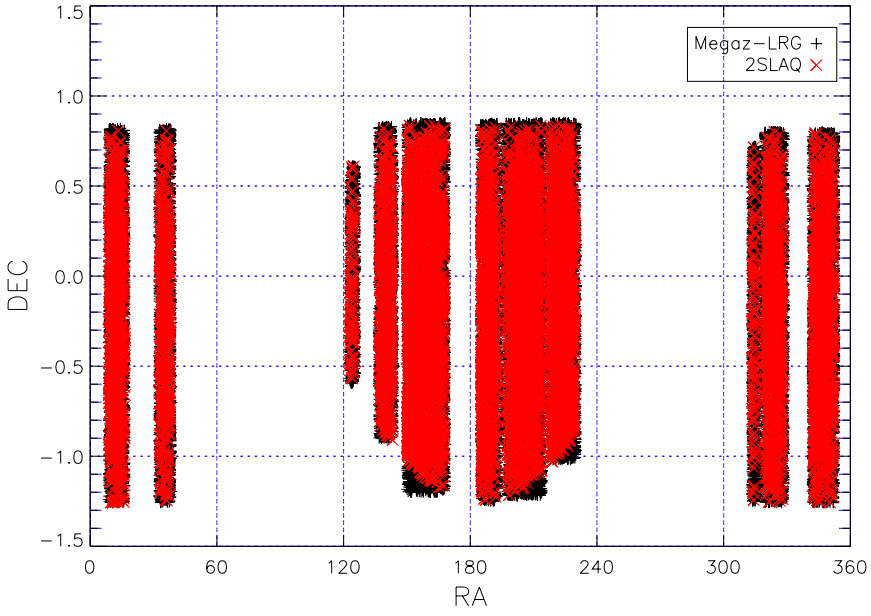


Figure 3.10: Distribution of Mega$z$-LRG galaxies within the 2SLAQ area.

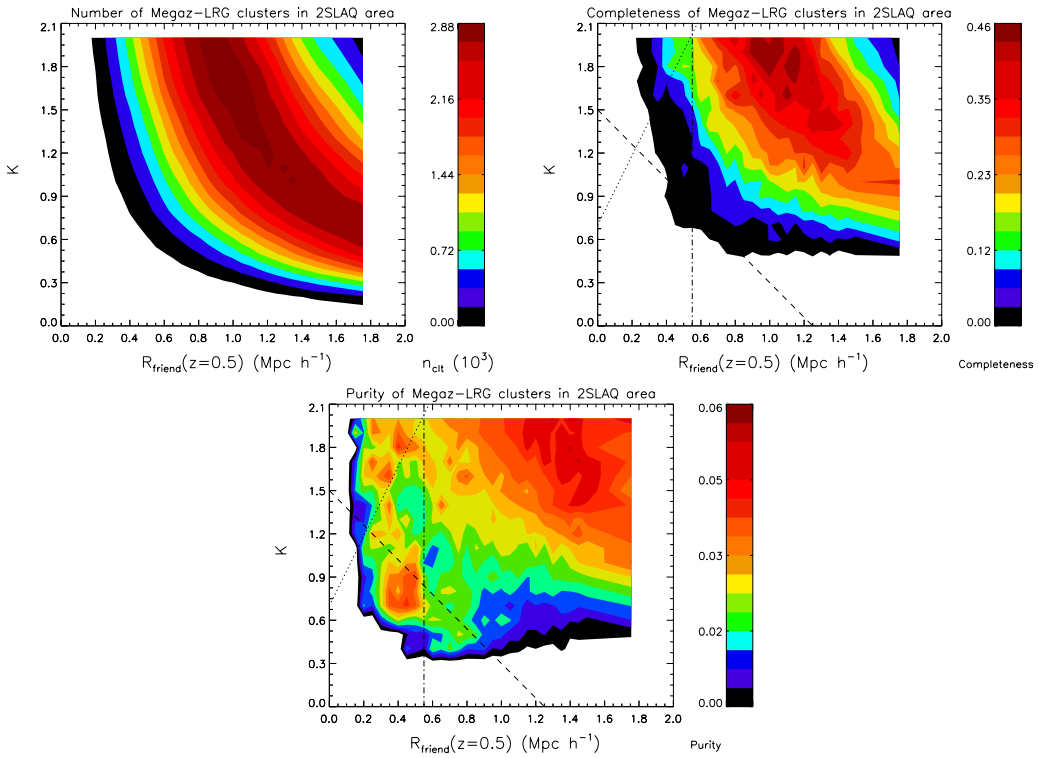


Figure 3.11: Number of Mega$z$-LRG clusters detected in the 2SLAQ area (top left panel), and the unique completeness (top right panel) and purity (bottom panel) of the groups and clusters relative to 2SLAQ spectroscopic cluster catalogue all as a function of R_{friend}(z = 0.5) and K.

and the unique completeness (top right panel) and purity (bottom panel) of the groups and clusters relative to 2SLAQ spectroscopic cluster catalogue all as a function of R_{friend}(z = 0.5) and K. The limits imposed from fig.3.9 are displayed on top of the completeness and purity plots. This figure clearly shows that point at which the cluster catalogue satisfies the limits imposed from 3.9 and the point at which it is complete are inconsistent.

Based on these results, two Mega$z$-LRG cluster catalogues are chosen for further analysis. The first catalogue corresponds to parameters $R_{friend}(z = 0.5) = 0.5h^{-1}$Mpc and $K = 0.8$ (Mega$z$-LRG v1), which falls within the limits imposed from fig.3.9, but has very low completeness ($\sim 5\%$ matching to $N_{mem} \geq 3$ or $\sim 33\%$ matching to $N_{mem} \geq 6$). The second catalogue corresponds to parameters $R_{friend}(z = 0.5) = 0.75h^{-1}$Mpc and $K = 1.5$ (Mega$z$-LRG v2), which produces more unphysical structures but has a much higher completeness ($\sim 35\%$ matching to $N_{mem} \geq 3$ or $\sim 72\%$ matching to $N_{mem} \geq 6$) according to fig.3.11.

The choice of two different catalogues can be justified for the following reasons. Firstly, it may be necessary to produce some highly unphysical structures due to merging in order to obtain the

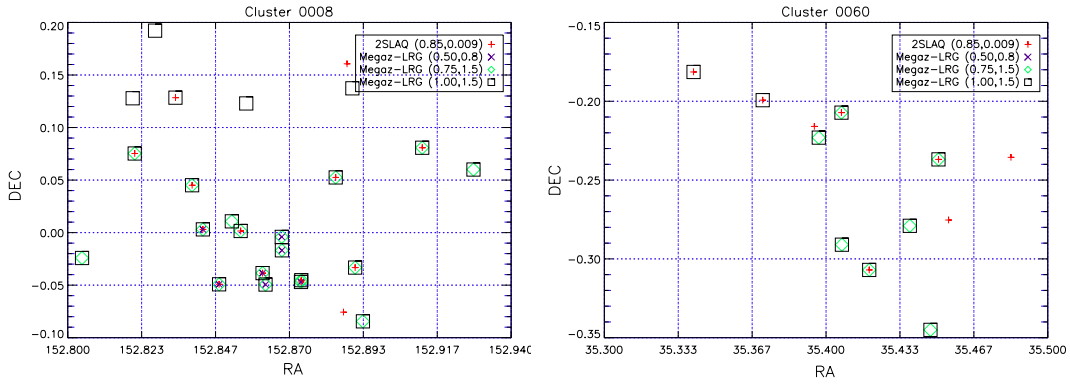


Figure 3.12: Angular positions of the 2SLAQ spectroscopic cluster member galaxies for clusters 0008 and 0060 (red crosses), and member galaxies for Mega$z$-LRG cluster catalogues with $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 0.5h^{-1}$Mpc and $K = 0.8$ (blue 'X's), $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 0.75h^{-1}$Mpc and $K = 1.5$ (green diamonds) and $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 1.0h^{-1}$Mpc and $K = 1.5$ (black squares) within same area.

desired level of completeness in the cluster catalogue. This may be due to the high number density of objects in Mega$z$-LRG. A secondary cluster analysis of the substructures of these extremely large clusters may reveal the true objects. Secondly, it is also difficult to discern how well matching Mega$z$-LRG photometric clusters to 2SLAQ spectroscopic clusters will reveal the true amount of completeness and purity in the catalogue. Therefore examining the statistical properties of both catalogues may help to identify which constraints produce the most realistic cluster catalogue. These points are more clearly illustrated in the following subsection.

3.5.1.3 Case Study

Fig.3.12 shows the angular positions of the 2SLAQ spectroscopic cluster member galaxies for clusters 0008 and 0060 (red crosses). This catalogue was produced using parameters $R_{\mathrm{friend}}(z = 0.5) = 0.87h^{-1}$Mpc and $v_{\mathrm{friend}}(z = 0.5) = 900$ kms^{-1} in the previous chapter (chapter 2). Member galaxies for Mega$z$-LRG cluster catalogues with $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 0.5h^{-1}$Mpc and $K = 0.8$ (blue 'X's), $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 0.75h^{-1}$Mpc and $K = 1.5$ (green diamonds) and $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 1.0h^{-1}$Mpc and $K = 1.5$ (black squares) are over plotted within the same area.

This figure shows that in the case of cluster 0008 the Mega$z$-LRG catalogue with parameters $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 0.5h^{-1}$Mpc and $K = 0.8$ identifies four out of the eleven members, which is sufficient for identifying the structure. The same Mega$z$-LRG catalogue does not identify any members of cluster 0060, however the catalogue with parameters $\mathrm{R}_{\mathrm{friend}}(z = 0.5) = 0.75h^{-1}$Mpc

and $K = 1.5$ identifies three out of the nine members. This clearly shows that there is a higher completeness for the larger parameter set, although the catalogue contains some highly unphysical structures as seen in fig.3.9. The richest object found in the version 1 Mega$z$-LRG cluster catalogue contains 89 LRGs and has a size of 2.23 h^{-1}Mpc, which is fairly reasonable considering the number of objects. This set of parameters marks the point at which cluster merging is only just starting and therefore the structures are much more pure, but also much less complete. The richest object found in version 2 contains 35,023 LRGs and has a size of 1442.25 h^{-1}Mpc, which is highly unphysical and is clearly the result of over-merging. It should be noted that this object is not representative of the entire catalogue and is in no way considered to be a genuine cluster.

3.5.2 Basic Results

Running the photometric DFoF algorithm using a linking length of $R_{\text{friend}}(z = 0.5) = 0.85$ Mpc and a photometric redshift linking parameter of $K = 1.8$ on the 13,133 galaxies in the 2SLAQ photometric redshift catalogue produced a total of 569 groups and clusters containing 2,180 member galaxies. Fig.3.13 shows the angular position of the centres of these groups and clusters (top left panel), their distribution with respect to redshift (top right panel), the distribution of cluster sizes, R_{clt}, as a function of redshift (bottom left panel) and a histogram of the cluster sizes (bottom right panel). Where R_{clt} is defined as the average projected distance from the cluster centre of the galaxy members. The 'x's indicate clusters with three members, the blue circles indicate clusters with between four and six members and the red square indicate clusters with seven or more members. These plots illustrate that on average the richest clusters have the largest sizes as expected. Around 65% of all the groups and clusters have sizes less than 1 h^{-1}Mpc, which is comparable with $\sim 77\%$ of spectroscopic 2SLAQ clusters, which implies that the majority of the clusters are virialised structures. A general increase can be seen in the average cluster sizes in the range $0.65 \leq z \leq 0.70$, however this is somewhat expected as $z \approx 0.70$ is the limit of the 2SLAQ catalogue.

Running the photometric DFoF algorithm using a linking length of $R_{\text{friend}}(z = 0.5) = 0.5$ h^{-1}Mpc and a photometric redshift linking parameter of $K = 0.8$ on the 1,939,760 galaxies in the Mega$z$-LRG DR7 photometric redshift catalogue produced a total of 21,255 groups and clusters containing 73,073 member galaxies. While running the code using linking parameters of $R_{\text{friend}}(z = 0.5) = 0.75$ Mpc and $K = 1.5$ produced a total of 114,874 groups and clusters containing 544,008 member galaxies. Fig.3.14 and fig.3.15 show the angular position of the centres of these groups and clusters (top left panel), their distribution with respect to redshift (top right panel), the distribution of cluster sizes, R_{clt}, as a function of redshift (bottom left panel) and a

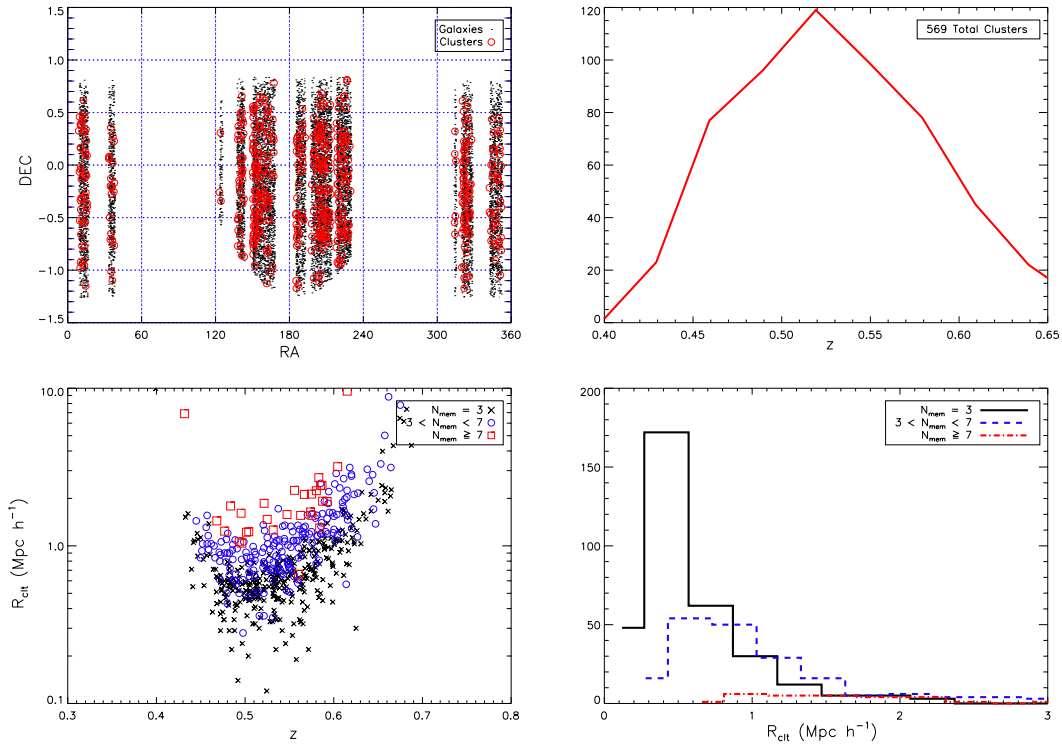


Figure 3.13: Angular position of the 2SLAQ photometric cluster centres (top left panel), their distribution with respect to redshift (top right panel), the distribution of the cluster sizes, R_{clt}, as a function of redshift (bottom left panel) and a histogram of the cluster sizes (bottom right panel).

histogram of the cluster sizes (bottom right panel). Around 97% of the $R_{\rm friend}(z = 0.5) = 0.5$ h^{-1}Mpc and $K = 0.8$ Mega$z$-LRG clusters, and $\sim 94\%$ of the $R_{\rm friend}(z = 0.5) = 0.75$ h^{-1}Mpc and $K = 1.5$ clusters have sizes less than 1 h^{-1}Mpc. This is not surprising considering the much higher number of objects detected, but does clearly show that the unphysical structures found are a very small fraction of the total number of clusters. In addition, these figures show that for both mocks the majority of larger clusters are found at the edges of the redshift distribution.

The results of fig.3.14 and fig.3.15 indicate that the Mega$z$-LRG catalogue with linking parameters $R_{\rm friend}(z = 0.5) = 0.75$ h^{-1}Mpc and $K = 1.5$ may be the better choice, as the catalogue is more complete and at most only $\sim 6\%$ of the clusters are unphysical.

3.5.3 Background Subtraction

As in the previous chapter, the DFoF groups and clusters for the three catalogues (2SLAQ photometric catalogue, and Mega$z$-LRG catalogues v1 and v2) were examined by looking for peaks in the photometric redshift distribution and colour-magnitude relation of SDSS galaxies around the

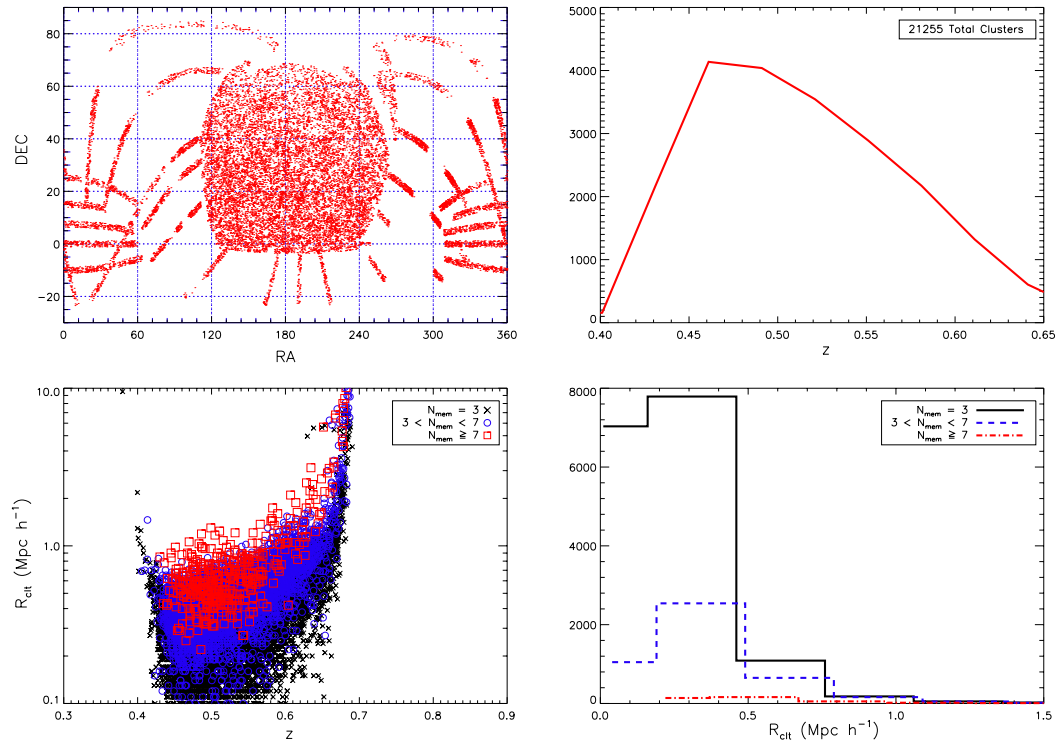


Figure 3.14: Angular position of the Mega$z$-LRG DR7 (with linking parameters $R_{\rm friend}(z = 0.5) = 0.5\ h^{-1}$Mpc and $K = 0.8$) photometric cluster centres (top left panel), their distribution with respect to redshift (top right panel), the distribution of the cluster sizes, $R_{\rm clt}$, as a function of redshift (bottom left panel) and a histogram of the cluster sizes (bottom right panel)

cluster centres. Using an SQL query regions of one square degree and centred on the FoF cluster candidates were downloaded from the SDSS DR7 catalogue. These squares were downloaded for each object in each catalogue and include all SDSS galaxies with photometric redshifts. The properties downloaded for each galaxy were: id, ra, dec, u-magnitude, g-magnitude, r-magnitude, i-magnitude, z-magnitude, model i-magnitude, photometric redshift, photometric redshift error and star likelihood. In each of these squares all galaxies within a 1 h^{-1}Mpc radius of the centre were taken as cluster members with background, regardless of redshift, and all galaxies between 3 and 7 h^{-1}Mpc were used as field galaxies. Galaxies between 1 and 3 h^{-1}Mpc were ignored. These limits were chosen because, on average, one would not expect galaxies father than 3 h^{-1}Mpc from a real cluster centre to be genuine members. Using the DFoF cluster redshifts, the data from the corresponding SDSS squares were stacked in redshift slices of $\Delta z = 0.05$ in the range $0.4 \leq z \leq 0.7$. Due to constrains on time and storage space background subtraction was only performed for Mega$z$-LRG v1 clusters with $N_{\rm mem} \geq 6$ and Mega$z$-LRG v2 clusters with $N_{\rm mem} \geq 10$, which corresponds to 860 and 5,088 objects respectively, where $N_{\rm mem}$ is the number of galaxy

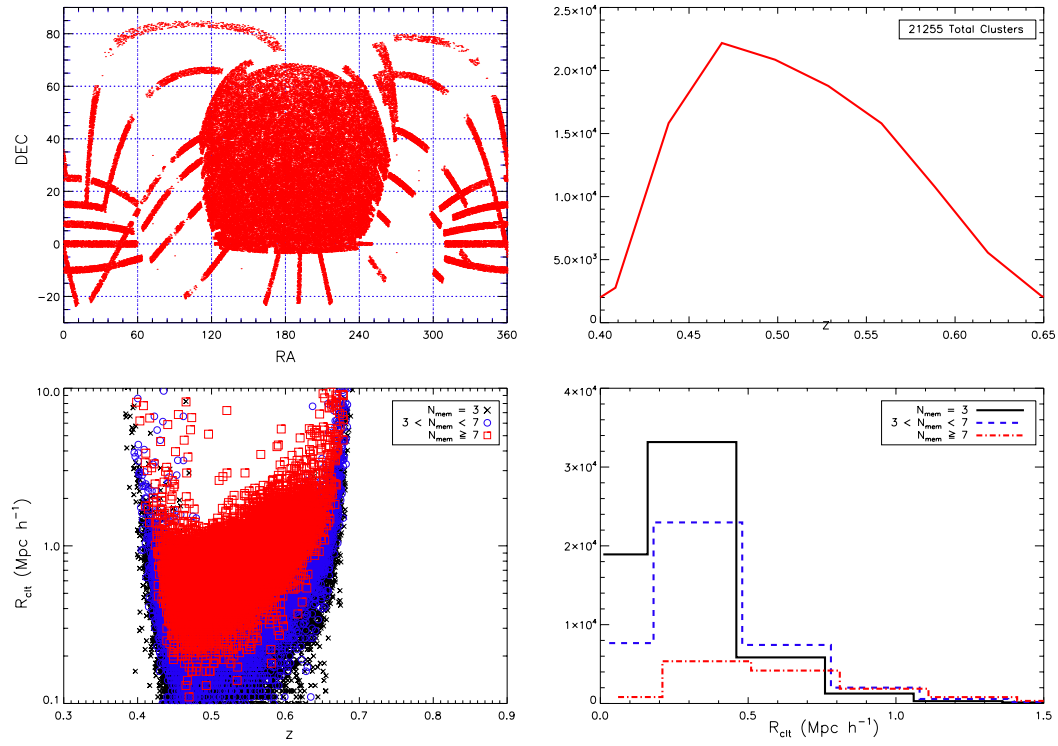


Figure 3.15: Angular position of the Mega$z$-LRG DR7 (with linking parameters $R_{\rm friend}(z = 0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$) photometric cluster centres (top left panel), their distribution with respect to redshift (top right panel), the distribution of the cluster sizes, $R_{\rm clt}$, as a function of redshift (bottom left panel) and a histogram of the cluster sizes (bottom right panel)

members in a group or cluster (*i.e.* the richness).

In each cluster redshift bin the stacked SDSS galaxies were binned by photometric redshift. The number of SDSS galaxies assigned to the field was then subtracted from the number of SDSS galaxies assigned as cluster members plus background in each redshift bin. This subtraction was done taking into account the difference in area between the two regions. Fig.3.16, fig.3.17 and fig.3.18 show the number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters in the 2SLAQ photoz catalogue and the Mega$z$-LRG catalogues respectively. The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies. In these plots the background subtraction should average out to zero if there is no overdensity in the field. Peaks can been identified in each of the plots at the cluster redshift range. These peaks are not as pronounced as those seen in fig.2.10 from the 2SLAQ spectroscopic catalogue, however this remains a strong indication that the DFoF cluster candidates are genuine structures. In fig.3.16, like in fig.2.10, peaks can clearly be seen out to a redshift $z \sim 0.6$ after which the plots become noisy, however fig.3.17

and fig.3.18 still show evidence of a peak at the cluster redshift in the range $0.6 \leq z \leq 0.65$. This is because a much larger number of objects at higher redshifts have been identified in the Mega$z$-LRG cluster catalogues. The noise seen above this redshift range can partially be attributed to the quality of the SDSS photometric redshifts.

To look for the colour-magnitude relation of the groups and clusters in each cluster catalogue, i-magnitude vs $(g-i)$-colour maps were made for both field galaxies and cluster+field galaxies in each SDSS square. The C-M maps were then stacked by cluster redshift in redshift slices. Finally the stacked field galaxy maps were subtracted from the stacked cluster galaxy maps taking into account the relative areas. Fig.3.19, fig.3.20 and fig.3.21 show the background subtracted C-M diagrams for SDSS galaxies. The maps show a trend in colour-magnitude space that resembles a cluster red sequence. The cluster red sequence is an observational property whereby cluster galaxies are more red than field galaxies at the same redshift. Similarly to fig.3.16, fig.3.19 shows a less noticeable trend than that see in fig.2.11 with the same cut-off at $z \sim 0.6$. Fig.3.20 and fig.3.21 show a much sharper trend (~ 0.3 in $g-i$) than fig.3.19 that becomes more noisy with redshift. These figures are another strong indication that a large fraction of the objects in the cluster catalogues are genuine structures.

The redshift peaks and red sequence track appears more prominent in the Mega$z$-LRG v2 cluster catalogue, which indicates that the catalogue is more complete than the Mega$z$-LRG v1 cluster catalogue. This implies that the completeness levels shown in fig.3.11 are, to a certain degree, representative of reality and that it may be better to produce a more complete and unphysical catalogue that can be reanalysed later on.

3.5.4 Radial Profile

Fig.3.22 shows the radial profile of 2SLAQ spectroscopic cluster galaxies (red solid line), Mega$z$-LRG photometric cluster catalogue members, with linking parameters $R_{\text{friend}}(z = 0.5) = 0.5$ h^{-1}Mpc and $K = 0.8$ (blue dashed line), and Mega$z$-LRG photometric cluster catalogue members, with linking parameters $R_{\text{friend}}(z = 0.5) = 0.75$ h^{-1}Mpc and $K = 1.5$ (green dot-dashed line). The red triple dot-dashed lines show the the 2SLAQ spectroscopic radial profile multiplied by the ratios of the number of 2SLAQ spectroscopic cluster galaxies (1,152) to the number of Mega$z$-LRG cluster galaxies (73,078 for v1, 544,022 for v2).

This provides further evidence that the Mega$z$-LRG v2 cluster catalogue uses a better choice of linking parameters as the 2SLAQ profile, when normalised, appears to be a good fit to the Mega$z$-LRG v2 profile.

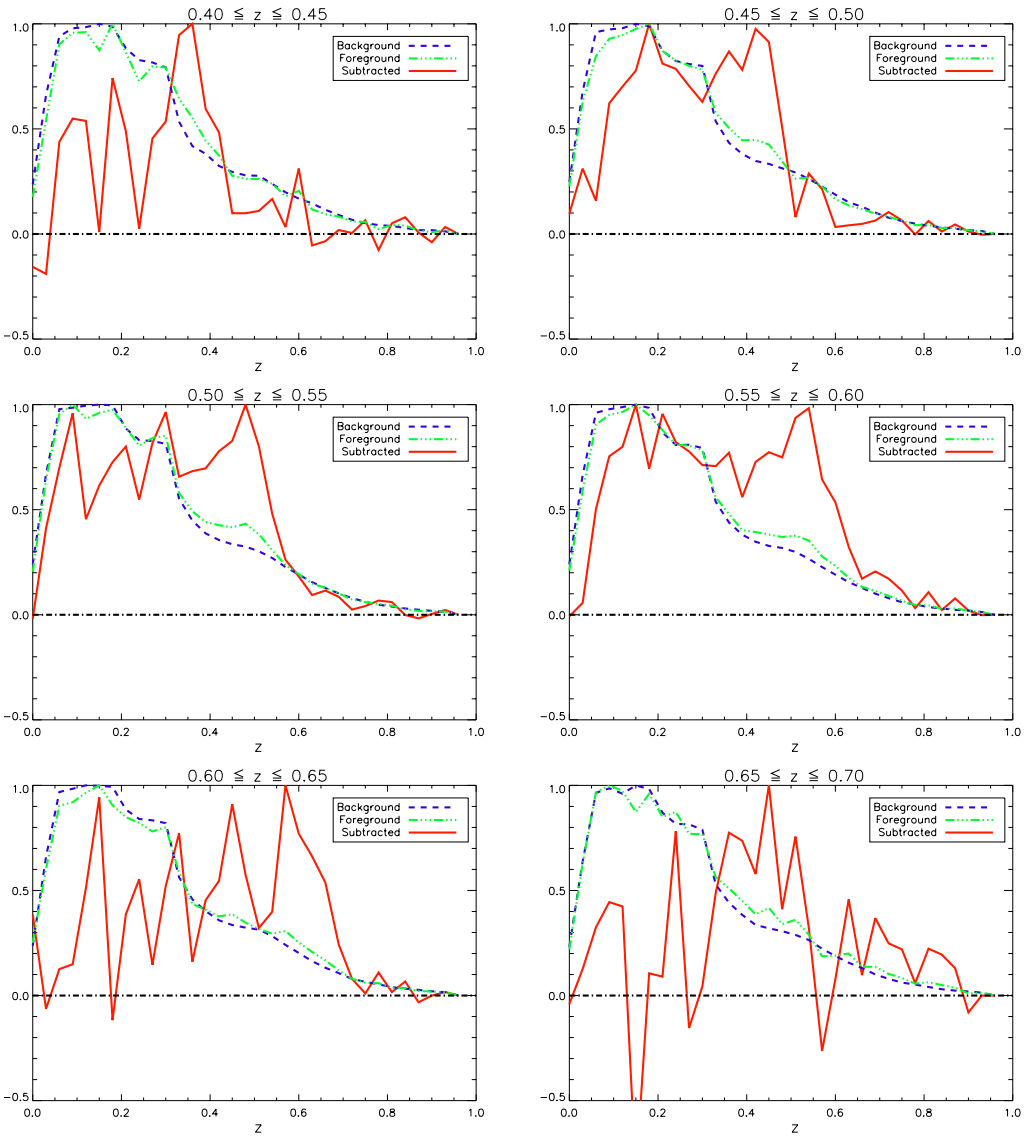


Figure 3.16: Number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters found in the 2SLAQ photometric catalogue. The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies.

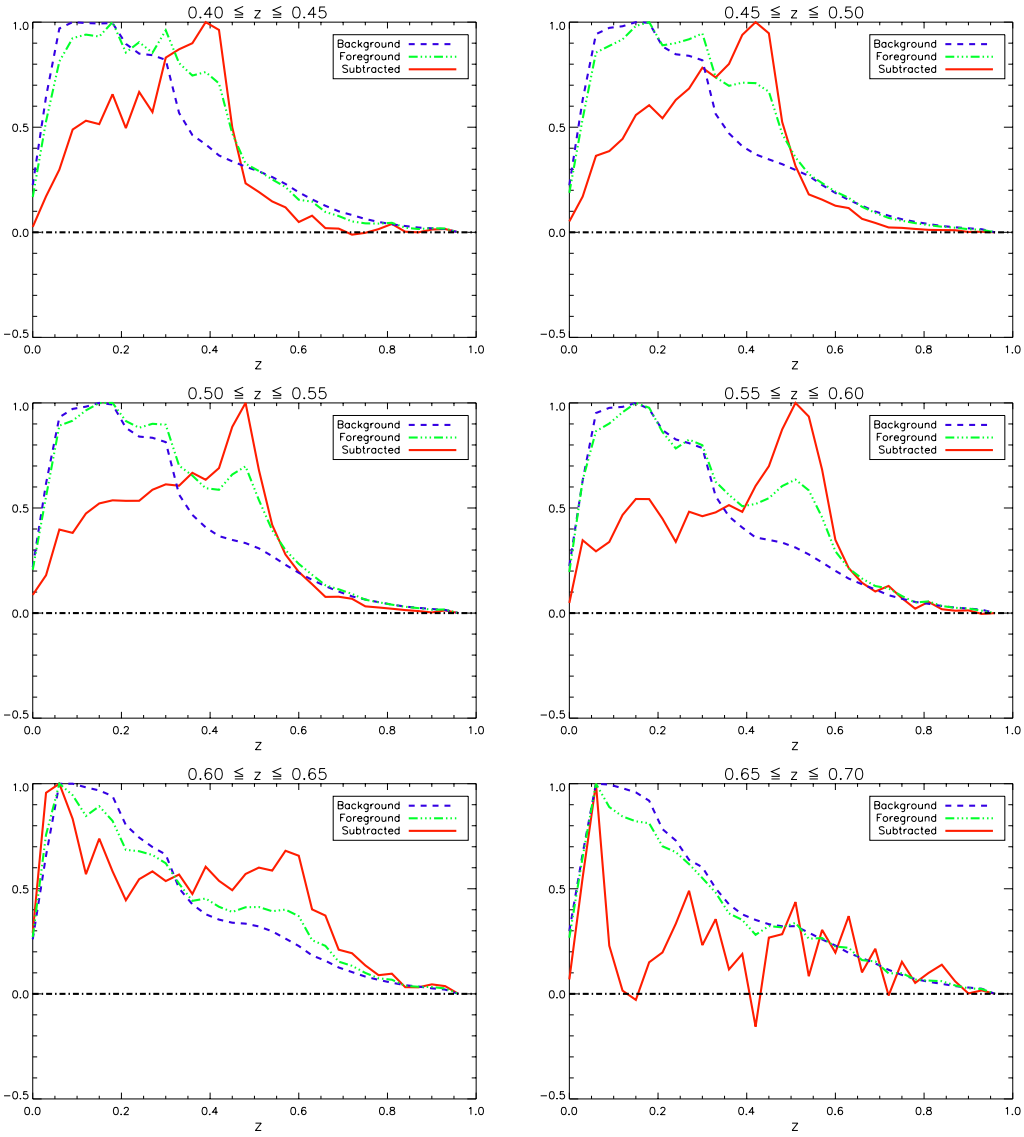


Figure 3.17: Number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters with $N_{\rm mem} \geq 6$ found in the Mega$z$-LRG catalogue (with linking parameters $R_{\rm friend}(z = 0.5) = 0.5\ h^{-1}$Mpc and $K = 0.8$). The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies.

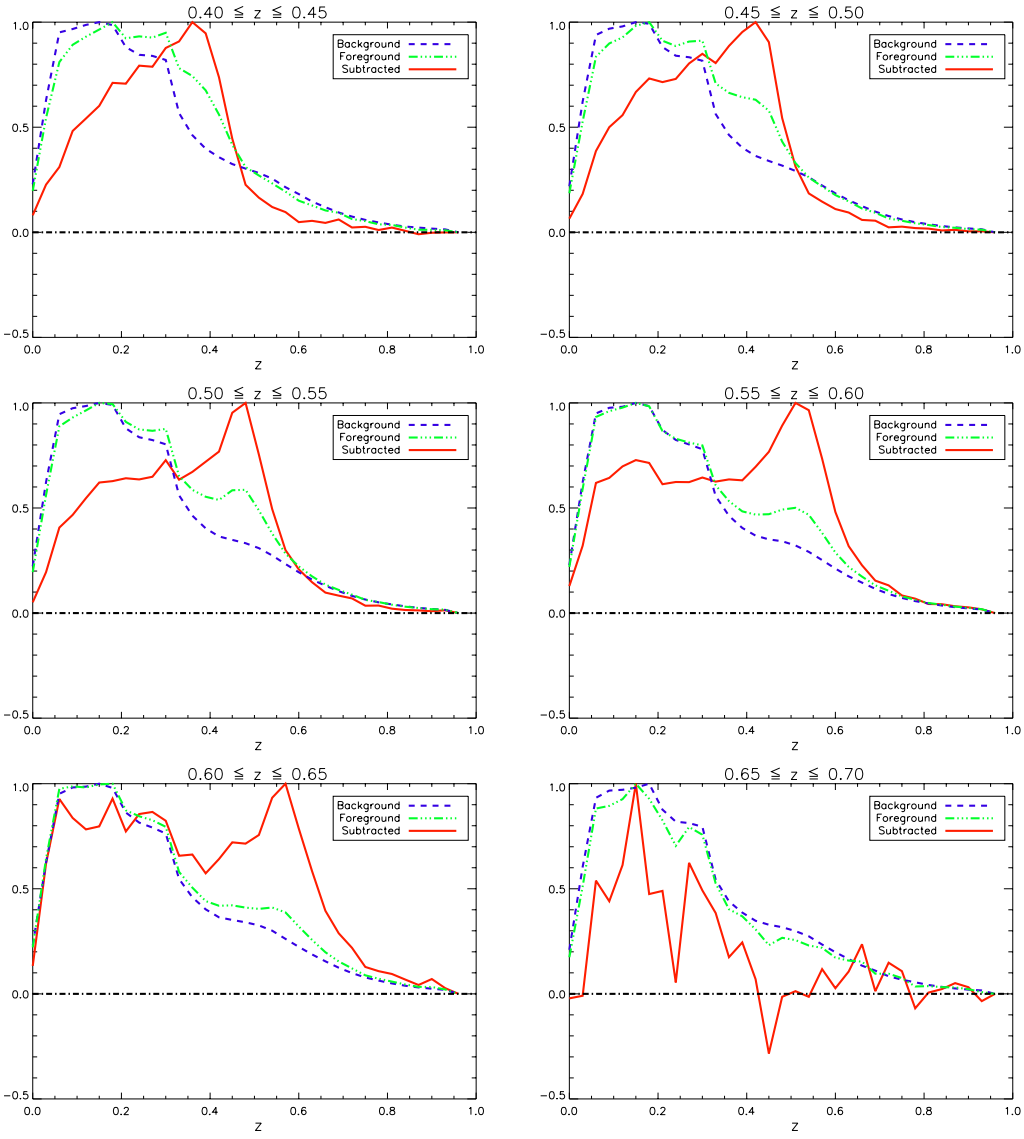


Figure 3.18: Number of SDSS background subtracted galaxies as a function of photometric redshift for all groups and clusters with $N_{\rm mem} \geq 10$ found in the Mega$z$-LRG catalogue (with linking parameters $R_{\rm friend}(z = 0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$). The blue dashed line are the background galaxies, the green dot-dashed line are the foreground galaxies and the red solid line are the background subtracted galaxies.

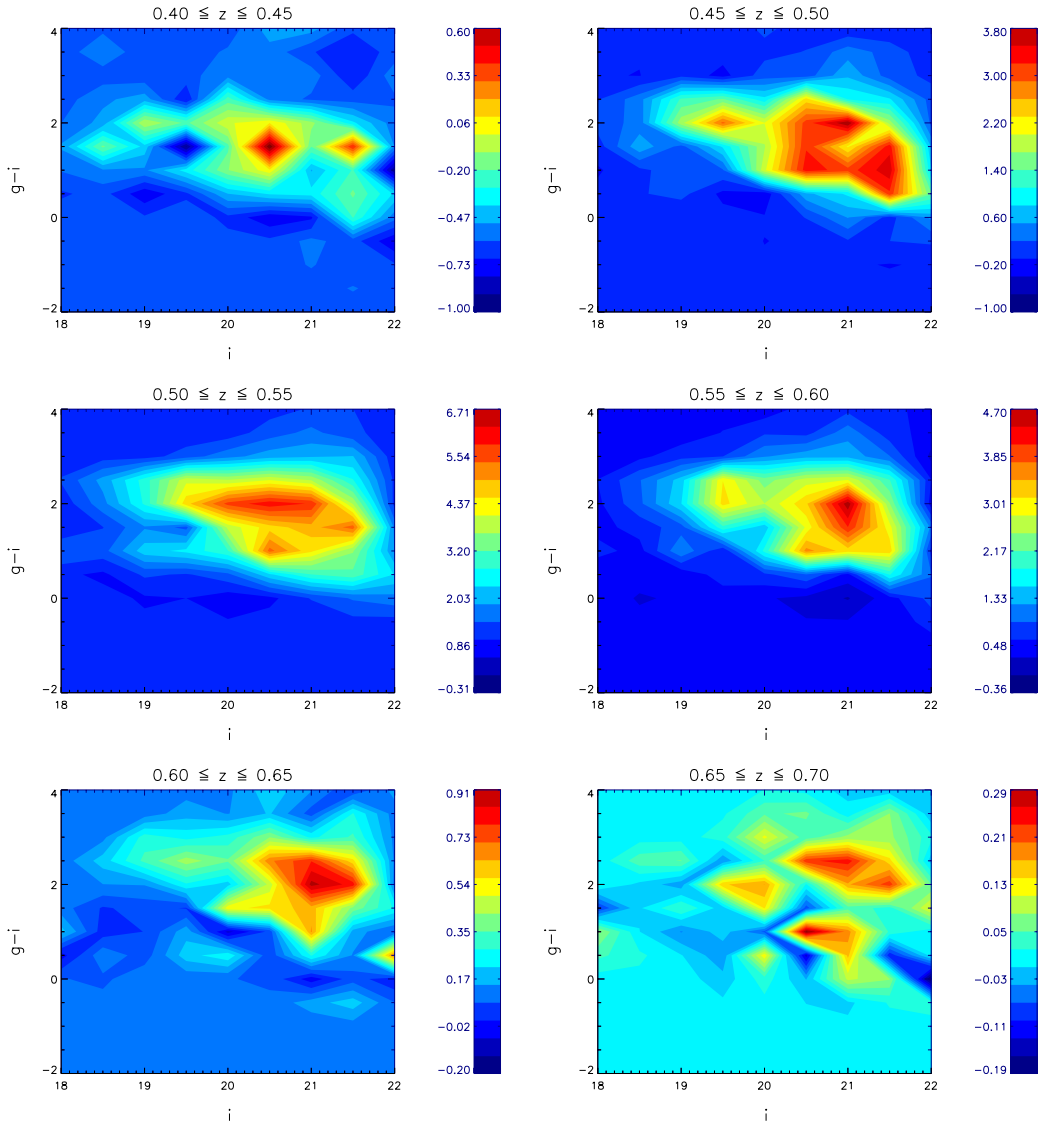


Figure 3.19: Colour-Magnitude diagrams for background subtracted SDSS galaxies for the 2SLAQ photometric cluster catalogue. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence.

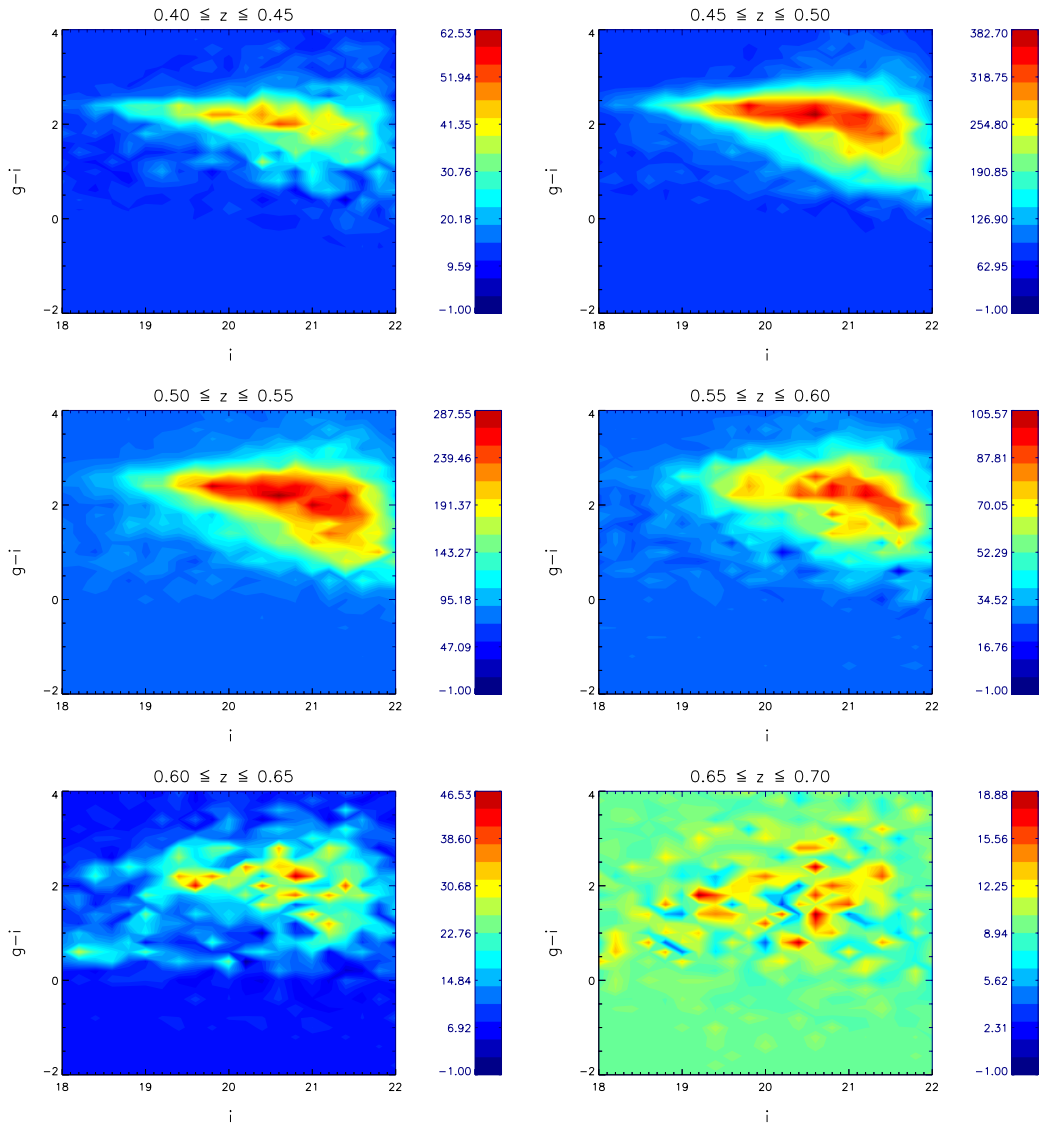


Figure 3.20: Colour-Magnitude diagrams for background subtracted SDSS galaxies for the Mega$z$-LRG cluster catalogue (with linking parameters $R_{\rm friend}(z = 0.5) = 0.5\ h^{-1}$Mpc and $K = 0.8$) with $N_{\rm mem} \geq 6$. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence.

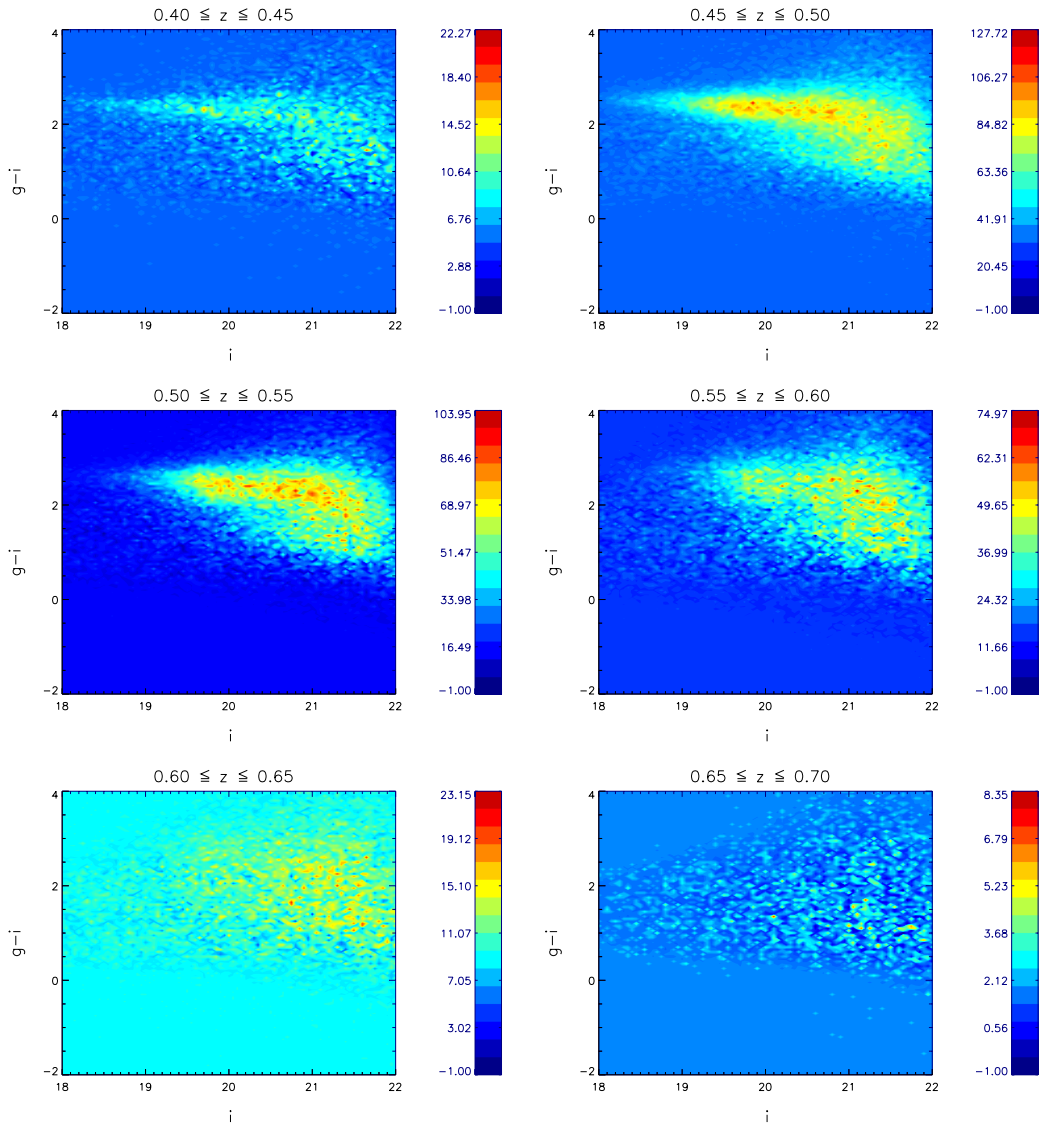


Figure 3.21: Colour-Magnitude diagrams for background subtracted SDSS galaxies for the Mega$z$-LRG cluster catalogue (with linking parameters $R_{\rm friend}(z = 0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$) with $N_{\rm mem} \geq 10$. The maps show a clear trend in colour-magnitude space that resembles a cluster red sequence.

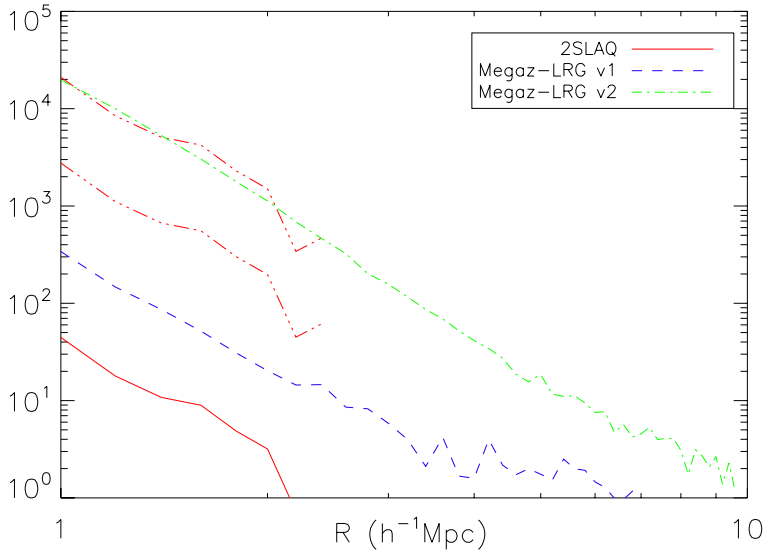


Figure 3.22: Radial profile of 2SLAQ spectroscopic cluster galaxies (red solid line), Mega$z$-LRG photometric cluster catalogue members, with linking parameters $R_{\rm friend}(z = 0.5) = 0.5\ h^{-1}$Mpc and $K = 0.8$ (blue dashed line) and Mega$z$-LRG photometric cluster catalogue members, with linking parameters $R_{\rm friend}(z = 0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$ (green dot-dashed line).

3.6 Correlation Function

The two-point correlation function for the 569 2SLAQ photometric groups and clusters obtained previously is calculated. Using the 2SLAQ photometric galaxy sample, a random catalogue is created that replicates the angular coverage on the sky, while assuming constant angular completeness with the observed region. This process is handled straightforwardly in the Healpix[1] package of software. The radial distribution of the random catalogue is obtained from a smooth spline fit to the galaxy redshift distribution, $n(z)$. The smoothing ensures the exclusion of large scale structure voids and filaments.

The correlation function is calculated with the Landy & Szalay estimator (Landy & Szalay 1993),

$$\xi = \frac{DD - 2DR + RR}{RR}, \tag{3.12}$$

where DD, DR and RR are the number of pairs of points in the data, D, and random, R, catalogues. The number of pairs are calculated in 14 bins, equally separated in log space, from $r = 5 - 90$ Mpc h^{-1}.

[1]This software is available from http://healpix.jpl.nasa.gov/

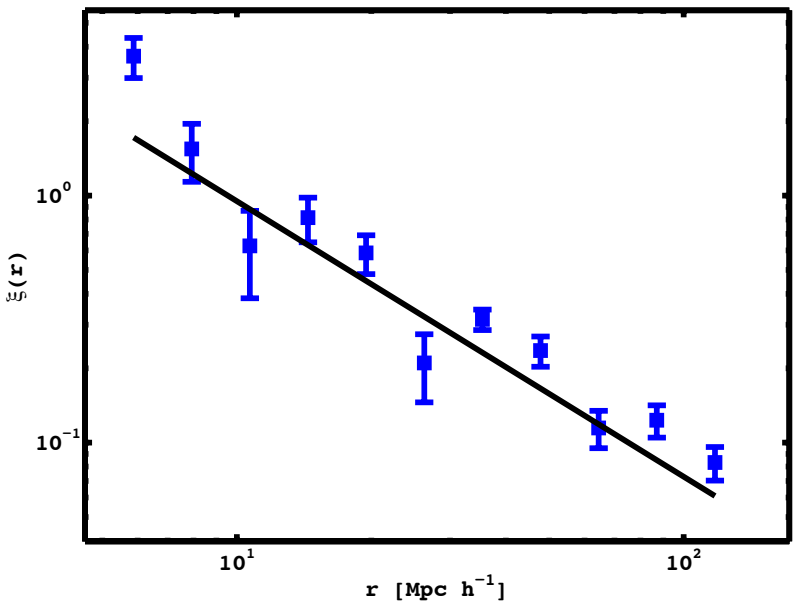


Figure 3.23: Two-point correlation function for the groups and clusters found in the 2SLAQ photometric catalogue using the DFoF code (blue squares with error bars) and best-fitting power law slope (black solid line). ξ(R) was measured using the estimator of Landy & Szalay (1993). Plot provided by Cristiano Sabiu.

In fig.3.23 the two-point correlation function of the 2SLAQ clusters (squares with error bars) and a best-fitting power-law slope (red solid line) are plotted. The error bars are estimated using the jackknife method, which involves dividing the survey into N sections with equal area or volume. The variance and mean are estimated from N measures of the statistic. Each measurement is performed on the survey with region i removed, where $i = 1, ..., N$.

The jackknife estimate of the variance is (Lupton 1993)

$$\sigma_{\xi}^2(r_i) = \frac{N_{jk} - 1}{N_{jk}} \sum_{j=1}^{N_{jk}} [\xi_j(r_i) - \overline{\xi}(r_i)]^2, \tag{3.13}$$

where N_{jk} is the number of Jackknife samples used and r_i represents a single bin in the statistic, ξ. In this analysis $N_{jk} = 20$ samples is set.

The power law model, $\xi(r) = (r/r_0)^{\gamma}$, is best fit with the parameters $r_0 = 9.55 \pm 4.23$ Mpc h^{-1} and $\gamma = 1.115 \pm 0.29$. The value of the reduced chi-squared is $\chi^2_{red} = 0.89$.

These results show less correlation than those found in §2.6, where the correlation function of the spectroscopic cluster catalogue was best fit with parameters $r_0 = 24 \pm 4$ Mpc h^{-1} and $\gamma = -2.1 \pm 0.2$ and the mock halo catalogue was best fit with parameters $r_0 = 23.05 \pm 0.72$ Mpc h^{-1} and $\gamma = -1.95 \pm 0.05$. This discrepancy is partially due to a higher number of false

detections or impurities in the photometric cluster catalogue, which dilutes the clustering length. Objects are detected at random locations where there is no genuine structure, which will cause the amount of correlation at a given scale to decrease. In addition to this, the differences between the correlation lengths of the spectroscopic and photometric cluster catalogues can also be attributed to the incompleteness of the photometric cluster catalogue as a function of the richness of the objects and hence as function of mass. Clusters with lower masses have lower clustering lengths and preferentially including them in the sample from up-scattered detections will also suppress the correlation length. This implies that the two-point correlation can actually be used to measure the amount of impurity and completeness in the photometric cluster catalogue with respect to the spectroscopic catalogue. This also indicates, however, that in its current form the photometric cluster catalogue cannot be used to make reliable constraints on cosmological parameters such as σ_8 and $\Omega_{\rm m}$ without further calibration.

3.7 Conclusions

An optical cluster finding algorithm based on that of Botzler *et al.* (2004) has been presented that uses a dynamical linking length that changes with the surface number density of galaxies at a given redshift range.

The cluster finder was run with a range of different linking parameters on the 2SLAQ catalogue with photometric redshifts obtained using ANNz (Collister & Lahav 2004). The completeness and purity of the resulting catalogues were tested by matching the galaxy members of the clusters found to the clusters found in the 2SLAQ spectroscopic cluster catalogue produced in chapter 2. Additionally, photometric redshifts are added to the spectroscopic mock catalogue, also produced in chapter 2, using the bias between the 2SLAQ spectroscopic and photometric redshifts. To add the photometric redshifts it was necessary to remove a certain number of objects to maintain the correct number density of galaxies. This was done in two different ways; first adding photometric redshifts and then removing objects and second by removing objects and then adding photometric redshifts. The results showed that the way in which objects were removed alters the level of completeness significantly. In addition, no correlation is found between the mock results and the 2SLAQ photometric cluster results. A more sophisticated mock would be required to constrain the linking parameters. This would entail adding realistic colours to the galaxies in the mock before producing the photometric redshifts. Since no combination of parameters exists, which optimises both the completeness and the purity simultaneously, linking parameters of $R_{\rm friend}(0.5) = 0.8\ h^{-1}$Mpc and

$K = 1.8$ were chosen purely from the 2SLAQ photometric cluster matching. This corresponds to a catalogue, which is 38% complete (80% when matched to clusters with $N_{\rm mem} \geq 6$) and includes 569 clusters containing 2,180 galaxies.

To constrain the linking parameters for the Mega$z$-LRG catalogue, the physical properties of the resulting structures from a given combination of parameters were analysed. Additionally, Mega$z$-LRG clusters found within the 145 deg^2 2SLAQ area were matched to the 2SLAQ spectroscopic clusters. Based on these results two cluster catalogues were chosen, one with parameters $R_{\rm friend}(0.5) = 0.5\ h^{-1}$Mpc and $K = 0.8$, which contains 21,255 groups and clusters (containing 73073 galaxies) that are more physical structures with low completeness, and another with parameters $R_{\rm friend}(0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$, which contains 114,874 groups and clusters (containing 544,008 galaxies) that are more unphysical structures with higher completeness. This occurs because as the linking parameters are increased, a large number of smaller structures are found contributing to the completeness and at the same time larger structures are merged together to form extremely large unphysical structures. Later analysis shows that a maximum of 6% of these structures are unphysical all of which appear to exist near the edges the Mega$z$-LRG redshift distribution.

The validity of the cluster catalogues was tested by obtaining SDSS galaxies in a 1 deg^2 region around each cluster centre. The background signal from all the SDSS galaxies within 1 h^{-1}Mpc of the cluster centre were subtracted and the results stacked in bins according to cluster redshift. Each of these bins was examined in SDSS photometric redshift space and colour-magnitude space. This analysis shows that the 2SLAQ photometric groups and clusters are reliable out to $z \sim 0.6$, which is consistent with the 2SLAQ spectroscopic results. The Mega$z$-LRG results show reliable cluster detection out to $z \sim 0.65$, however the results appear more pronounced for the Mega$z$-LRG cluster catalogue with $R_{\rm friend}(0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$, which shows a trend of ~ 0.3 in $g - i$ colour space. This is consistent with the earlier completeness analysis.

The radial profiles of all three cluster catalogues were examined and although all three profiles show roughly the same slope, the profile corresponding to the Mega$z$-LRG cluster catalogue with $R_{\rm friend}(0.5) = 0.75\ h^{-1}$Mpc and $K = 1.5$ appears to fit better with the 2SLAQ spectroscopic cluster profile. Another strong indication that this set of parameters is more appropriate for the Mega$z$-LRG catalogue.

Finally, the two-point correlation function of the photometric 2SLAQ cluster catalogue is tested. A best-fitting power law model, $\xi(r) = (r/r_0)^{\gamma}$, with parameters $r_0 = 9.55 \pm 4.23$ Mpc h^{-1} and $\gamma = -1.115 \pm 0.29$ is found. The value of the reduced chi-squared is $\chi^2_{\rm red} = 0.89$. These values

show less correlation than that found with the spectroscopic 2SLAQ clusters, which is the result of a higher number of impurities and incompleteness at low richnesses in the photometric 2SLAQ cluster catalogue.

The principal difficulty encountered when applying the (Botzler *et al.* 2004) technique is that is appears to depend on the number density of the galaxy catalogue. While the code reproduces the 2SLAQ spectroscopic cluster catalogue to reasonable degree, it becomes difficult to produce a complete cluster catalogue before significant merging occurs in a denser catalogue like Mega$z$-LRG. To produce a complete and physical cluster catalogue further analysis is required. This could be done by examining the substructure of the over-merged structures or by trying to remove interloping field galaxies, which allow objects to merge, before running the DFoF code. Hence for catalogues with a surface number density of galaxies such as Mega$z$-LRG, this method fails to produce a catalogue that is both pure and complete. It is, however, possible to independently produce a catalogue that is either pure or complete.

CHAPTER 4

AAT SPECTROSCOPIC OBSERVATIONS OF A SELECTION OF PHOTOMETRIC FOF TARGETS

4.1 Introduction

In the current cosmological model of the Universe it is believed that most of the mass resides in a network of cold dark matter haloes. Small perturbation in the early Universe grow through gravitational collapse and mergers, hierarchically forming structure over time. This model has successfully been used to explain much of the distribution and structure of the objects we observe in the Universe now. Our lack of knowledge regarding the source and nature of dark matter, however, limits our ability to fully understand the origin and fate of structure in the Universe. It is therefore essential to test and refine this model using all available data.

One of the most useful tools in probing the evolution and properties of CDM haloes are cosmological N-body simulations, in which particles within a given volume are allowed to evolve under gravity. Specifically N-body simulations provide estimates for the number density of CDM haloes as a function of mass, their spatial distribution and clustering, their radial density profiles and the their growth and evolution with redshift. All of these predictions can easily be compared to observations to test the reliability of the current CDM model.

Clusters of galaxies are ideal objects to compare with the predictions of N-body simulations as they directly trace the distribution of dark matter in the Universe. More fundamentally, the mass function of galaxy clusters is a particularly sensitive probe of cosmological models. The clusters should be sampled over a large volume to produce the most accurate tests of cosmology. The clusters detected in the 8000 deg^2 Mega$z$-LRG catalogue as described in §3 fit this criterion,

however before any comparison can be made it is necessary to calibrate the relation between the dark matter halo mass and the observed cluster richness. To accomplish this, a proposal was made to carry out a spectroscopic follow-up of a sample of the Mega$z$-LRG clusters using the Anglo-Australian Telescope. The spectra could then be used to determine the redshifts of the galaxies, which in turn provide the cluster velocity dispersions and from these estimates can be made of the cluster dynamical masses.

The project, described in this chapter, aims to follow-up a subset of around 100 galaxy clusters from the Mega$z$-LRG DR5 cluster catalogue. The cluster sizes and velocity dispersions can then be measured to estimate the halo dynamical masses, which in turn can be used to calibrate the scatter between mass and richness (van der Marel *et al.* 2000; Yee & Ellingson 2003). In addition to this, the galaxy redshifts can be used to establish the reliability of the cluster finding algorithm by identifying missing or interloping cluster members.

The following sections provide details on the observations made at the Anglo-Australian Observatory (AAO) and the analysis of this data. Section 4.2 briefly describes the telescope and instruments used at the AAO. Section 4.3 highlights the main points in the observing proposal submitted to the AAO. Section 4.4 presents the observations obtained and their analysis. Finally, section 4.5 discusses the conclusions and plans for future work regarding the AAO data.

4.2 The Anglo-Australian Telescope

The Anglo-Australian Telescope (AAT) is a 3.9 m equatorially mounted reflecting telescope located at the Anglo-Australian Observatory (AAO). The AAO is situated at an altitude of 1100 m on the Siding Spring mountain near Coonabarabran in New South Wales, Australia. The median seeing recorded at this site is 1.5 arcseconds. The AAT was commissioned in 1974 and is equipped with the Two Degree Field (2dF) facility and the AAOmega spectrograph, which is described in further detail below (§4.2.1).

The AAT has been an important instrument for several important completed and ongoing projects such as the 2dF Galaxy Survey (Colless 1999), WiggleZ (Glazebrook *et al.* 2007) and GAMA (Driver *et al.* 2008).

4.2.1 AAOmega Spectrograph

AAOmega is a bench mounted, fibre fed, double-beam spectrograph that can be used for either multi-object spectroscopy or integral field spectroscopy. Table 4.1 highlights some of the main

Table 4.1: AAOmega Observing Mode Properties

Property	Multi-object	Integral Field
Field of view	2 deg	11x22 arcsec
Number of fibres	392	512
Angular size of fibre	2”	0.7”
Wavelength range	0.37-0.95μm	0.37-0.95μm
Dispersion	0.2-1.5 Å/pixel	0.2-1.5 Å/pixel
Spectral resolution (FWHM)	3.5 pixels	2.1 pixels
Spectral resolution ($\lambda/\Delta\lambda$)	1300-10000	2000-15000
Magnitude limit	B=22	B=22

properties of these two observing modes. AAOmega provides improved resolution, throughput and stability compared to the 2dF, RGO and SPIRAL spectrographs that were previously used with the AAT.

AAOmega uses a duel-beam system with separate blue and red arms split by a dichroic at 570nm as shown in fig.4.1[1]. The spectrograph uses Volume Phase Holographic (VPH) gratings, which can be up to 50% more efficient than standard reflection gratings, and the point spread function (PSF) of the instrument is extremely Gaussian and uniform, which is essential for accurate sky subtraction.

4.3 Observing Proposal

With the aim of expanding upon the work discussed in §3, an observing proposal was submitted to the AAO in spring 2007 for four nights to carry out a spectroscopic followup on a selection of Mega$z$-LRG cluster candidates. Some of the key points made in that proposal are discussed in the following subsections including a description of the input data (§4.3.1) and the observing strategy (§4.3.2).

4.3.1 Input Data

The Mega$z$-LRG catalogue, which is discussed in greater detail in §3, is a collection of over a million luminous red galaxies (LRGs) with photometric redshifts in the range $0.4 \leq z \leq 0.7$ and

[1] http://www.aao.gov.au/local/www/aaomega/

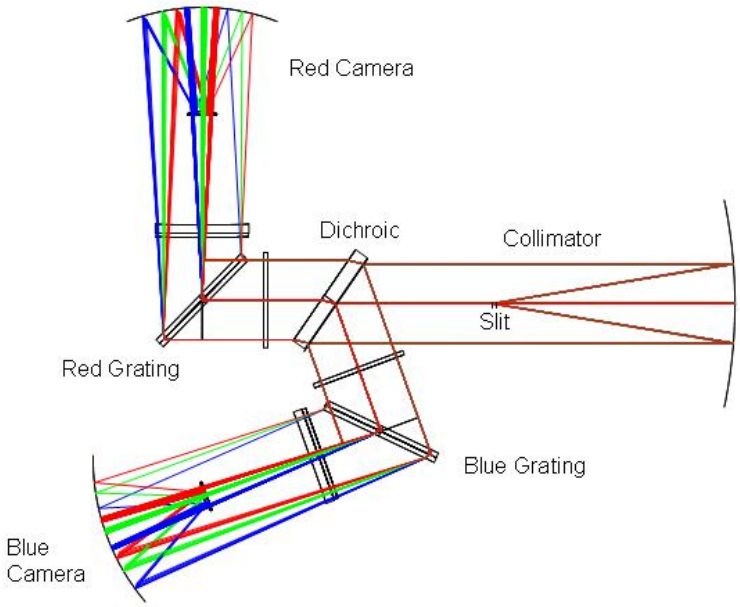


Figure 4.1: AAOmega optical layout in high dispersion mode.

magnitudes with $i < 19.8$ in a volume of $3.3h^{-3}\text{Gpc}^3$ produced from the imaging component of the Sloan Digital Sky Survey (SDSS) (Cannon *et al.* 2006). The fifth data release (DR5) of the Mega$z$-LRG catalogue contains a total of 1,468,028 galaxies. LRGs are important for cosmology as they reside in the densest environments and thus are good tracers of cosmic structure. The large volume of the Mega$z$-LRG catalogue enables one to probe the most massive peaks in the in the density field, which provides the most sensitive tests of the cosmological model. This catalogue has previously been used to test cosmological models by Blake *et al.* (2007, 2008); Thomas *et al.* (2010).

A cluster catalogue was produced from the southern strips of the Mega$z$-LRG DR5 catalogue using a friends-of-friends (FoF) procedure that uses photometric redshifts, which is based on that of Botzler *et al.* (2004) (see §3). The southern strips were selected by taking only galaxies in the ranges $-60 \leq$RA≤ 70 and $-15 \leq$DEC≤ 20 as shown in fig.4.2, which corresponds to a total of 138,308 objects. It should be noted, however, that the FoF code used to produce these cluster candidates was an older version than that described in §3 and is therefore, to a certain degree, less accurate. This is mainly because the older version of the code was run before mock calibrations were used and therefore the linking parameters were selected based on the physical properties of the resulting FoF clusters. The linking parameters were chosen such that the average cluster sizes were $R_{\text{clt}} \leq 2\ h^{-1}$Mpc and the average cluster velocity dispersions were $\sigma_v \leq 1000$ kms^{-1}. In

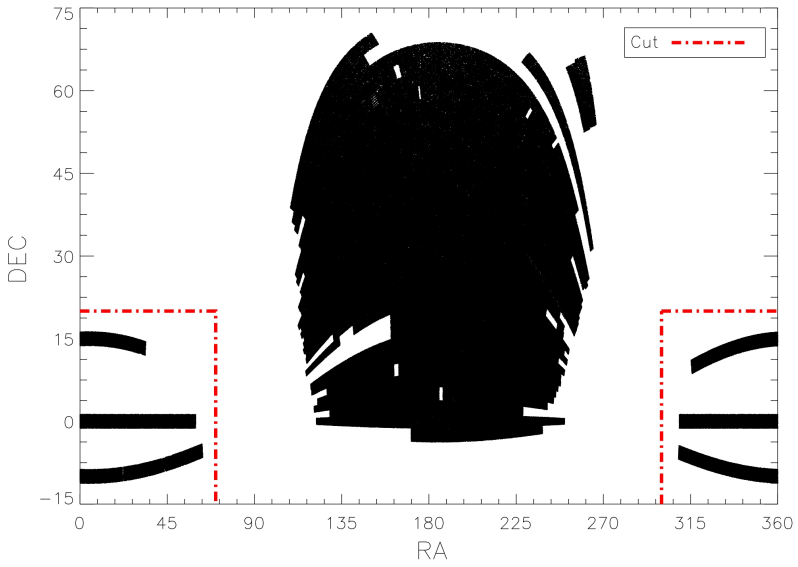


Figure 4.2: Distribution of Mega$z$-LRG DR5 galaxies. The red dot-dashed line shows cuts made to the full catalogue to leave only the southern strips.

Table 4.2: Richness Cuts

Sample	Minimum Richness	Number of Clusters
A	7	248
B	5	947
C	4	2894

the current version of the DFoF code this catalogue would have linking parameters smaller than those used in chapter 3 and would therefore be pure but not very complete. Given that a mock was not used for calibration, it is difficult gauge the exact level of completeness or purity of the cluster catalogue, however for the purposes of a spectroscopic follow up, the purity is the principal quantity of interest. Therefore, the cluster catalogue was divided into three samples based on cluster richness as shown in table 4.2. Sample A are the richest structures and thus should be the most reliable targets for observation. Fig.4.3 shows the positions of all of the clusters found in the southern strips of the Mega$z$-LRG DR5 catalogue, where the red circles correspond to sample A clusters, the blue plus signs correspond to sample B clusters and the yellow ‘x’s correspond to sample C clusters.

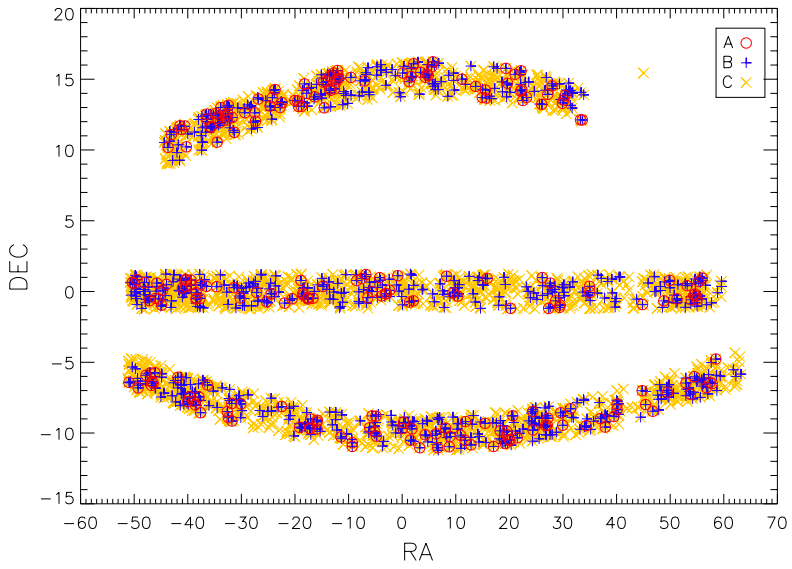


Figure 4.3: Positions of clusters in the souther strips of the Mega$z$-LRG DR5 catalogue. Red circles correspond to sample A clusters, blue plus signs correspond to sample B clusters and yellow 'x's correspond to sample C clusters.

4.3.2 Observing Strategy

In order to increase the efficiency of the observations the densest regions were targeted. Eight fields were selected in which there were multiple sample A clusters within a 2 deg diameter. Fig.4.4 shows the positions of the cluster candidates in the eight observing fields. The large red circles correspond to sample A clusters, the medium blue circles correspond to sample B clusters, the small yellow circles correspond to sample C clusters and the green 'x's correspond to 2SLAQ clusters. The dashed line shows the 2 deg field of view of the spectrograph. Regions also containing clusters from the 2SLAQ catalogue were favored as spectra for these galaxies are already available.

As is it difficult to obtain accurate velocity dispersions with only a few LRGs, each field was supplemented with SDSS galaxies ($i < 20$) lying along the colour-magnitude red sequence of the clusters. This approach has been used by previous cluster surveys such as CNOC Carlberg *et al.* (1996) and RCS Blindert *et al.* (2004) to estimate cluster velocity dispersions. SDSS galaxies were selected for each Mega$z$ cluster based on two ranking criteria. The first criterion was that if a galaxy satisfied

$$|(r-i)_{\rm gal} - (r-i)_{\rm RS}| < 0.1 \tag{4.1}$$

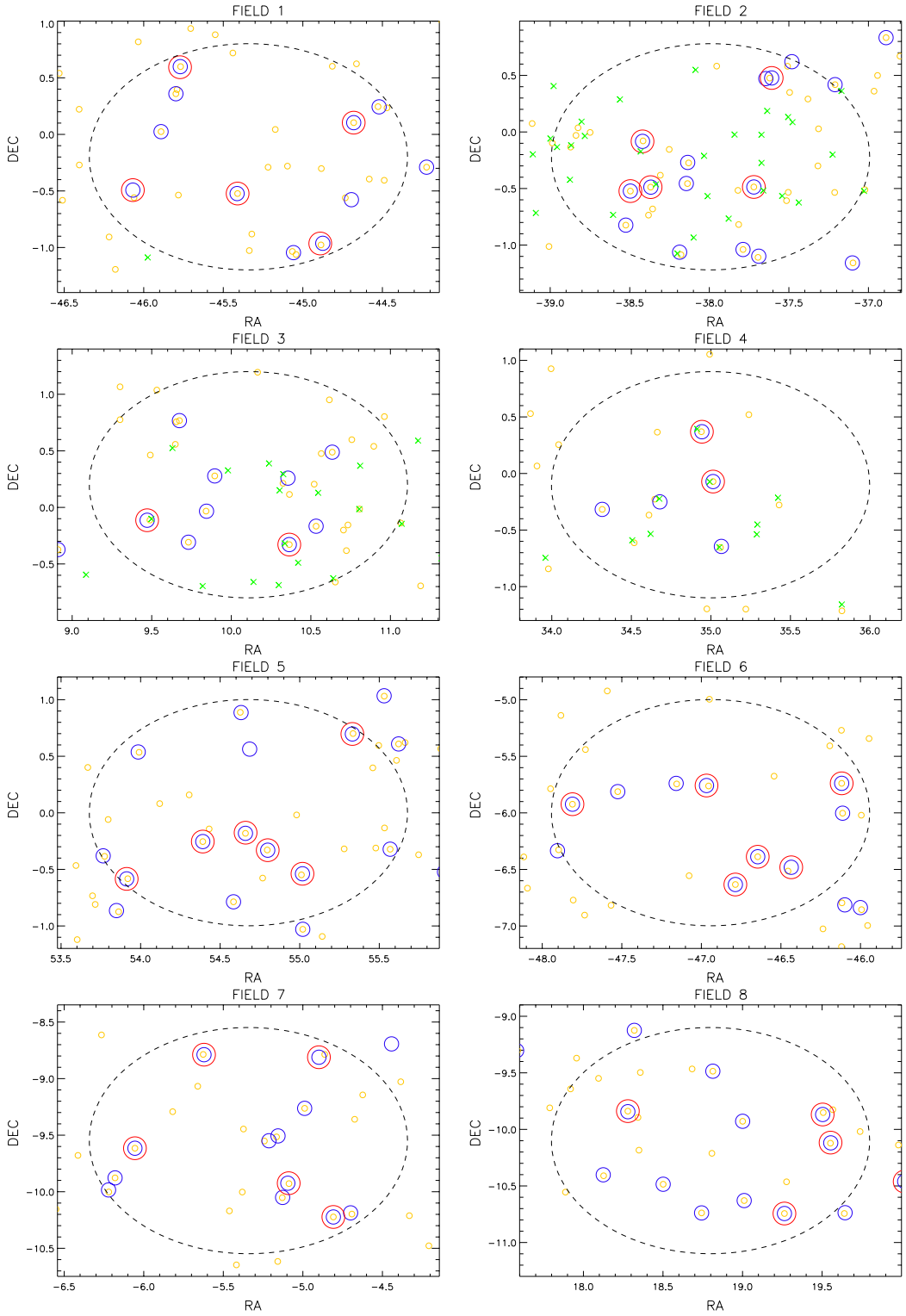


Figure 4.4: Positions of clusters in target fields. The large red circles correspond to sample A clusters, the medium blue circles correspond to sample B clusters, the small yellow circles correspond to sample C clusters and the green 'x's correspond to 2SLAQ clusters. The dashed line shows the 2 deg field of view of the spectrograph.

Table 4.3: AAOmega Diffraction Gratings

Grating	Wavelength Range	Dispersion	MOS Resolution
580V	3700-5800 Å	0.1 nm/pixel	1300 R
385R	5600-8800 Å	0.16 nm/pixel	1300 R

and

$$i_{\text{bright}} \leq i_{gal} \leq 19.8, \tag{4.2}$$

where $(r-i)_{\text{gal}}$ and i_{gal} are the colour and magnitude of the galaxy, $(r-i)_{\text{RS}}$ is the colour of the red sequence track and i_{bright} is the brightest galaxy in the cluster, then a rank of 1, 3 or 5 was assigned to the galaxy if it was located within 1, 2 or 3 h^{-1}Mpc of the cluster centre respectively. $(r-i)_{\text{RS}}$ was calculated for each i-mag and redshift using the model of Bruzual & Charlot (2003). The second criterion was that if a galaxy satisfied

$$((r-i)_{\text{RS}} - 0.5) \leq (r-i)_{\text{gal}} \leq ((r-i)_{\text{RS}} - 0.1), \tag{4.3}$$

and

$$i_{\text{blue}} \leq i_{gal} \leq 19.8, \tag{4.4}$$

where $i_{\text{blue}} = (i_{\text{bright}} + 19.8)/2$ or $i_{\text{blue}} = i_{\text{bright}} + 0.5$ depending upon which of the two values is smaller, then a rank of 2, 4 or 6 was assigned to the galaxy if it was located within 1, 2 or 3 h^{-1}Mpc of the cluster centre respectively. The SDSS galaxies were then selected in rank order until each field had at maximum 392 targets, which is the number of AAOmega spectrograph fibres.

The main technical challenge for these particular observations using AAOmega is that the spectrograph fibres cannot be placed more closely than 30 arcseconds. This limitation can cause problems with fibre collisions when attempting to observe galaxies that are separated by $\sim$ 1-2 arcminutes. On average the virial radius of clusters is $\sim 1h^{-1}$ Mpc, which projects to an angular diameter of $\sim$ 2.5 arcminutes on the sky at $z \approx 0.55$. Therefore, two separate fibre configurations were assumed for each field, which ensures that in each field spectra are obtained for each cluster galaxy.

The 580V and 385R diffraction gratings were chosen for the blue and red arms of the AAOmega spectrograph respectively. The basic properties of these gratings are shown in table 4.3, which principally shows that the wavelength coverage is 3700 Å$\leq \lambda \leq$ 8800 Å.

Table 4.4: Fractional Lunar Illumination

Night	FLI Range
Dark	$0.00 \leq \text{FLI} < 0.25$
Grey	$0.25 \leq \text{FLI} < 0.65$
Bright	$0.65 \leq \text{FLI} \leq 1.00$

A signal-to-noise (SNR) ratio of 3 per Å has been shown to produce high redshift completeness for red galaxies from AAT data (Cannon *et al.* 2006). The AAOmega Exposure Time Calculator shows that a SNR $\approx$ 3.6 per Å can be obtained for an $I = 20$ galaxy in grey time and 1.5 arcsecond seeing with a one hour exposure. Therefore, in order to observe each field twice (for both configurations) for one hour, a total of four grey nights were required to carry out the project.

A grey night is defined by the fractional lunar illumination (FLI) as shown in table 4.4, where FLI denotes the fraction of time in which the moon is below the horizon as given by

$$2 \times \text{FLI} = 1 - \cos\beta\cos(\lambda - \lambda_{\odot}), \tag{4.5}$$

where λ and $\lambda_{\odot}$ are the longitudes of the Moon and the Sun respectively.

4.4 Analysis & Results

Observations were carried out at the AAO between the 7^{th} and 10^{th} of September 2007. During this period two and half nights were clear and a total of 4116 spectra were obtained from the 8 fields. The following subsections describe the process followed for reducing the data (§4.4.1), obtaining the galaxy redshifts (§4.4.2) and the results found (§4.4.3).

4.4.1 Data Reduction

The observations obtained with the AAT were reduced using the standard AAO pipeline software called 2dfdr [2]. 2dfdr is a package that was originally developed by Jeremy Bailey to reduce 2df data. The package has more recently been updated for reducing data from other instruments such as 6df, SPIRAL, SDSS and AAOmega.

The 2dfdr system performs bias and dark subtraction, flat fielding, tram-line mapping to fibre locations on the CCD, fibre extraction, arc identification, wavelength calibration, fibre throughput

[2] http://www.aao.gov.au/AAO/2df/aaomega/aaomega_software.html

calibration and sky subtraction. Each of these data reduction procedures are described in the following subsections[3].

4.4.1.1 Debiassing

Debiassing or bias subtraction is performed by subtracting a bias frame from the data image. A bias frame is an image obtained with the instrument with no actual exposure time. This removes noise due to electronics in the device from the data image. As there is no exposure time the bias frame only contains the intrinsic bias of the sensor without a build up of charge.

4.4.1.2 Dark Subtraction

Dark subtraction is performed by subtracting a dark frame from the data image. A dark frame is an exposure obtained with the instrument in darkness. This removes noise due to the small electric current that flows though a CCD when it is not collecting photons, which is commonly referred to as dark current. Most CCDs contain a certain number of pixels with imperfections that cause them to collect more of these thermal electrons than the rest or none at all. Dark subtraction removes the noise caused by these hot and/or dead pixels.

4.4.1.3 Flat Fielding

Flat fielding is performed by dividing the data image by a flat field frame. A flat field frame is obtained by taking an exposure of a uniform radiation source with the instrument. In practice, flat field frames are generally obtained by taking an exposure of a blank screen on the inside of the telescope dome with the lights turned on or by taking an exposure of the sky during twilight hours. The disadvantage with the first approach is that in can be difficult to illuminate the dome in a way that is uniform. The sky at twilight is a much better approximation of uniform ration, however the problem with this approach is that the brightness of the sky changes rapidly during twilight hours, so it can be difficult to obtain multiple flat fields. Fig.4.5 shows an example of a flat field obtained inside a telescope dome. Flat fielding removes the noise in an image caused by the pixel-to-pixel variation in the CCD. This variation is owing to the fact that every pixel in a CCD has a different gain or quantum efficiency, η, which is given by

$$\eta = \frac{N_e}{N_\nu}, \tag{4.6}$$

where N_e is the number of electrons produced and N_ν is the number of photons absorbed.

[3] see http://www.starlink.ac.uk/star/docs/sc14.htx/sc14.html for more details

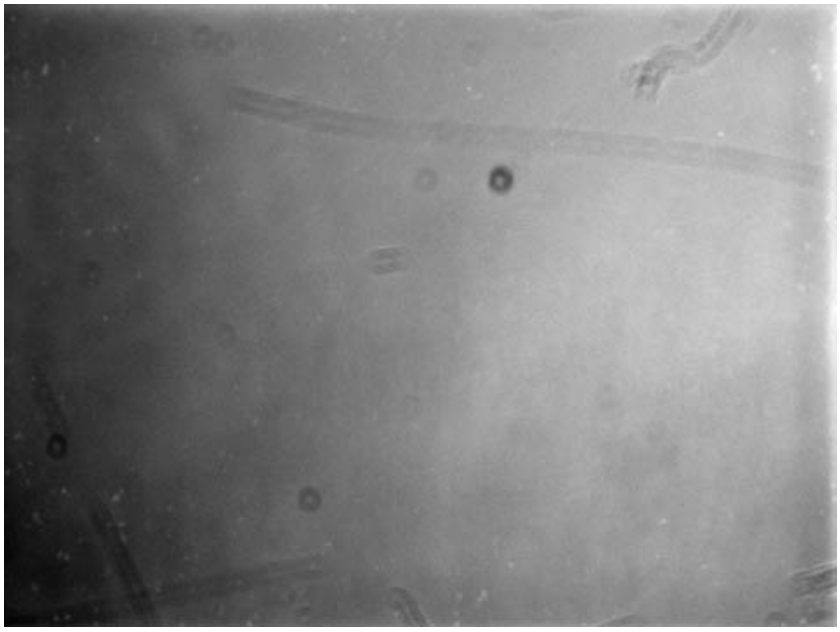


Figure 4.5: Example of a flat field frame obtained inside a telescope dome.

4.4.1.4 Tram-Line Mapping

Tram-line mapping or tramlining is the process of tracking the positions of the fibre spectra on the CCD. Fibre spectrographs produce frames containing a series of horizontal lines, which are the individual spectra obtained through each of the fibres as shown in fig.4.6. In order to extract the individual spectra, it is necessary to define the aperture of each spectrum on the CCD. This is done by tracing each spectrum to identify the amount of curvature caused by the optics of the system and the degree of misalignment with the CCD grid. In addition, it is necessary to determine the width of the spectrum in the number of pixels it spans. The tram-line map can be made using a flat field frame to ensure that the spectra are well defined.

4.4.1.5 Fibre Extraction

Fibre extraction is the process of extracting the one-dimensional spectra from the CCD frame. This involves first subtracting background scattered light from the data, then fitting overlapping Gaussian profiles to a fibre flat field frame, fitting these profiles to the data image and finally using a tram-line map to identify the positions of the spectra and extract them.

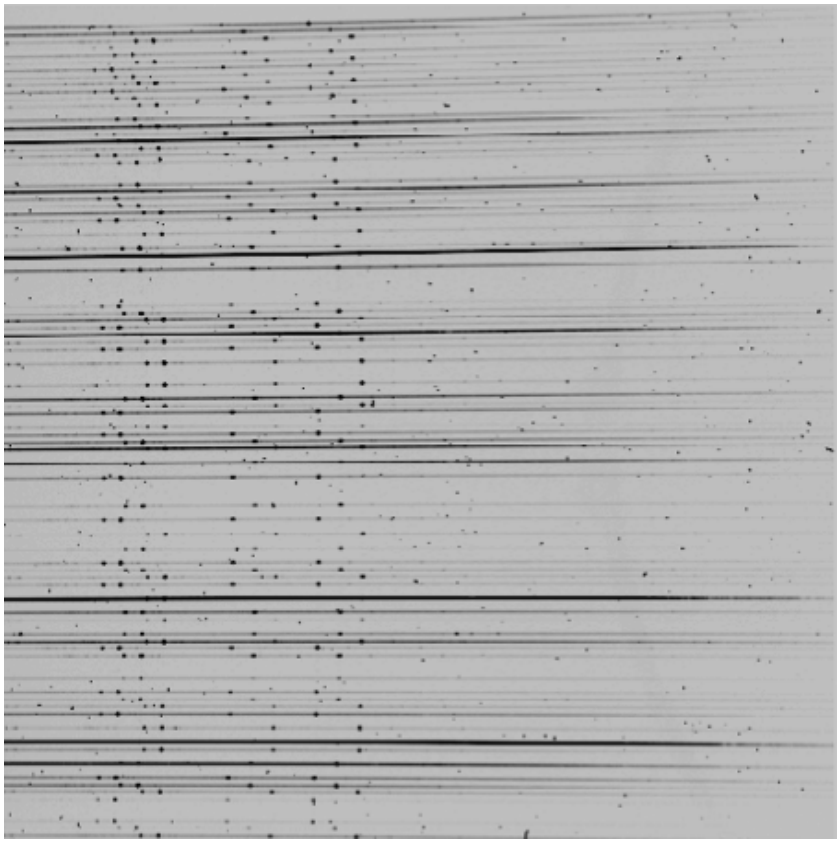


Figure 4.6: Example of a tram-line map observed with the WYF- FOS/AUTOFIB2 fibre spectrograph on the WHT.

4.4.1.6 Arc Identification

Arc identification is the process of identifying emission or absorption line features in the spectrum of a calibration source or arc lamp. Arc lamps consist of two electrodes separated by a gas, which for calibration purposes can typically be *e.g.* CuAr, CuNe, He, HeAr, FeAr, FeNe or ThAr. This identification procedure is generally done in a semi-interactive way using data reduction software such as FIGARO[4], IRAF[5] and MIDAS[6], however some automated methods also exist (Zuiderwijk 1995).

4.4.1.7 Wavelength Calibration

Wavelength calibration is the process of converting the CCD pixel positions of spectra to wavelengths using calibration frames. Calibration frames are obtained by taking exposures of an arc

[4]http://www.aao.gov.au/figaro/

[5]http://iraf.noao.edu/

[6]http://www.eso.org/sci/software/esomidas/

lamp, which has a spectrum with emission lines of know wavelength. A relation can then be made between the CCD pixel positions and wavelengths of the arc lamp emission lines by fitting a low order polynomial. Finally, this relation can be applied to the data image spectra to identify the wavelengths of the spectral features.

4.4.1.8 Fibre Throughput Calibration

Fibre throughput calibration is performed by dividing an individual spectrum by its corresponding normalised fibre throughput. The normalised fibre throughput for a given fibre is obtained by taking the mean signal in the fibre divided by the mean signal over all fibres. Generally the relative fibre throughput is obtained by taking exposures offset from the target field, so that the fibres are configured in the same way but are receiving sky signal. Another method uses sky emission lines to normalise the fibre throughputs. This process corrects for the fibre-to-fibre differences in the spectrograph.

4.4.1.9 Sky Subtraction

Sky subtraction is performed by subtracting a sky frame from an observed spectrum. A sky frame is obtained by taking an exposure of a region of blank sky. This can be done simultaneously to target exposures by allocating a certain number of spectrograph fibres to blank patches of sky, or separately before and/or after the target exposures. The amount of sky correction required depends on the angular size of the field of view of the fibre. Sky subtraction removes noise caused by emission from the night sky, which is principally caused by the aurora, zodiacal light, which is sunlight reflected off interplanetary dust, atmospheric emission, which is mainly from water lines as seen in fig.4.7, and faint background sources.

4.4.2 Redshifting

Galaxy spectra are typically characterised by a strong continuum component with a break at 4000 Å, which is caused by the absorption of high energy radiation by metals in the atmospheres in stars and the absence of hot blue stars. Other key features include the calcium H and K lines located at 3934 Å and 3969 Å respectively. These features are shifted to the range $5600 \leq \lambda \leq 6800$ Å for the redshift range $0.4 \leq z \leq 0.7$ and therefore are often used to identify galaxy redshifts.

The reduced galaxy spectra obtained with the AAT were redshifted using RUNZ. RUNZ is a Fortran 77 redshift-fitting software package written by Will Sutherland for determining redshifts in

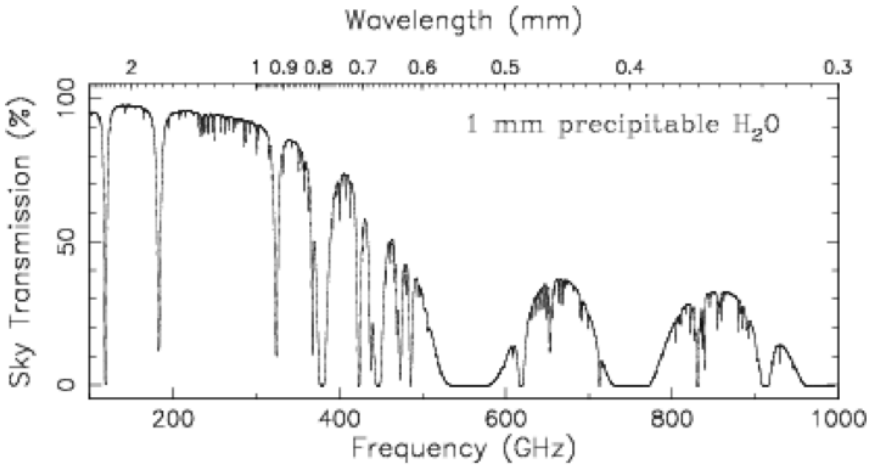


Figure 4.7: Example spectrum of the Earth's atmosphere containing water lines.

the 2df Galaxy Redshift Survey. The package was later modified for use on 6df (by Will Saunders), on 2SLAQ (by Russell Cannon) and on Wigglez (by Scott Croom). Fig.4.8 shows a screen shot of the RUNZ environment with an example galaxy spectrum. Redshifts are chosen for each spectrum by identifying absorption features (*e.g.* calcium H and K lines) or emission features (*e.g.* oxygen OII line) and fitting a template. Fig.4.9 shows a close up of an example galaxy spectrum (top panel) and a template fitted to the galaxy spectrum (bottom panel). In this example the H, K and G lines are clearly identified. Fig.4.10 shows an example of stellar spectrum (top panel) and galaxy spectrum with a visible OII emission line (bottom panel). M class stars, which are very cool stars with surfaces temperatures $T \leq 3900$ K, are often mistaken for galaxies because on the sky they appear as bright red objects. The spectra of M stars, like the one shown in the top panel of fig.4.10, are quite distinct from galaxies and contain large absorption features due to molecules like TiO.

In the final step of the RUNZ code each galaxy redshift is assigned a quality, as shown in table 4.5, to record its reliability. Qualities of 2 or below are assigned to galaxies for which no redshift could be obtained or for which the redshift obtained was not reliable enough to be used for further analysis. A quality of 6 is assigned to a spectrum that corresponds to a M class star. Qualities of 3-5 are assigned to galaxies for which a reliable redshift can be obtained and only these objects are used for further analysis.

4.4.3 Results

Each of the eight observing fields was redshifted at least once. Fields 1, 2, 3, 5, 6 and 7 were independently redshifted multiple times by different people in order to increase the reliability

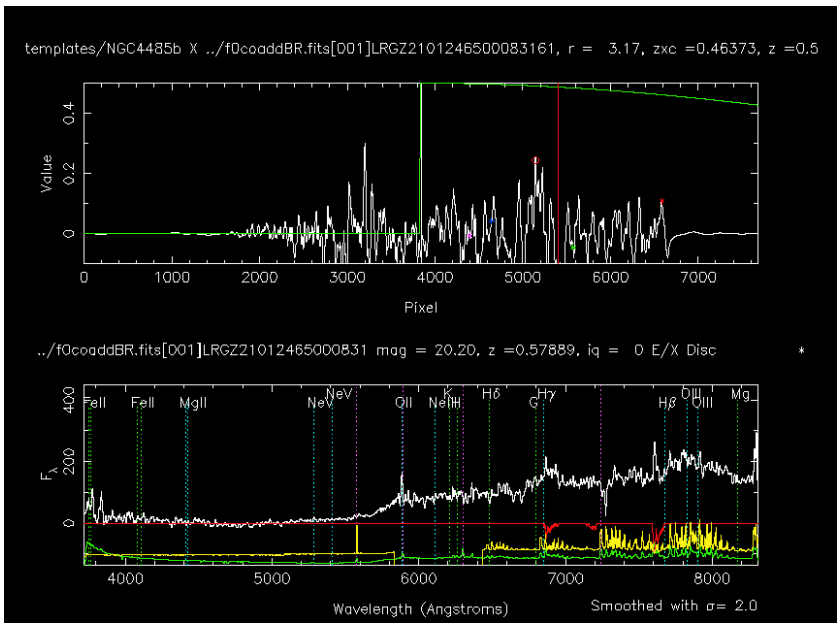


Figure 4.8: Screen shot of the RUNZ environment with an example galaxy spectrum.

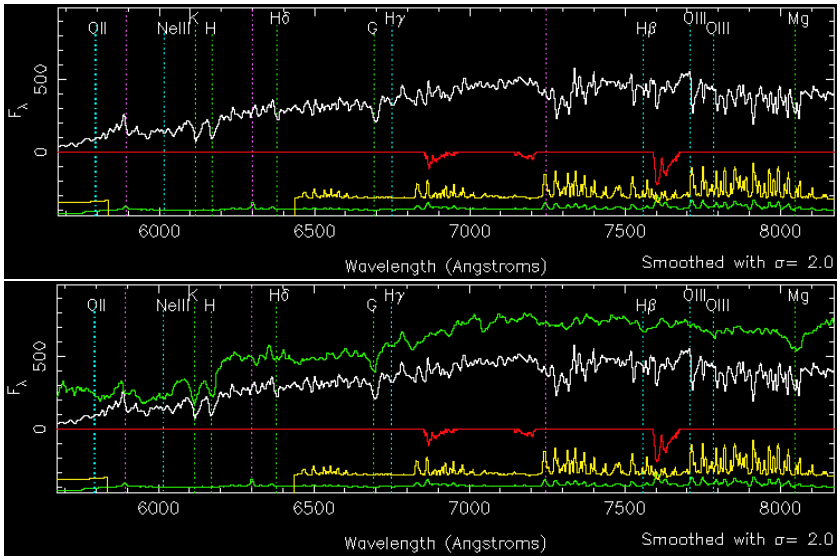


Figure 4.9: Close up of an example galaxy spectrum (top panel) and a template fitted to the galaxy spectrum (bottom panel) in RUNZ. The H, K and G absorption features can clearly be identified.

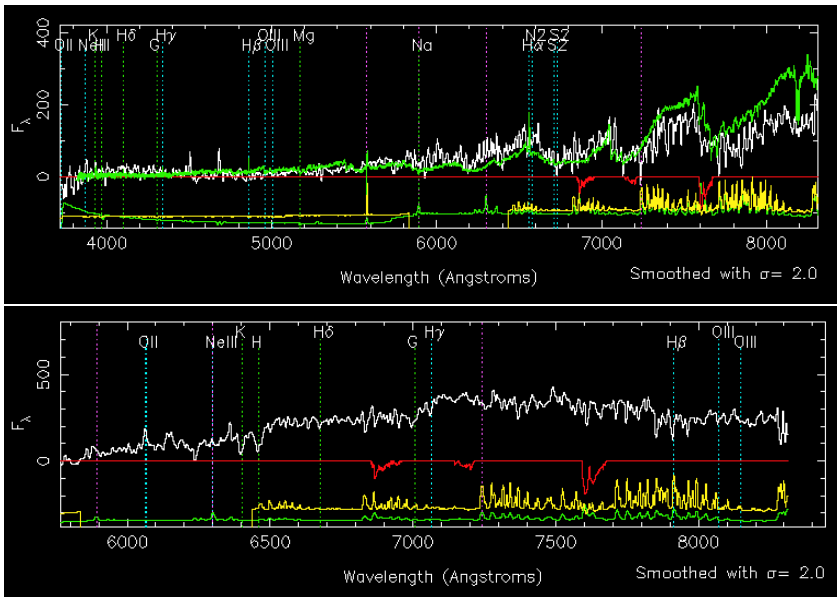


Figure 4.10: Close up of an example stellar spectrum (top panel) and a galaxy spectrum with visible OII emission line (bottom panel) in RUNZ.

Table 4.5: RUNZ Redshift Quality

Redshift Quality	Reliability
0	Redshift requires further examination
1	Poor quality spectrum, no redshift estimate
2	Possible redshift, requires re-observation
3	Probable redshift, $\sim 75\%$ confidence
4	Reliable redshift, $\sim 95\%$ confidence
5	Reliable redshift and high quality spectrum
6	Reliable redshift and high quality spectrum, but not extragalactic.

Table 4.6: Observed Spectra

Field	Number of Spectra	Number of Redshifts	Redshifter
1	531	329	SF, SB, KP
2	655	347	SF, SB, FBA
3	571	402	SF, SB, KP
4	401	306	SB
5	493	278	SF, CB
6	529	310	CB, KP
7	592	382	SF, SB, FBA
8	343	25	CB

of each redshift. Data from field 8 was not used as, owing to the poor quality of the data, an insufficient number of spectra were obtained for analysis. Field 4 was only redshifted once due to time constraints. Field 6 was only partially re-redshifted, therefore there are not enough secondary redshifts for comparison. Redshifting was carried-out by Filipe Abdalla (FBA), Chris Blake (CB), Sarah Brough (SB), Kevin Pimbblet (KP) and myself (SF).

Comparison between redshifts obtained independently by different redshifters, where available, shows a high level of consistency as seen in fig.4.11. For this study, a reliable redshift is defined as one that has a quality assigned by the redshifter between 3 and 5. In the case of disagreement between two redshifters, the value with the highest assigned quality was used. In the vary rare cases in which there was disagreement and both redshifters had assigned the same quality to the redshift, the spectrum was re-analysed. When three independent redshifts were available, a value was chosen that showed agreement between at least two redshifters regardless of the assigned qualities. In the rare case of disagreement between all three redshifters, the redshift with the highest assigned quality was chosen. Fig.4.11 shows that in the worst case $\sim 4\%$ of the redshifts may be incorrect for fields 1, 2, 3, 5 and 7. For the fields that were only redshifted once this number could be higher, however given the level of consistency between different redshifters, it is assumed that this value is a reasonable estimate of the error in the measured redshifts.

Table 4.6 shows the total number of spectra obtained for each observing field, the number of reliable redshifts obtained and by whom the field was redshifted. A total of 2379 reliable redshifts were obtained for the eight observing fields. Fig.4.12 shows $N(z)$ for the stacked galaxy redshifts, which peaks in the same range as the Mega$z$-LRG catalogue, $0.4 \leq z \leq 0.7$.

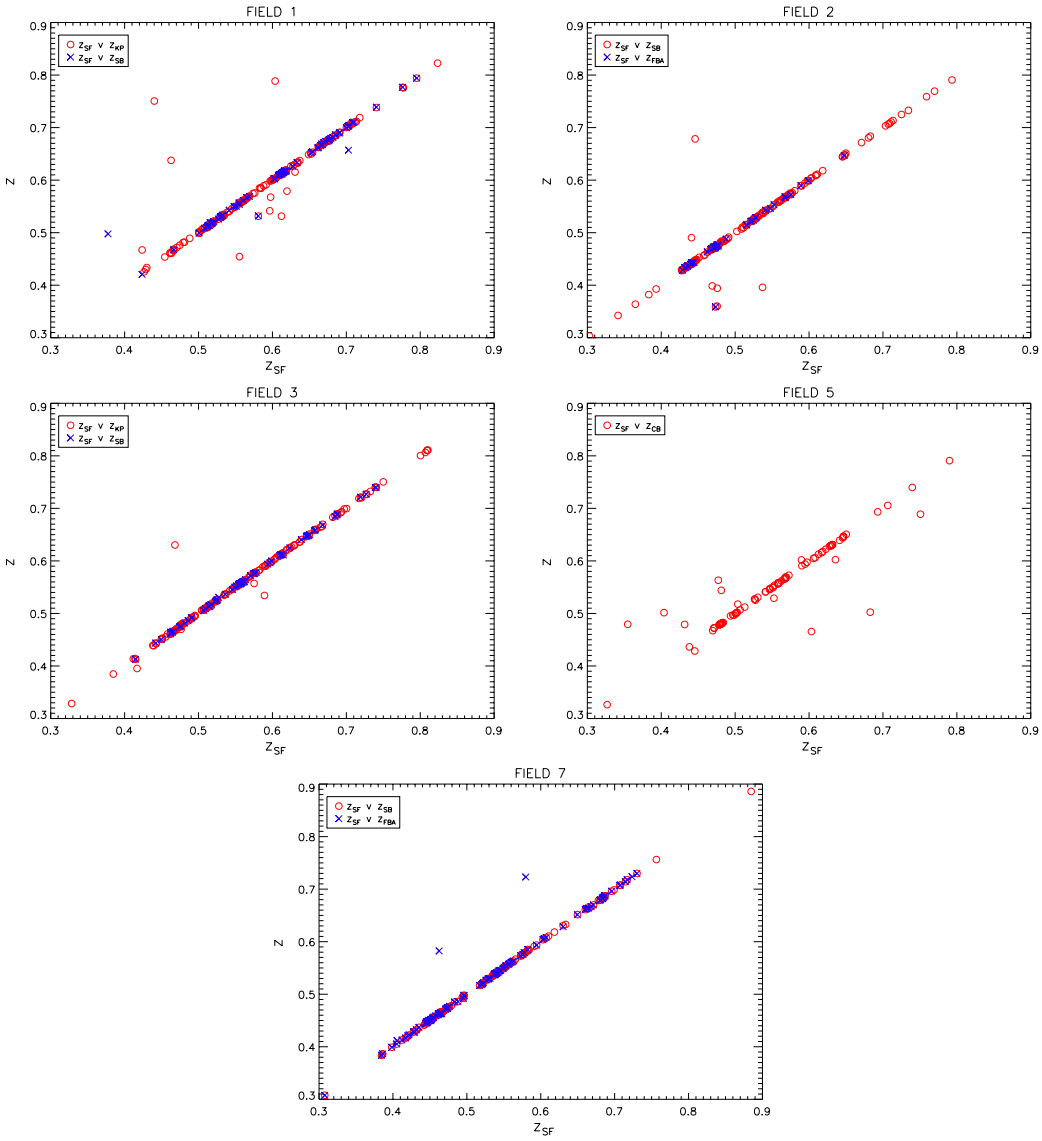


Figure 4.11: Comparison between redshifts obtained independently by different redshifters for fields 1, 2, 3, 5 and 7. The x-axis in each panel corresponds to a redshift obtained by SF and the y-axis to a value obtained by a second or third redshifter.

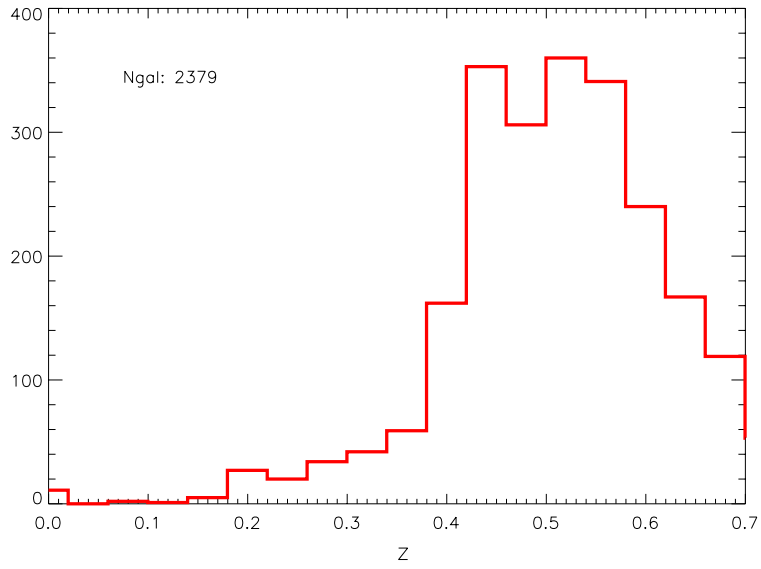


Figure 4.12: $N(z)$ for combined redshift data from each of the eight observing fields.

Fig.4.13-4.18 show $N(z)$ for all sample A clusters, which includes all the redshifted Mega$z$-LRG galaxies and any 2SLAQ galaxies that are within the observing regions. $N(z)$ was calculated in each panel by taking all galaxies within 1 h^{-1}Mpc of the sample A cluster centre. Each panel in fig.4.13-4.18 clearly shows a peak in z around the sample A cluster redshifts (indicated by a black dashed line in the plots). In certain panels, such as the first panel in fig.4.17 (field 5), the peak in z is very sharp and perfectly aligned with the cluster redshift. In other panels, such as the first panel in fig.4.14 (field 2), there appear to be multiple structures at different redshifts. It can also be seen that in some panels the z peak and the cluster redshift are not perfectly aligned, this can partially be attributed to photometric redshift bias. It should, however, also be noted that these cluster centres were found using an older version of the FoF code and the new and improved version should produce more accurate results.

Velocity dispersions were estimated for each of the Sample A clusters, where possible, using only the galaxies in the sharpest peak in the $N(z)$ distribution. The left panel of fig.4.19 shows a histogram of the velocity dispersions obtained compared to the velocity dispersions of the 2SLAQ mock galaxy catalogue presented in §2.4. The AAT data shows a peak at higher velocities (~ 850 kms^{-1}), however this is to be expected as the AAT targets were selected from the largest FoF cluster candidates, while the mock catalogue shows a range of velocities from a more complete sample of haloes of different sizes. The left panel of fig.4.19 shows a comparison between the velocity dispersions of AAT and mock data as function of richness. Although the AAT sample is

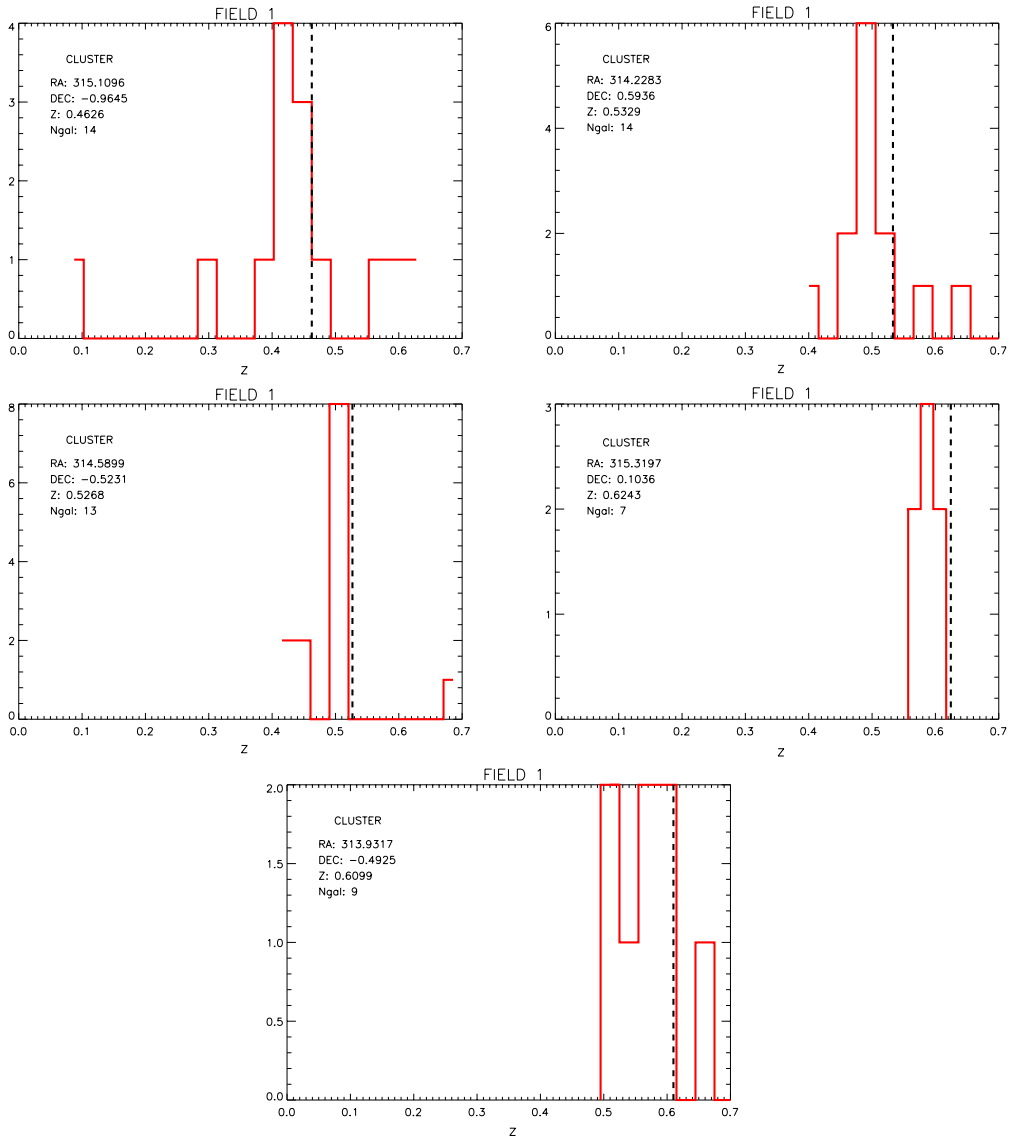


Figure 4.13: $N(z)$ of galaxies within 1 h^{-1}Mpc for sample A clusters in field 1. The black dashed line shows the location of the cluster redshift.

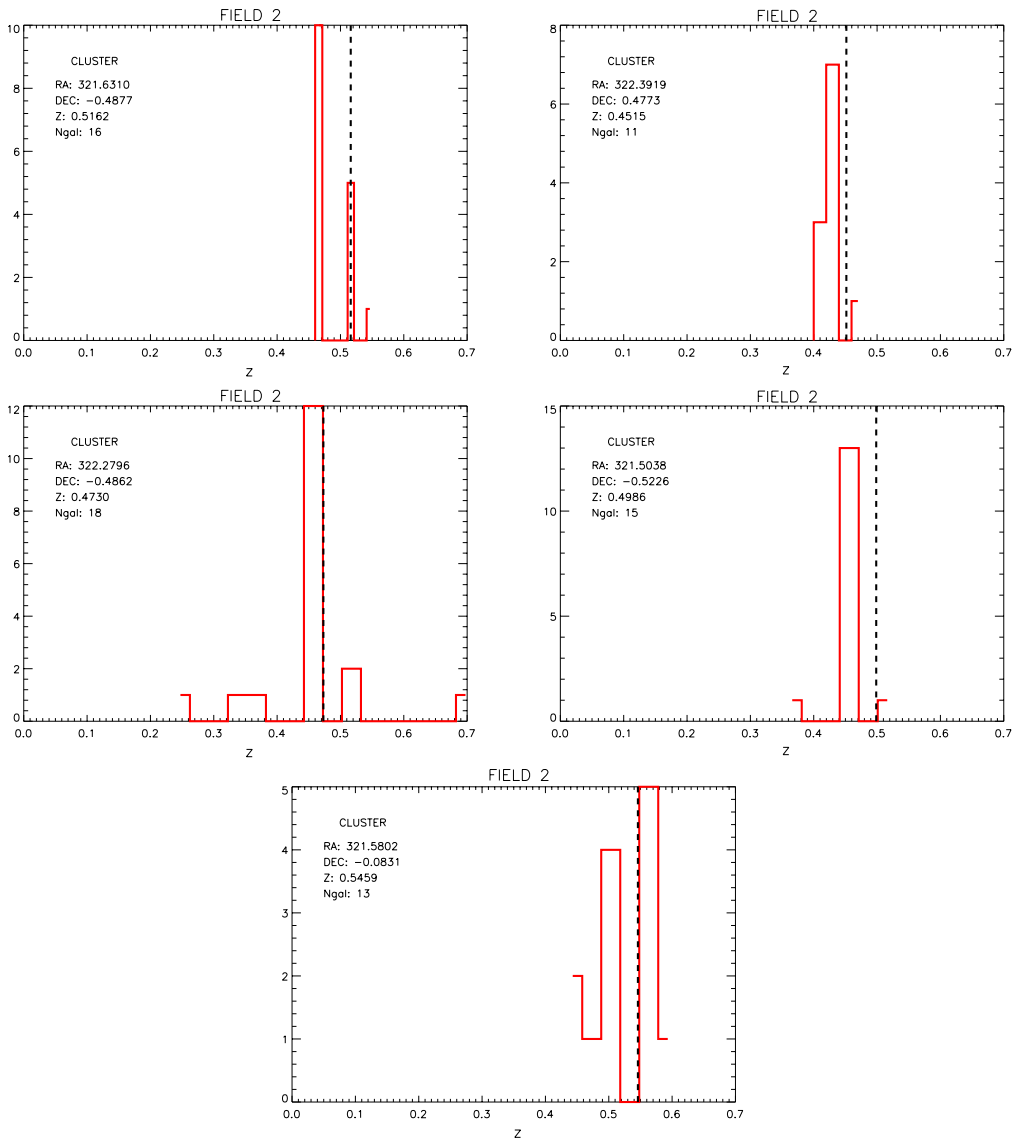


Figure 4.14: $N(z)$ of galaxies within 1 h^{-1}Mpc for sample A clusters in field 2. The black dashed line shows the location of the cluster redshift.

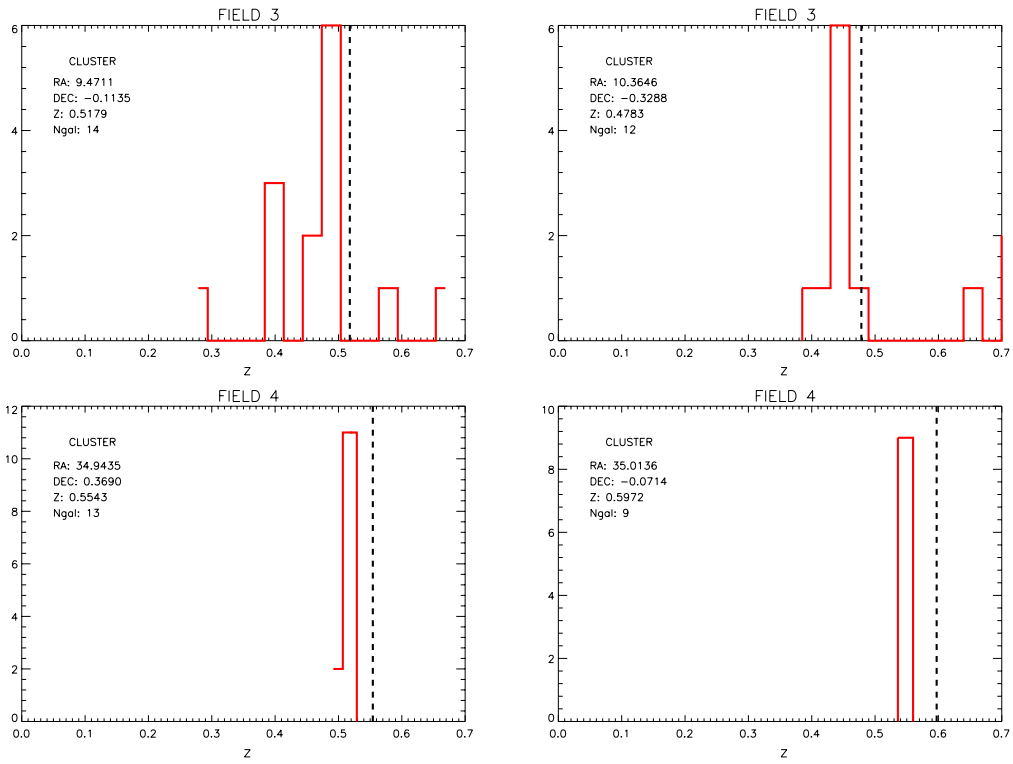


Figure 4.15: $N(z)$ of galaxies within 1 h^{-1}Mpc for sample A clusters in fields 3 and 4. The black dashed line shows the location of the cluster redshift.

much smaller, the two data sets show good agreement. This indicates that the velocity dispersions measured are reasonable given the number of LRGs in each cluster.

Mass estimates were also obtained for the clusters using equation 2.19. Only 25 redshifts have been obtained for field 8 due to poor weather on the last night of observing, therefore this field has not been analysed. Table 4.7 lists the positions of all the Sample A clusters and the number of Mega$z$-LRG and 2SLAQ galaxies within 1 h^{-1}Mpc of this position, the number of galaxies within the highest z peak, the average redshift of these galaxies and the their velocity dispersion. Regions where there are multiple significant peaks in z are identified with '*' and regions with no visible peak in z are identified with '-'. Clusters 5 and 18 are clear non-detections, and assuming that cluster with only a few galaxies in the peak and relatively high velocity dispersions (*e.g.* clusters 1, 17, 20, 23 and 27) are also inconclusive, then 24 out of the 31 clusters can be claimed as detections, which is a success rate of $\sim 77.4\%$. This is a crude analysis, but provides an estimate for how many real collapsed structures can be detected using the photometric friends-of-friends technique.

Fig.4.20 shows the distribution of the sample A cluster redshifts as a function of the average peak galaxy redshifts. The error bars show are the standard deviation of the galaxy redshifts in a

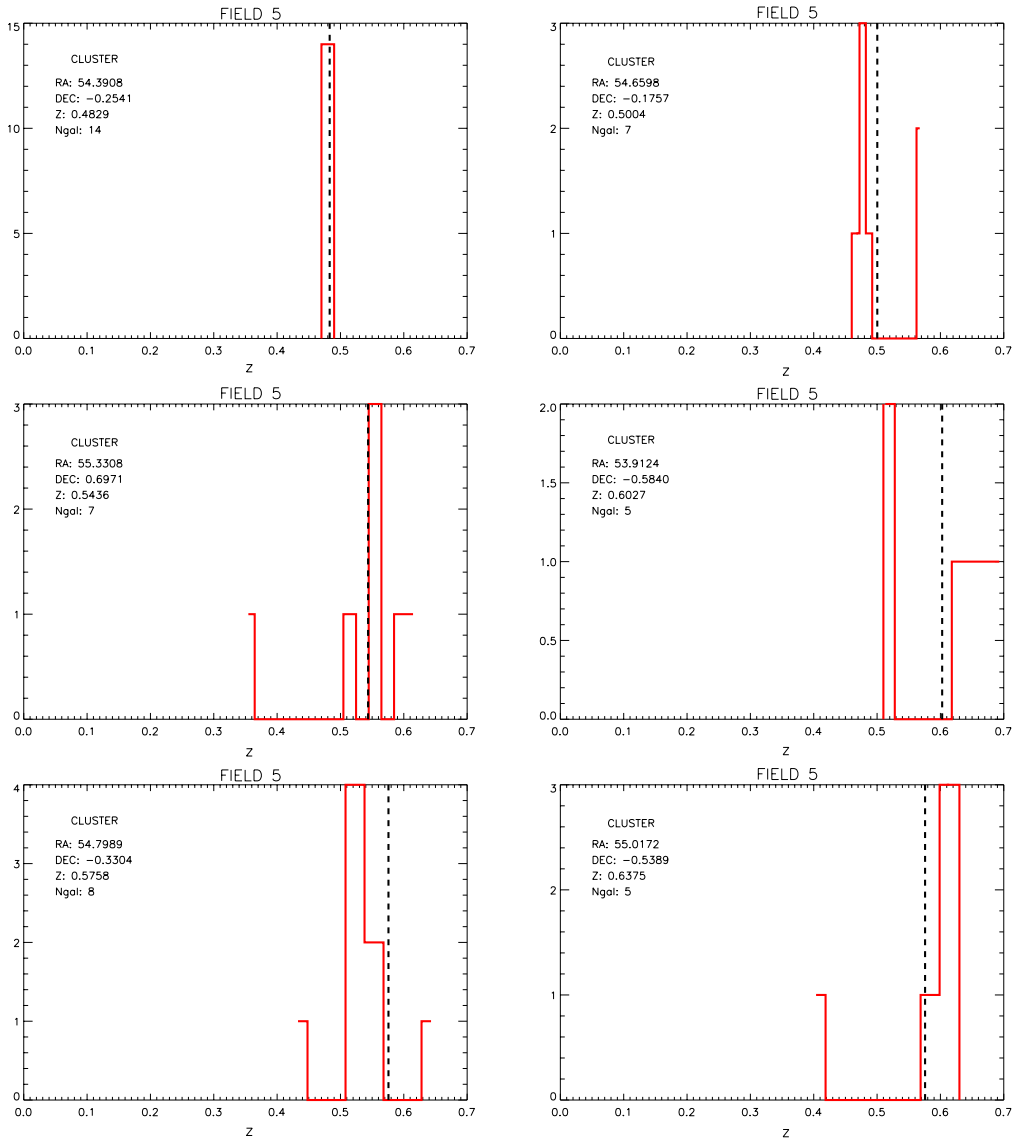


Figure 4.16: $N(z)$ of galaxies within 1 h^{-1}Mpc for sample A clusters in field 5. The black dashed line shows the location of the cluster redshift.

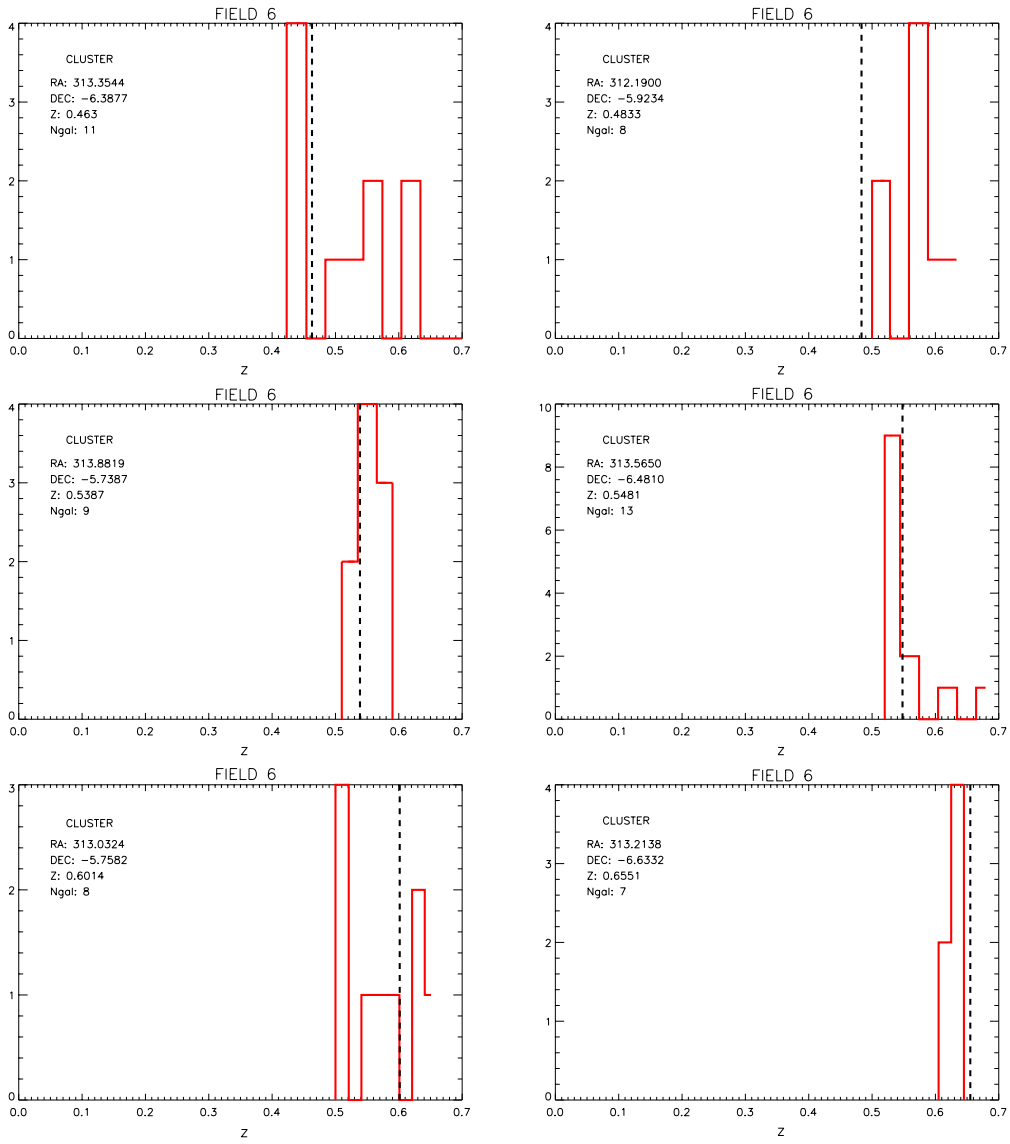


Figure 4.17: $N(z)$ of galaxies within 1 h^{-1}Mpc for sample A clusters in field 6. The black dashed line shows the location of the cluster redshift.

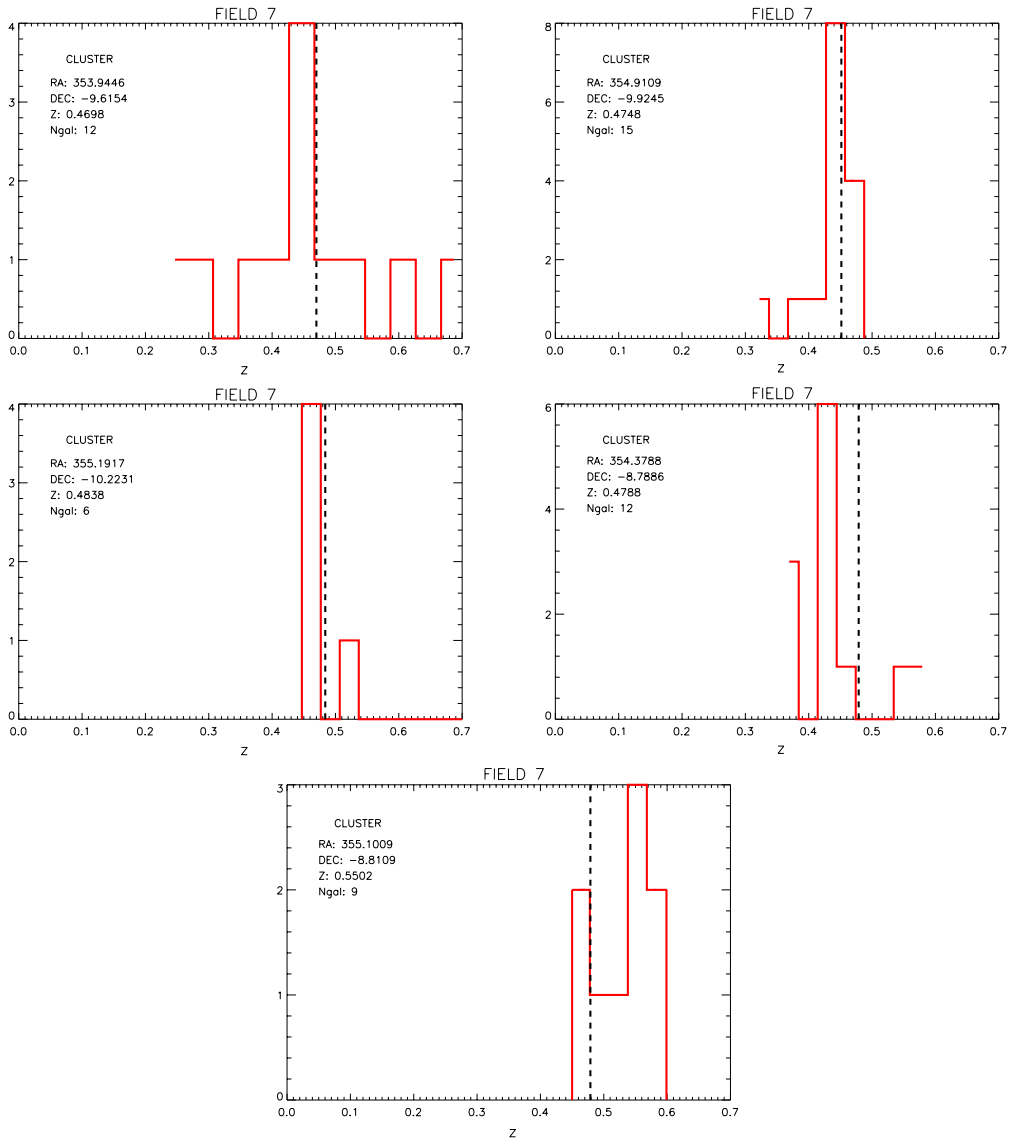


Figure 4.18: $N(z)$ of galaxies within 1 h^{-1}Mpc for sample A clusters in field 7. The black dashed line shows the location of the cluster redshift.

Table 4.7: Estimated velocity dispersions and masses for sample A clusters. ('*' Secondary peak in field, '-' non-detection)

N_o	RA_{clt} (Deg)	Dec_{clt} (Deg)	z_{clt}	Field	N_{tot}	N_{peak}	$\bar{z}_{gal}$	σ_{gal} (kms^{-1})	$\log_{10}$ M_{est} ($M_\odot$)
1	315.1096	-00.9645	0.4626	1	14	5	0.4205	1601.09	15.697
2	314.5899	-00.5231	0.5268	1	13	8	0.5129	0825.69	14.805
3	314.2283	+00.5936	0.5329	1	14	6	0.5191	0966.38	15.008
4	315.3197	+00.1036	0.6243	1	7	3	0.6000	0333.46	13.596
5	313.9317	-00.4925	0.6099	1	9	-	-	-	-
6	322.3919	+00.4773	0.4515	2	11	7	0.4377	0843.51	14.857
7	322.2796	-00.4862	0.4730	2	18	12	0.4750	0472.03	14.089
8	321.5038	-00.5226	0.4986	2	15	13	0.4715	1571.11	15.656
9	321.6310	-00.4877	0.5162	2	16	5	0.5237	0594.39	14.373
*9	321.6310	-00.4877	0.5162	2	16	10	0.4720	0768.81	14.725
10	321.5802	-00.0831	0.5459	2	13	5	0.5677	0287.37	13.412
*10	321.5802	-00.0831	0.5459	2	13	4	0.5209	1978.67	15.941
11	010.3646	-00.3288	0.4783	3	12	6	0.4609	0725.62	14.653
12	009.4711	-00.1135	0.5179	3	14	7	0.4917	0767.87	14.717
13	034.9435	-00.3690	0.5543	4	13	10	0.5371	0955.71	14.988
14	035.0136	-00.0714	0.5971	4	9	8	0.5549	1142.78	15.215
15	054.3908	-00.2541	0.4829	5	14	14	0.4799	1167.77	15.267
16	054.6598	-00.1757	0.5004	5	7	3	0.4817	0257.25	13.295
17	055.3308	+00.6971	0.5436	5	7	3	0.5648	1488.02	15.556
18	053.9124	-00.5840	0.6027	5	5	-	-	-	-
19	054.7989	-00.3304	0.5758	5	8	4	0.5510	0019.77	09.930
20	055.0172	-00.5389	0.6375	5	5	4	0.6155	0912.70	14.903
21	313.3544	-06.3877	0.4630	6	11	3	0.4396	0153.57	12.637
22	312.1900	-05.9234	0.4833	6	8	4	0.5822	0547.69	14.248
23	313.8819	-05.7387	0.5387	6	9	4	0.5632	0924.39	14.936
24	313.5650	-06.4810	0.5481	6	13	8	0.5355	0825.54	14.797
25	313.0324	-05.7582	0.6014	6	8	3	0.5123	0506.23	14.168
26	313.2138	-06.6332	0.6551	6	7	4	0.6362	0386.83	13.778
27	353.9446	-09.6154	0.4698	7	12	4	0.4513	1009.31	15.086
28	354.9109	-09.9245	0.4748	7	15	8	0.4633	0510.07	14.193
29	355.1917	-10.2231	0.4838	7	6	4	0.4622	0094.64	11.999
30	354.3788	-08.7886	0.4788	7	12	6	0.4348	0537.36	14.270
31	355.1009	-08.8109	0.5502	7	9	3	0.5611	0233.26	13.143

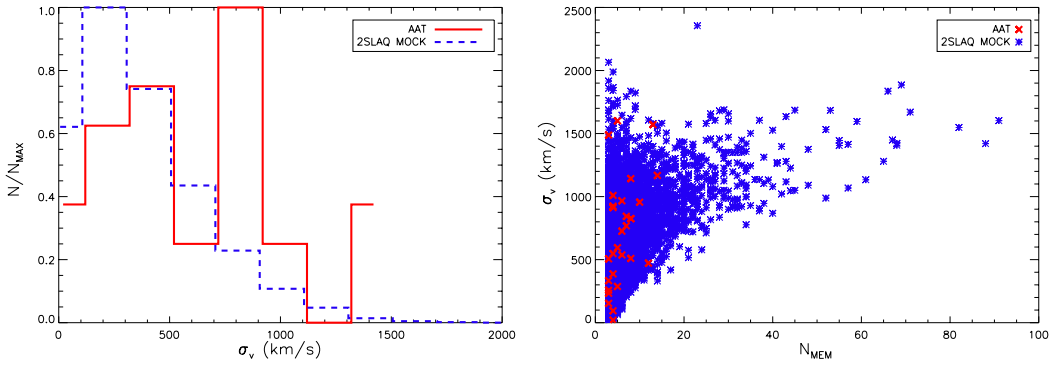


Figure 4.19: Left panel: Histogram of the velocity dispersions obtained from the $N(z)$ peak galaxies around each cluster compared to the velocity dispersions of the 2SLAQ mock galaxy catalogue. Right panel: velocity dispersions of the measured and mock data as function of cluster richness.

given cluster. This plot shows good agreement between the two redshifts, although the FoF code marginally overestimates the cluster redshift for the most part. This is an important discrepancy that would need to be addressed before any cosmological analysis could be performed using the photoz clusters.

It should be noted that the work presented in the section is in essence a pilot study principally aimed at determining the reliability of the FoF code to detect genuine structure. The velocity dispersions and mass estimates obtained are only meant to give a rough approximation for the clusters using the limited number of galaxy spectra. More reliable estimates could be obtained with more in depth spectroscopic analysis using the most recent version of the FoF code.

4.5 Conclusions & Future Work

Eight fields were targeted for observation based on cluster candidate positions found using a friends-of-friends (FoF) algorithm that incorporates photometric redshifts on the Mega$z$-LRG DR5 catalogue. The best candidates were chosen as those with seven or more LRGs and were labeled sample A clusters. The fields were supplemented with SDSS galaxies lying along the red-sequence of these clusters. 4116 spectra were obtained over four grey nights at the Anglo-Australian Observatory using AAOmega from the eight observing fields. From these 4116 spectra, 2379 galaxy redshifts were measured using the RUNZ package. Measuring $N(z)$ in regions 1 h^{-1}Mpc around a selection of the sample A clusters shows clear peaks in z around the cluster redshifts. Estimates are made for the cluster velocity dispersions and dynamical masses by taking

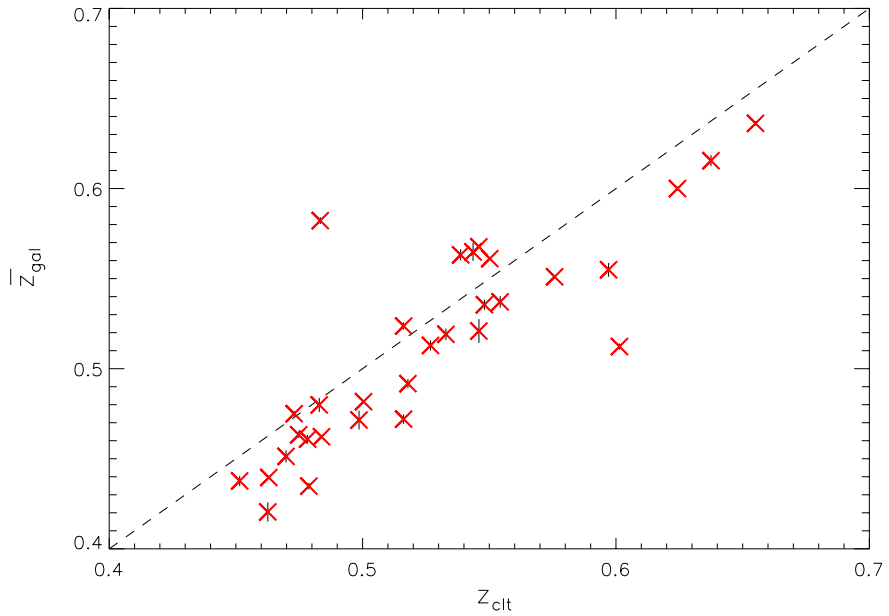


Figure 4.20: Average photometric redshift of sample A cluster galaxies, $z_{\rm clt}$, versus average spectroscopic redshift of galaxies around cluster centres, $\bar{z}_{\rm gal}$. Error bars are the standard deviation of the spectroscopic redshifts of galaxies in a given cluster.

only the galaxies within the $N(z)$ peak. The majority of these estimates are consistent with what would be expected for genuine structures, although these estimates are very approximate. A crude analysis of the expected physical properties for the clusters gives an estimate of a success rate of $\sim 77.4\%$.

These results can potentially be improved by running a better calibrated version of the FoF code on the Mega$z$-LRG catalogue. This may improve the accuracy of detecting the cluster centres. This code could also be run on a more recent release of Mega$z$-LRG. A more in depth spectroscopic analysis of this data could provide reliable velocity dispersion and dynamical mass estimates, which could be compared with lensing and/or X-ray data.

Further work that can be done using this data includes:

1. **Halo mass function:** One of the central predictions of CDM models is the trend seen in the number density of haloes as a function of mass and how this trend changes as a function of redshift (Sheth & Tormen 1999). For massive haloes, such as those considered in this project, the mass function is particularly steep and a sensitive probe of cosmology. The observations proposed here, which principally correspond to $z \approx 0.55$ (a look-back time of 5.4 Gyr), can be compared to measurements of low-redshift spectroscopic galaxy surveys such as the

SDSS (Rines *et al.* 2007).

2. **Clustering of massive haloes:** The scaling of the clustering strength of haloes with mass is an important prediction of CDM simulations (Younger *et al.* 2005). Since massive clusters are relatively scarce objects, large surveys like Mega$z$-LRG are required to measure the clustering properties accurately. The enormous volume probed by Mega$z$-LRG enables one to trace the rarest peaks in the density field.

3. **Halo occupation distribution:** The halo occupation distribution (HOD) is a useful point of comparison between models and data of CDM haloes (Seljak 2000). The HOD can be measured directly from the Mega$z$-LRG cluster catalogue, in addition to the radial profile of the satellite galaxies in halo, which is assumed to trace the dark matter halo density profile.

4. **Comparison with other cluster mass observables:** Dynamical masses estimated from the Mega$z$-LRG cluster velocity dispersions can be compared to those estimated from X-ray temperatures (*e*.g. Viana *et al.* 2003) and gravitational lensing (*e*.g. Cypriano *et al.* 2004). Each of these techniques on their own are subject to systematic errors: the velocity dispersion may contain an offset between cluster galaxies and dark matter particles; X-ray temperatures may have a non-equilibrium structure due to on-going merger events; and lensing measurements are subject to projection effects from additional mass along the line-of-sight. Therefore, cross-correlation between these three methods is essential for producing reliable mass estimates.

CHAPTER 5

THE DARK ENERGY SURVEY

5.1 Introduction

The term dark energy has been introduced to account for around 2/3 of the density in the Universe (Carroll *et al.* 1992) and to explain the observed acceleration in the expansion of the Universe (Perlmutter *et al.* 1998). Understanding the nature of dark energy has since become one of the most fundamental questions in cosmology. The Dark Energy Survey (DES) is one project that is aiming to shed some light on this problem (The Dark Energy Survey Collaboration 2005). DES will employ four distinct methods to probe the properties of dark energy, namely baryon oscillations of the power spectrum, the abundance and spatial distribution of clusters, weak gravitational lensing and type Ia supernovae. This project is a large undertaking involving institutions from all over the world.

The following sections outline the details of the DES and work that has been carried out so far using DES simulations from Stanford (DC4) and the Institut de Ciències de l'Espai (ICE). Section 5.2 describes the DES project and instruments. Section 5.3 describes two different simulations of the data expected from DES. Finally, section 5.4 presents work that has been carried out so far using the DES simulations.

5.2 Survey Overview

DES is an optical-near infrared survey of 5000 deg^2 of the southern galactic cap out to $\sim 24^{th}$ magnitude in the SDSS *g, r, i* & *z* bands. The survey will be conducted with the Blanco Telescope using the newly developed Dark Energy Camera (DECam) and is expected to run from 2012 to

2017. The overall objective of the survey will be to measure the dark energy and dark matter densities and the dark energy equation of state to high accuracy.

The project was originally proposed in response to a competition announced in December of 2003 by the National Optical Astronomy Observatory (NOAO) to build an instrument for the Blanco telescope in return for 30% observing time over a 5 year period. Once completed, the constraints from DES will allow certain models of dark energy to be discarded and should provide a better understanding of what effects dark energy actually has on the Universe.

This section will give a brief description of the Blanco telescope (§5.2.1), DECam (§5.2.2) and the science objectives of DES (§5.2.3).

5.2.1 The Blanco Telescope

The Blanco telescope is a 4 m reflector telescope based at the Cerro Tololo Inter-American Observatory (CTIO), which is at an altitude of 2200 m and is situated just east of La Serena, Chile. The telescope was commissioned in 1974 and is named after the Puerto Rican astronomer Victor Manuel Blanco.

CTIO is located in the Atacama desert where there is virtually no rainfall and the median seeing is in the around $0.6 - 0.7$ arcseconds, which makes it an ideal location for obtaining high quality imaging data.

5.2.2 The Dark Energy Camera

DECam is a 519 mexapixel optical CCD camera with SDSS *g, r, i* & *z* filters (fig.5.1). The camera was designed to have a large field of view (2.2°), high throughput and good image quality. The optical system consists of 5 lenses, C1, C2, C3, C4 & C5, with diameters 980 mm, 690 mm, 652 mm, 604 mm and 512 mm respectively. DECam has a low noise CCD readout and is cryogenically cooled to 180 K.

There are 62 2k×4k (0.27"/pixel) CCDs arranged in a hexagon covering 3 deg^2 (fig.5.2). Each individual CCD is a 250 μm thick silicon device, which is optimised to detect high redshift *z*-band images.

The full DECam apparatus, which weighs around four tons, will be placed at the prime focus of the Blanco Telescope in Chile.

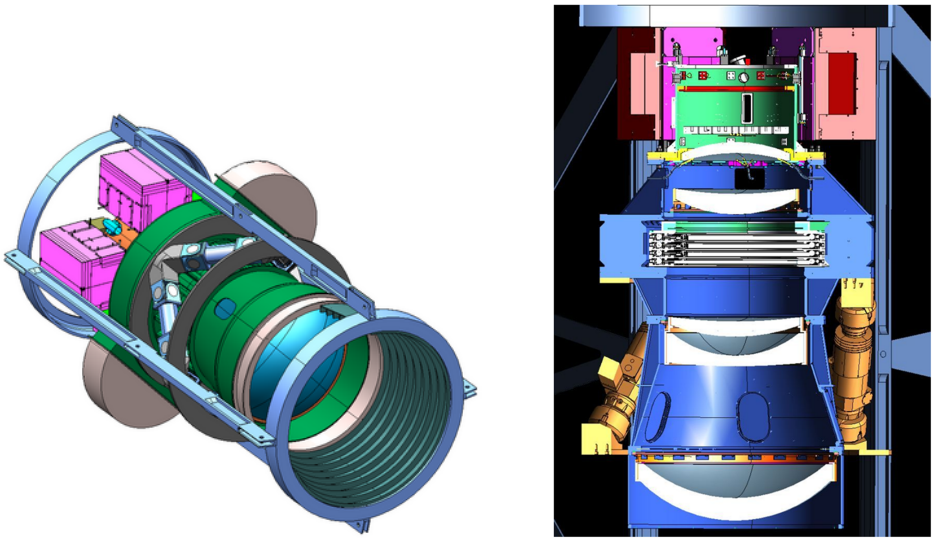


Figure 5.1: DECam representation.

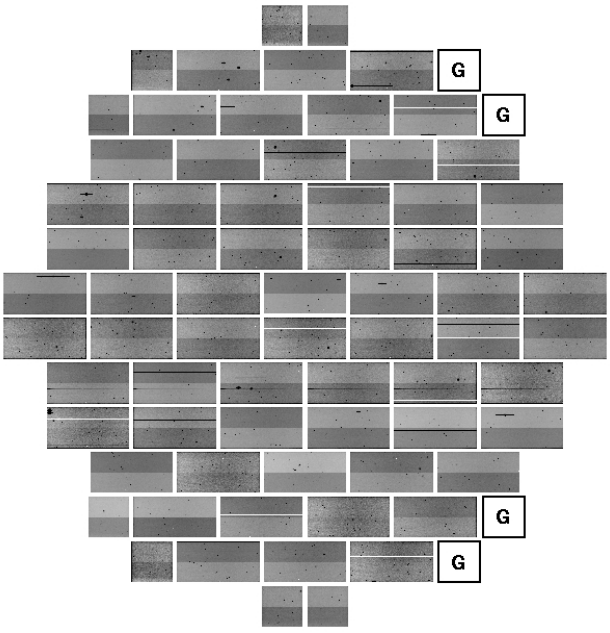


Figure 5.2: DECam CCD layout.

5.2.3 Science Objectives

DES aims to measure the dark energy and dark matter densities and the dark energy equation of state using four independent methods: galaxy clusters, weak lensing tomography, galaxy angular clustering and supernovae distances. Each of these probes are subject to different systematics and therefore combining them will allow a cross check of the systematic errors, which will enable higher precession measurements of dark energy.

Galaxy clusters are thought to be a good probe of cosmology because they only involve gravitational dynamics and dark matter. DES aims to obtain detailed optical measurements of galaxy clusters, which will allow accurate photometric redshift estimates. The galaxy cluster data will be complimented with results from the South Pole Telescope (SPT) survey, which uses the Sunyaev-Zel'dovich effect to detect clusters. The SPT is a 10 m telescope located at the Amundsen-Scott South Pole Station in Antarctica. The telescope observes in a frequency range 70-300 GHz.

Weak lensing is sensitive to the expansion history of the Universe through the shear-shear angular power spectrum and galaxy-shear correlations. DES aims to measure shapes for $\sim 3 \times 10^8$ galaxies with improved image quality.

Measurements of the peaks and wiggles found in the matter power spectrum from galaxy angular clustering can be used a distance measures. DES aims to measure angular clustering in photometric redshift bins out to $z \sim 1.1$.

The light curves of Type Ia supernovae can be used as standard candles in cosmological distance scales. DES will use 10% of its observing time to measure *r, i* & *z* light curves for $\sim$1900 supernovae in the range $0.3 < z < 0.75$ in a region of 40 deg^2.

The combination of the high precision DECam data for each of these probes will provide the most accurate test of dark energy to date.

5.3 Simulations

Once DES begins observations, a huge amount of data will become available, which will require a lot of work to reduce and analyse. It is therefore productive to establish reliable pipelines and resources before operations begin. To this end, several high quality simulations of the data expected from DES have been produced, which can be used to test and refine the data pipelines. These simulations are not only of galaxy catalogues, but also of the expected images from the CCD fields.

This section will briefly discuss the mock galaxy catalogues produced in Barcelona (§5.3.1)

and Stanford (§5.3.2).

5.3.1 Barcelona

The Barcelona simulations were produced by Jorge Carretero, Francisco Javier Castander, Anna Cabré, Ricard Casas, Martin Crocce, Pablo Fosalba and Enrique Gaztañaga at the Institut de Ciències de l'Espai (ICE).

The N-body simulation is a snapshot at $z = 0$ (*i.e.* with no evolution) of 2048^3 particles in a box of 3072 Mpch^{-1} in length made with the GADGET-2 code. The cosmological parameters used are for a flat universe with $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$, $h = 0.7$, $\sigma_8 = 0.8$ and $n_s = 0.9$. The dark matter haloes are found with a friends-of-friends algorithm, which requires a minimum of 20 particles to form a halo. The mass of each dark matter particle is $m_p = 23.42 \times 10^{10} M_\odot h^{-1}$.

5.3.1.1 Galaxy Assignment

Galaxies are assigned to haloes following the approach of Skibba *et al.* (2006). This method uses the HOD interpretations from the SDSS measurements of the luminosity dependence of clustering by Zehavi *et al.* (2005), which was motived by the work of Kravtsov *et al.* (2004). Zehavi *et al.* (2005) find that the HOD model assumes a simplified form when the galaxies are separated into central and satellite members.

From this approach, the mean number of galaxies with luminosity greater than L in haloes of mass M is given by

$$N_{\rm gal}(> L_r|M_{\rm halo}) = 1 + N_{\rm sat}(> L_r|M_{\rm halo}) = 1 + \Big[\frac{M_{\rm halo}}{M_1(L)}\Big]^{\alpha(L)}, \tag{5.1}$$

where M_{min} is the minimum halo mass for hosting at least one central galaxy, M_1 is the mass of a halo that on average hosts one satellite galaxy and α is the power-law slope of the satellite mean occupation function. Equation 5.1 holds If $M \geq M_{\rm min}(L)$, otherwise $N_{\rm gal}(M) = 0$ (Kravtsov *et al.* 2004; Zehavi *et al.* 2005). In practice, $M_{\rm min}(L)$ is a monotonic function of L. The results are well approximated by

$$\Big(\frac{M_{\rm min}}{10^{12} h^{-1} M_\odot}\Big) \approx \exp\Big(\frac{L}{9.9 \times 10^9 h^{-2} L_\odot}\Big) - 1, \tag{5.2}$$

where $M_1(L) \approx 23 M_{\rm min}(L)$, and $\alpha \sim 1$ (Skibba *et al.* 2006).

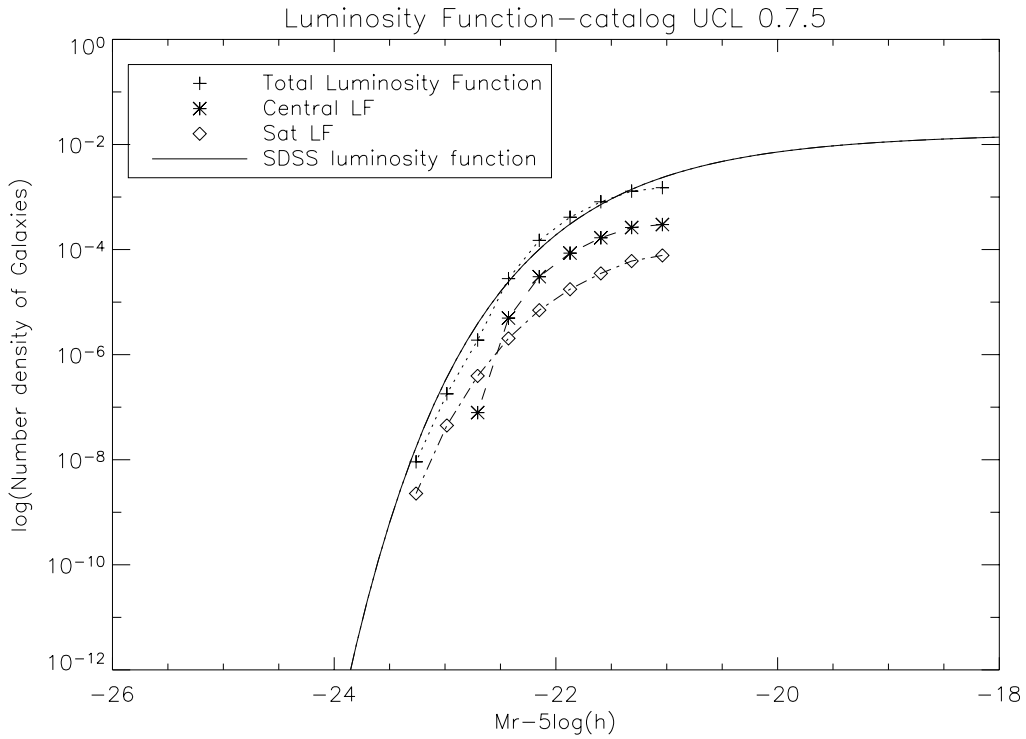


Figure 5.3: Luminosity function of Barcelona mock catalogue produced by Jorge Carretero Palacios. Smooth curve is SDSS luminosity function (Blanton *et al.* 2003).

5.3.1.2 Luminosities

All haloes in the simulation have at least one central galaxy with luminosity given by 5.2. Absolute $r-$band magnitudes are calculated via

$$M_r - 5\log_{10}(h) = -2.5\log_{10}(L_r) + M_{r\odot}, \tag{5.3}$$

where $M_{r\odot} = 4.76$ (Blanton *et al.* 2003).

The luminosity function for the central and satellite galaxies is shown in fig.5.3, where the there are 30,468,972 galaxies in total, 24,659,420 of which are centrals and the remaining 5,809,552 are satellites. This plot shows that the mock luminosity function is a good fit to SDSS.

5.3.1.3 Colours

Central galaxy colours are modelled via

$$\langle g - r|M_r\rangle_{\rm red} = 0.932 - 0.032(M_r + 20), \tag{5.4}$$

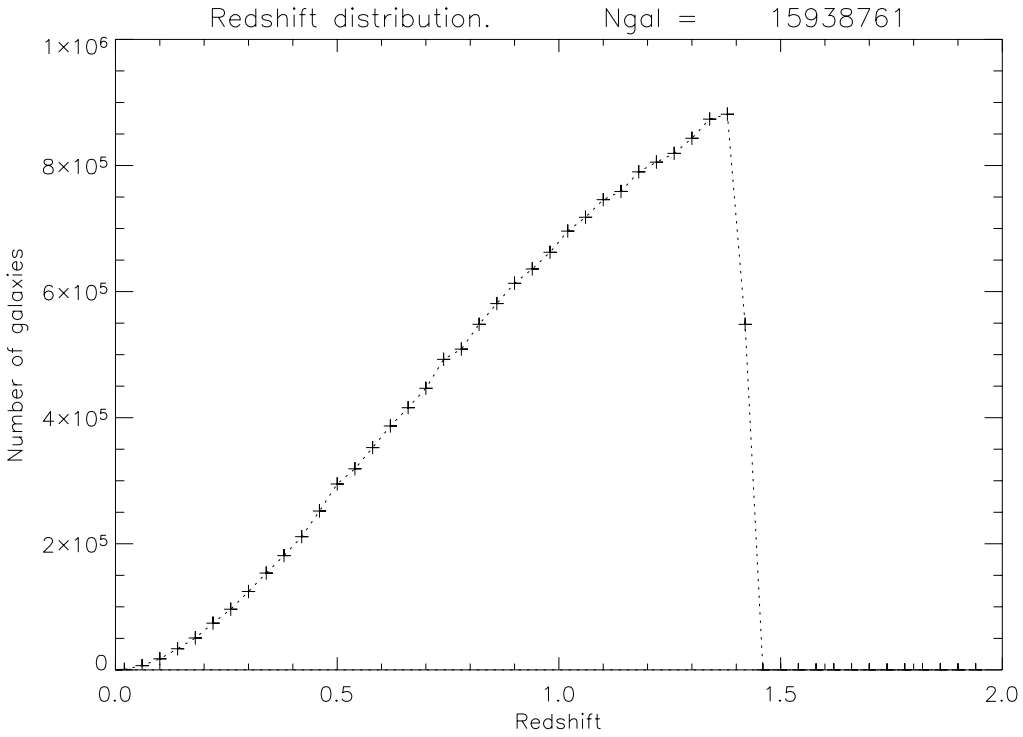


Figure 5.4: $N(z)$ of Barcelona mock catalogue produced by Jorge Carretero Palacios.

$$rms(g-r|M_r)_{\rm red} = 0.07 - 0.01(M_r + 20), \tag{5.5}$$

$$\langle g-r|M_r\rangle_{\rm blue} = 0.62 - 0.11(M_r + 20), \tag{5.6}$$

and

$$rms(g-r|M_r)_{\rm blue} = 0.12 - 0.02(M_r + 20), \tag{5.7}$$

where 5.4 and 5.5 are for central galaxies in the red sequence, and 5.6 and 5.7 are for blue central galaxies (Blanton *et al.* 2003). Satellite galaxies are better approximated by

$$\langle g-r|M_r\rangle_{\rm sat} = 0.83 - 0.08(M_r + 20). \tag{5.8}$$

5.3.1.4 Positions and Velocities

Satellite galaxies are placed in the dark matter haloes following an NFW profile (Navarro *et al.* 1996), where the spherically averaged density of the halo is given by

$$\frac{\rho(r)}{\bar{\rho}} = \frac{\Delta_{\rm vir}(z)}{3\Omega(z)} \frac{c^3 f(c)}{x(1+x)^2}, \tag{5.9}$$

and $x = c(M_{\rm halo} r / r_{\rm vir})$. c is the concentration parameter of the halo, which is approximated by

$$c(M_{\rm halo}) \approx \frac{9}{(1+z)} \Big(\frac{M_{\rm halo}}{M_{*0}} \Big), \tag{5.10}$$

(Bullock *et al.* 2001). M_{*0} is the standard nonlinear mass scale with the assumed value $M_{*0} = 2 \times 10^{13} M_{\odot}$.

5.3.1.5 Galaxy Types

Galaxies are assigned a probability of having a certain type based on their redshifts. The mock catalogues of Oyaizu *et al.* (2006) and Lin *et al.* (2004) are used to produce a catalogue of galaxy types. The distribution of galaxy types is derived using data from the GOODS/HDF-N field (Capak *et al.* 2004; Cowie *et al.* 2004; Wirth *et al.* 2004) and CWW template SEDs (Coleman *et al.* 1980).

5.3.2 Stanford

The Stanford Data Challenge 4 (DC4) simulation was produced by Risa Wechsler and Michael Busha using a 573 deg^2 subset of the ΛCDM Hubble Volume Simulation and covers 5000 deg^2 on the sky out to redshift $z \sim 1.4$. The mock is magnitude limited to $0.4L^*$ in the *r*-band, and uses $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$ and $\sigma_8 = 0.9$. Galaxies are added to the simulation using the ADDGALS routine, which assigns the properties of real SDSS galaxies to the mock galaxies.

5.3.2.1 Hubble Volume Simulations

The Hubble Volume Simulations are N-body simulations of 1×10^9 dark matter particles, where the mass of each particle is $m_p = 2.2 \times 10^{12} M_{\odot} h^{-1}$, in boxes of 3000 Mpc$h^{-1}$ and 2000 Mpch^{-1} in length. The project was aimed at studying clusters of galaxies, filaments and voids (Evrard *et al.* 2002). The 3000 Mpch^{-1} box was run with standard ΛCDM cosmology ($\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $\sigma_8 = 0.9$), while the 2000 Mpch^{-1} box was run with a cosmology dominated by non-relativistic cold dark matter, τCDM ($\Omega_M = 1$, $\Omega_\Lambda = 0$ and $\sigma_8 = 0.6$).

Fig.5.5 shows a lightcone from the ΛCDM Hubble Volume Simulation, which illustrates the evolution of clustering along the line of sight.

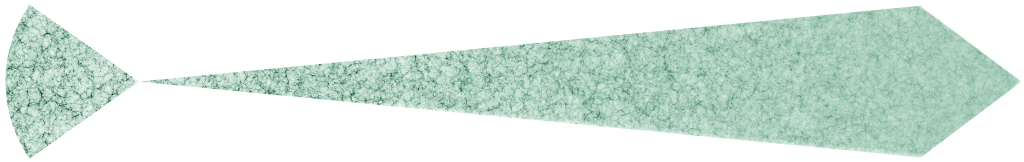


Figure 5.5: Lightcone, which shows the evolution of clustering in the Hubble Volume Simulation (Evrard *et al.* 2002).

5.3.2.2 Galaxy Assignment

The ADDGALS routine takes real galaxies from the luminosity function of Blanton *et al.* (2003) in the SDSS *r*-band with passive evolution of $M^* = 1.3$ per unit redshift irrespective of galaxy type. Each galaxy is assigned to a dark matter particle in the simulation by matching the luminosity-dependent clustering in the local Universe. No redshift evolution is used beyond the local SDSS colour-environment corrections.

The mock galaxies are assigned to SDSS galaxies with the same L/L^* and local galaxy density. Each mock galaxy then has apparent magnitudes in the SDSS *g, r, i, z* & *Y* bands. Brightest Cluster Galaxies (BCGs) are inserted with $L \sim M^{0.3}$ and $\sigma(\log L) = 0.15$, based on Hansen *et al.* (2009), only for haloes with $M > 1 \times 10^{14} M_{\odot}$.

5.4 Data Analysis

This section details work that has been done thus far using the DES simulations described in the previous section. Section 5.4.1 presents work on estimating photometric redshifts for the DES simulations. Section 5.4.2 highlights work done regarding the galaxy colours in the DES simulations. Finally, section 5.4.3 shows the start of work involving cluster detection in the DES simulations using galaxy photometric redshifts.

5.4.1 Photometric Redshifts

Photometric redshifts were estimated for the Stanford DC4 mock using ANNz, the neural network code of Collister & Lahav (2004). 50,000 random galaxies were selected from the mock with SDSS *g, r, i, z* & *Y* photometric bands. Since each mock galaxy has an associated redshift, these random galaxies can be used to train the neural network. 30,000 of these galaxies were used as the training set and the remaining 20,000 were used as a validation set to prevent over-fitting. The remaining mock galaxies not used for training were run through the trained network assuming no knowledge

of the spectroscopic redshifts. The ANNz architecture was 2 layers of 10 nodes. The resulting photometric redshifts obtained from ANNz versus the spectroscopic redshifts are shown in the top left panel fig.5.6. The fit to the $x = y$ line seen in this plot is very good, unfortunately it does not look very realistic. Comparing the top left panel with the top right panel of fig.5.6 from Banerji *et al.* (2008) shows that the scatter is at least two times smaller in the Stanford mock. The Banerji *et al.* (2008) photometric redshifts were produced by running ANNz on DES mocks produced by Lin *et al.* (2004) and Oyaizu *et al.* (2006) with SDSS *g, r, i, z* & *Y* photometric bands.

In an attempt to account for the discrepancies between the top left and right panels of fig.5.6, a second set of photometric redshifts were produced from a version of the Stanford DC4 mock without any magnitude cuts. 150,000 random galaxies were selected from the uncut mock, 100,000 were used as the training set and 50,000 were used as the validation set. The bottom left panel of fig.5.6 shows the photometric redshifts obtained from the uncut mock using ANNz versus the spectroscopic redshifts. This plot shows a very similar scatter to that seen in the top right panel of fig.5.6, which implies that the problem lies with the apparent magnitudes in the Stanford DC4 mock catalogue.

The source of the problem was found to be the result of magnitude errors that were not effected by evolution. Later versions of the DC4 mock and now the DC5 mock have since fixed this problem as shown in the bottom right panel of fig.5.6. This plot was created by Stephanie Jouvel using the photometric redshifts available in the current DC5 mock, which were produced using the neural network algorithm of Oyaizu *et al.* (2008*a*,*b*). The red and green lines correspond to $z_p - z_s = 0.15$ and $z_p - z_s = 0.3$ respectively.

Photometric redshifts were also produced for the Barcelona mock using ANNz. A sample of 47,000 random galaxies were used as the training set and another 23,000 were used as the validation set. The ANNz architecture used was the same as that used for the Stanford mock, 2 layers of 10 nodes. The resulting photometric redshifts versus spectroscopic redshifts are shown in fig.5.7. This figure shows a similar scatter to that seen in fig.5.6, which implies the Barcelona mock suffers from the same problem as the DC4 versions of the Stanford mock. To investigate this an analysis of the colour versus colour space of the Barcelona and DC4 Stanford mocks was made, which is discussed in the following subsection.

5.4.2 Colours

Fig.5.8 shows *g*-*r* vs. *r*-*i* colours (top left panel), *r*-*i* vs. *i*-*z* colours (top right panel) and *i*-*z* vs. *z*-*Y* colours (bottom panel) for the Barcelona mock, where the colour bar indicates the number of

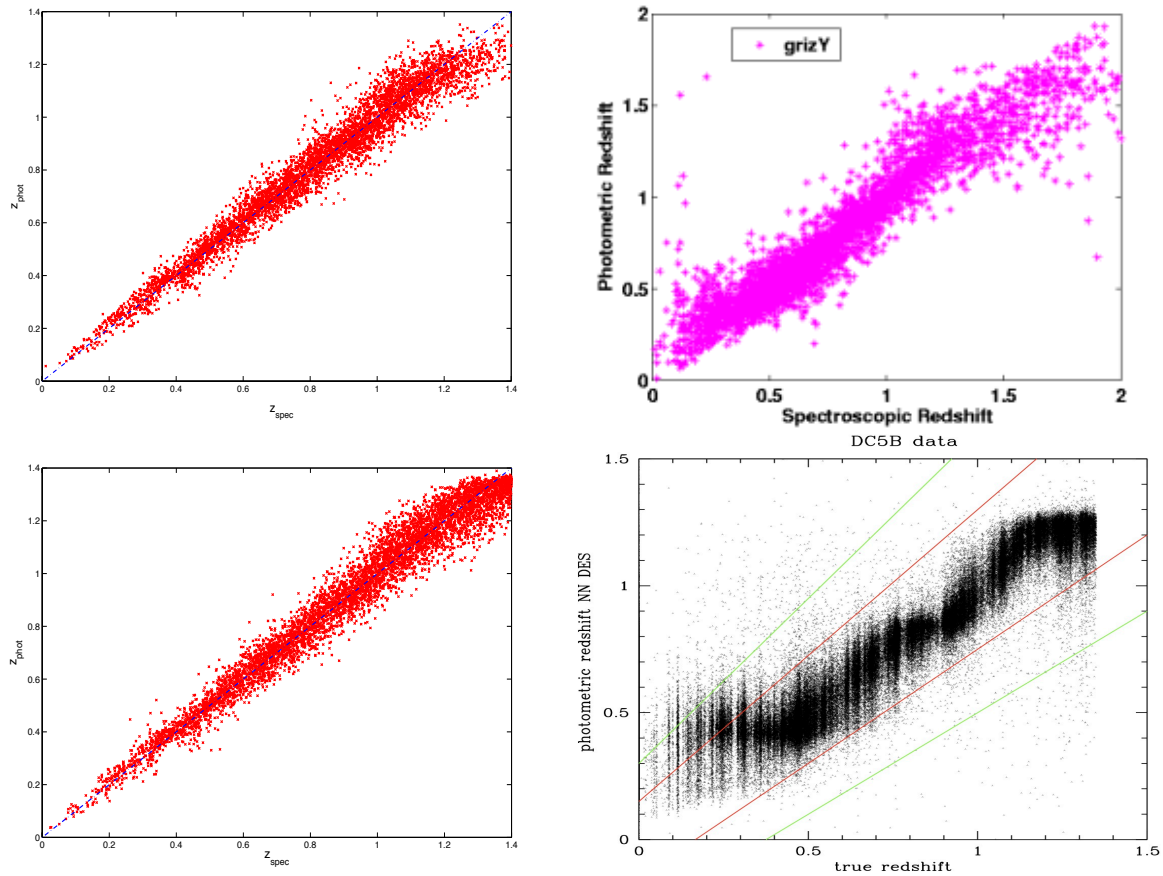


Figure 5.6: DC4 mock photometric redshifts versus spectroscopic redshifts using SDSS *g, r, i, z* & *Y* photometric bands (top right panel). Photometric redshifts versus spectroscopic redshifts using SDSS *g, r, i, z* & *Y* photometric bands from Banerji *et al.* (2008) produced using mocks produced by Lin *et al.* (2004) and Oyaizu *et al.* (2006) (top right panel). Uncut DC4 mock photometric redshifts versus spectroscopic redshifts using SDSS *g, r, i, z* & *Y* photometric bands (bottom right panel). DC5 mock photometric redshifts Oyaizu *et al.* (2008*a*,*b*) versus spectroscopic redshifts produced by Stephanie Jouvel (bottom right panel).The red and green solid lines correspond to $z_p - z_s = 0.15$ and $z_p - z_s = 0.3$ respectively.

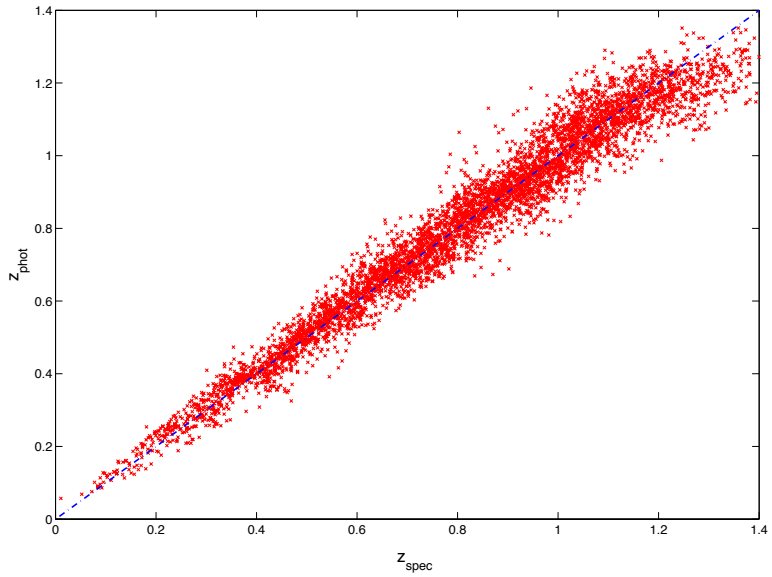


Figure 5.7: Barcelona mock photometric redshifts versus spectroscopic redshifts using SDSS *g, r, i, z* & *Y* photometric bands.

galaxies. These plots show the spectral energy distribution (SED) tracks for the mock galaxies. To test the reliability of the galaxy colours, these SED tracks are compared to the Stanford DC4 mock and to the mock of Lin *et al.* (2004).

Fig.5.9 shows the overlap of the colour space contours between the Barcelona mock and the Lin *et al.* (2004) mock (top panels), and between the Barcelona mock and the Stanford DC4 mock (bottom panels). This figure shows that there is relatively good agreement between the Barcelona mock and that of Lin *et al.* (2004), whereas there is much less overlap with the Stanford DC4 mock.

The overlap between the Barcelona and Lin *et al.* (2004) mocks is more clearly illustrated in the colour maps without magnitude errors shown in fig.5.10. Fig.5.11 shows the scatter in *g-r* vs. *r-i* colour space as a function of redshift for the Barcelona mock (left panel) and the Lin *et al.* (2004) mock (right panel). This plot shows that the colour distribution with respect to redshift is the same in both mocks, however the Lin *et al.* (2004) mock show a much more uniform spread of galaxy colours. Fig.5.12 shows the change in the *g-r* colour as a function of redshift in the Barcelona mock.

These plot have helped to understand the problems in the Barcelona simulation, which has led to development of improved versions with more realistic galaxy colours and evolution. This project is still ongoing and it yet to be seen how realistic the photometric redshifts will be from the current version. The aim of this project is to produce a mock galaxy catalogue, which is competitive with

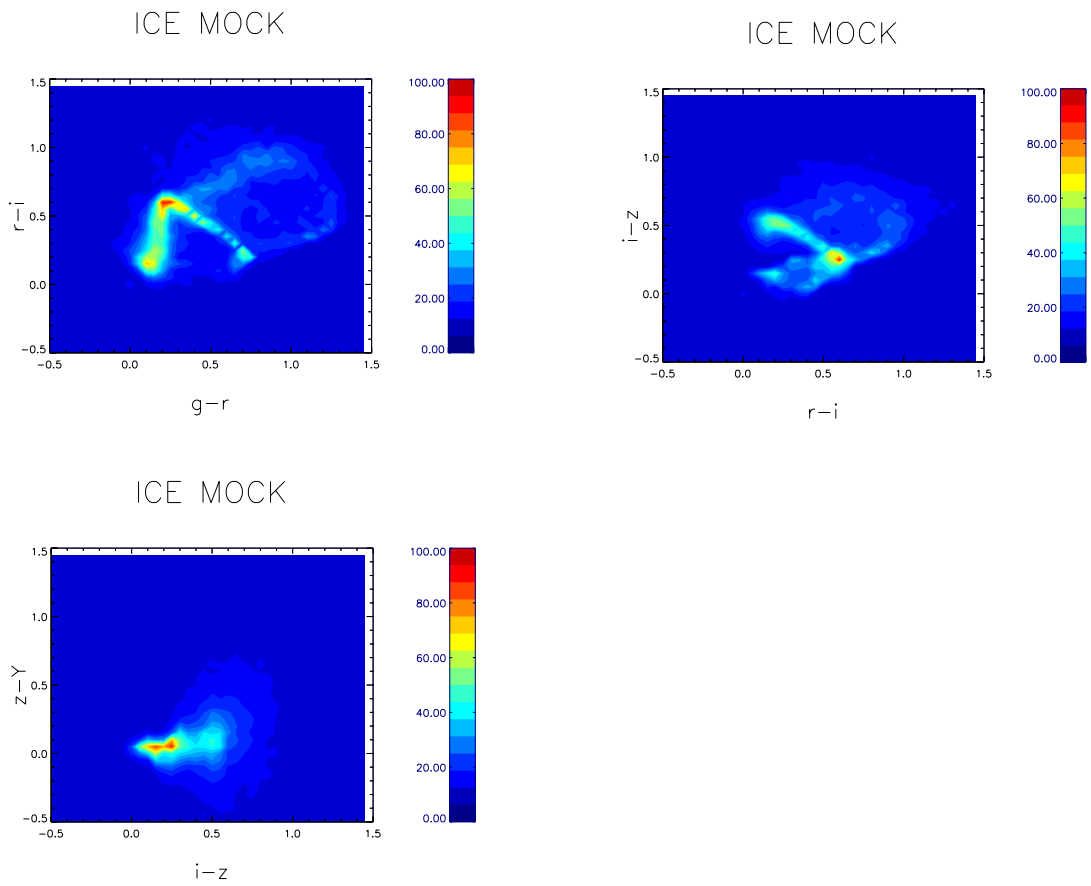


Figure 5.8: *g-r* vs. *r-i* colours (top left panel), *r-i* vs. *i-z* colours (top right panel) and *i-z* vs. *z-Y* colours (bottom panel) for the Barcelona mock galaxy catalogue. The colour bar indicates the number of galaxies.

and complementary to the Stanford mock for use in the Dark Energy Survey working groups.

5.4.3 Galaxy Clusters

In addition to the photometric redshift and colour analysis work, some attempts have been made to detect galaxy clusters in the Stanford mocks.

5.4.3.1 Initial Results

Initial work was carried out within the DES clusters working group, using the Stanford DC4 mock, for the DES Clusters Comparison Workshop held in Columbus, Ohio, in 2008. The objective of this meeting and those since then is to find the optimal cluster finder(s) for the Dark Energy Survey. To this end several groups were encouraged to prepare cluster catalogues using whichever

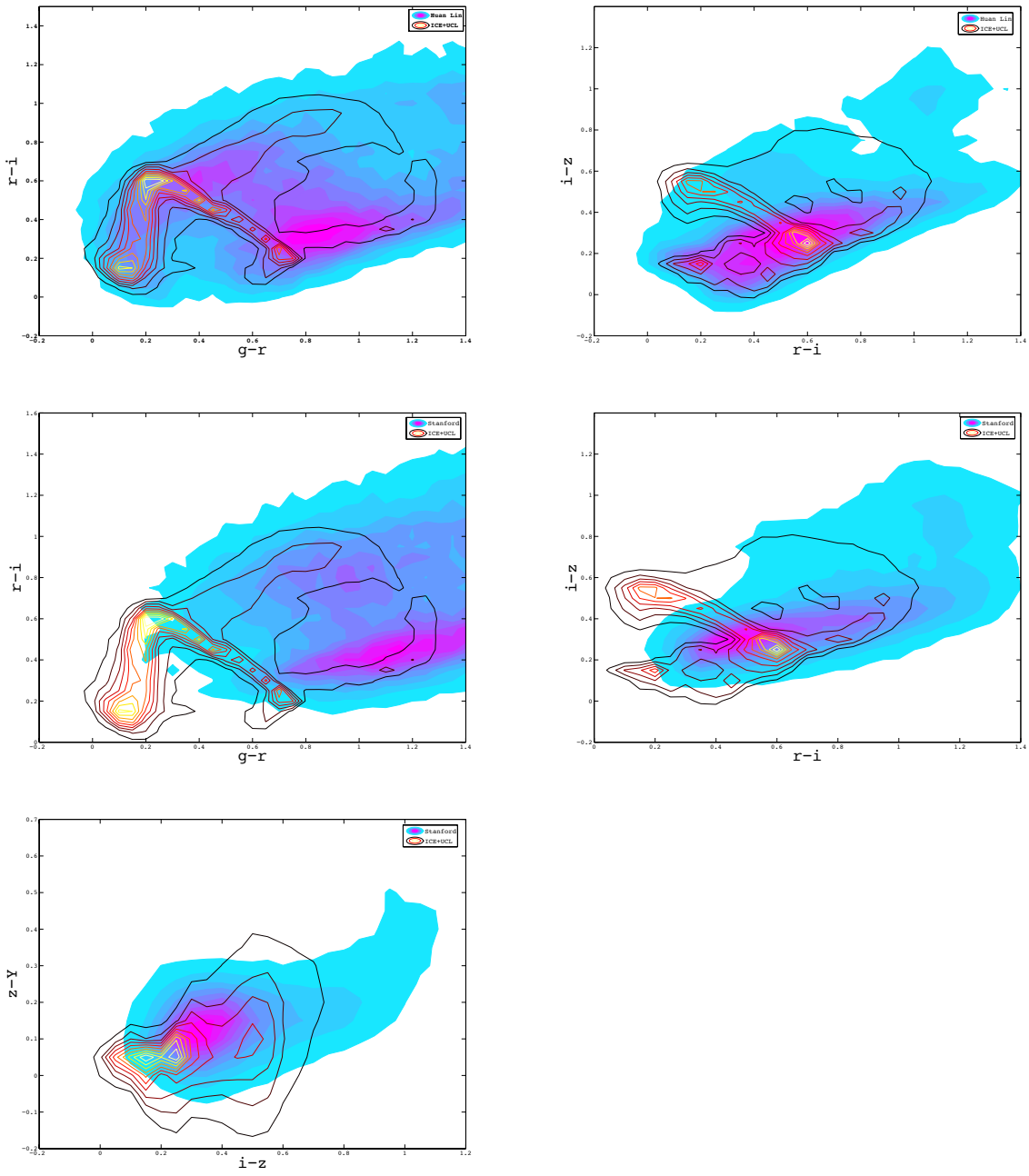


Figure 5.9: Colour space contours for the Barcelona mock (empty contours) and the Lin *et al.* (2004) mock (filled contours) (top panels). Colour space contours for the Barcelona mock (empty contours) and the Stanford DC4 mock (filled contours) (bottom panels).

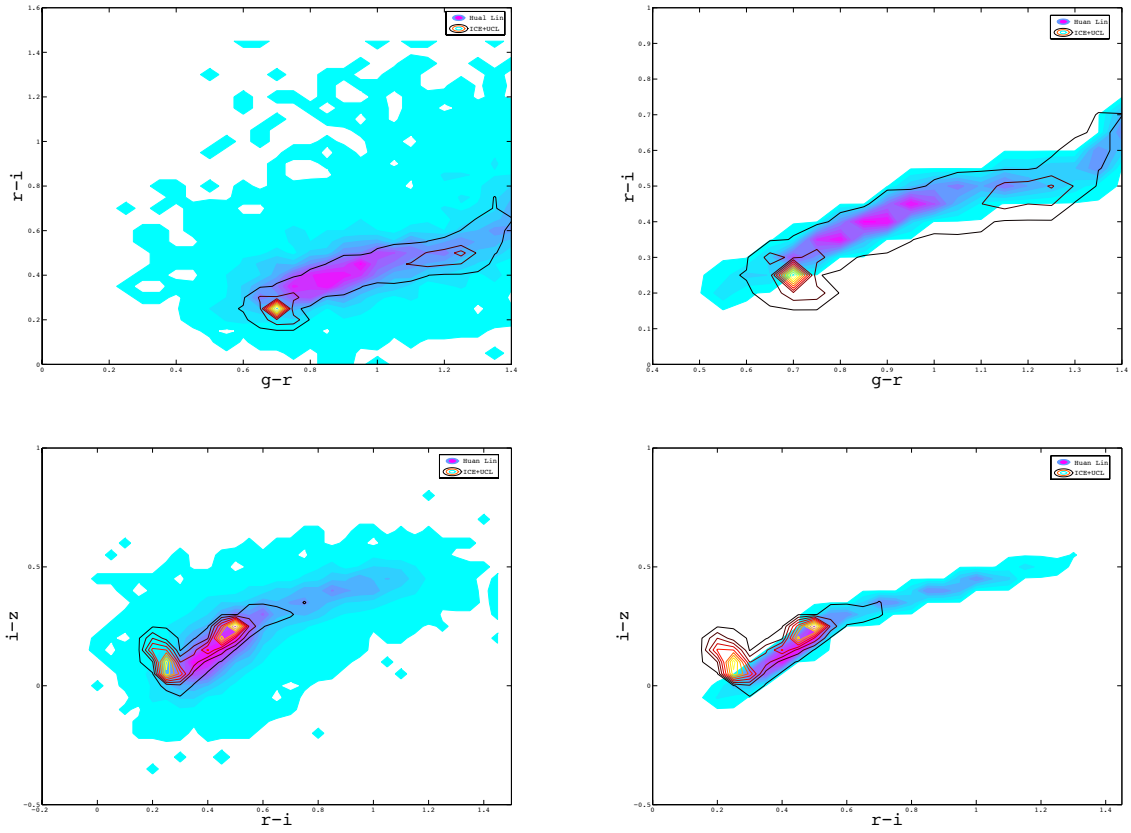


Figure 5.10: Colour space contours for the Barcelona mock (empty contours) and the Lin *et al.* (2004) mock (filled contours) without magnitude errors.

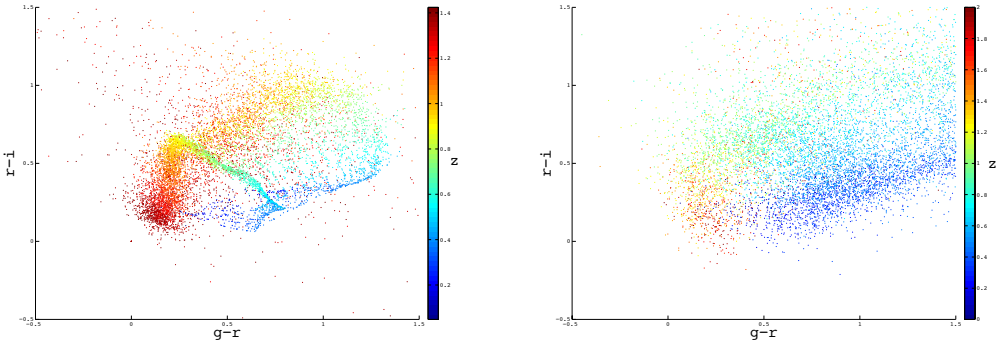


Figure 5.11: Scatter in *g-r* vs. *r-i* colour space as a function of redshift for the Barcelona mock (left panel) and the Lin *et al.* (2004) mock (right panel).

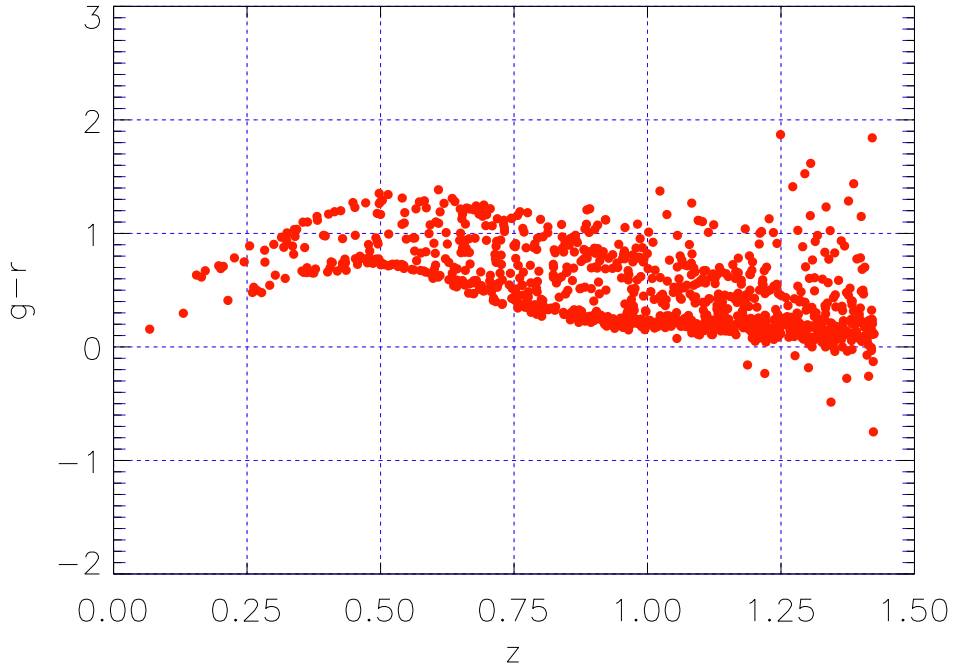


Figure 5.12: *g-r* colour as a function of redshift in Barcelona mock galaxy catalogue.

technique they prefer from the Stanford DC4 mock galaxy catalogue. Each group was provided with the catalogue of the mock galaxies and the positions of the dark matter haloes (or truth table) to determine the reliability (*i.e.* completeness & purity) of their catalogue.

Several cluster catalogues were produced from the Stanford DC4 mock using the friends-of-friends technique described in chapter 3 with various values of R_0. The completeness and contamination (1-purity) for each of these catalogues was determined using a cylindrical matching technique, which simply matches FoF cluster centres to mock halo centres in a cylinder, 1.5 Mpc in radius and $z = 0.004$ in length. This is done requiring that the matching from clusters to haloes be unique. The completeness and contamination of the catalogue were defined via

$$\text{Completeness} = \frac{N_{\text{matches}}}{N_{\text{haloes}}} \tag{5.11}$$

and

$$\text{Contamination} = \frac{(N_{\text{clusters}} - N_{\text{matches}})}{N_{\text{clusters}}}, \tag{5.12}$$

where N_{matches} is the number of matches between FoF cluster and mock haloes, N_{clusters} is the number of FoF clusters detected and N_{haloes} is the number of mock haloes.

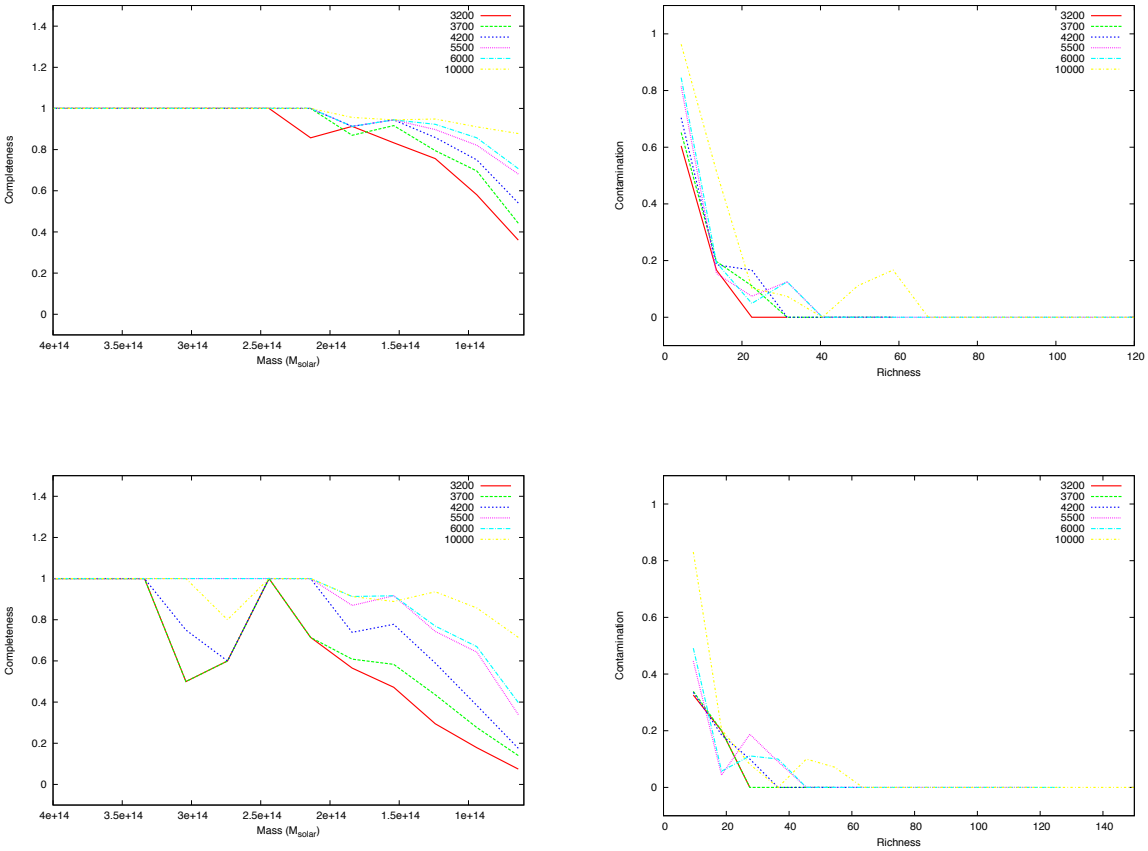


Figure 5.13: Completeness as a function of halo mass for all FoF clusters (top left panel). Contamination as a function of cluster richness for all FoF clusters (top right panel). Completeness as a function of halo mass for FoF clusters with $N_{mem} \geq 5$ (bottom left panel). Contamination as a function of cluster richness for FoF clusters $N_{mem} \geq 5$ (bottom right panel).

Fig.5.13 shows the completeness as a function of halo mass for all FoF clusters (top left panel) and the Contamination as a function of cluster richness for all FoF clusters (top right panel). The bottom panels of fig.5.13 show the completeness and contamination only for clusters with 5 or more members. The figure shows that majority of the contamination comes from the smallest groups of 3 or 4 members and that a reliable catalogue can be produced for haloes down to a mass of $2.5 \times 10^{14} M_{\odot}$. It should be noted that the FoF algorithm has been much improved since these results were presented, however the results presented here will be overly optimistic as the DC4 mock incorporates rather unrealistic photometric redshifts (see §5.4.1). These photometric redshifts approximate very closely the spectroscopic redshifts in the mock catalogue and therefore avoid many of the problems with merging encountered in chapter 3, which severely limits the levels of

completeness and purity that can be achieved without making richness cuts to the cluster catalogue.

The principal difficulty encountered in this meeting was that comparing different cluster catalogues is a non-trivial process. Each group presented a catalogue that was produced in a unique way, some of which contained information about cluster galaxy members, other that contained only information about the cluster centres. In addition to this, each group had presented the reliability of their cluster finder in slightly different way, making comparison very difficult.

5.4.3.2 Cluster Comparison

In order to gain some level of comparison between different cluster finding algorithms in the DES clusters working group, cluster catalogues were analysed that were obtained with some more recent versions of the Stanford DES mocks. These mocks contain more realistic photometric redshifts than previous versions among other minor improvements. The FoF code discussed throughout this thesis is compared, using a very crude cylindrical matching code (described in the previous subsection), with the Gaussian Mixture Brightest Cluster Galaxy (GMBCG) method of Hao *et al.* (2010) and the Voronoi Tessellation (VT) method of Soares-Santos *et al.* (2011).

The GMBCG method makes use of the assumption that the distribution of galaxy colours in a cluster can be well approximated by a mixture of two Gaussian distributions (Hao *et al.* 2009, 2010). The galaxy colours in a patch of sky that does not contain a cluster should be well represented by single wide Gaussian. If a cluster is present, the colours should show a bimodal distribution with one narrow Gaussian that includes galaxies in the cluster red sequence and one wide Gaussian that corresponds to background, foreground and other blue galaxies. A cluster catalogue found using v2.10 of the Stanford DES mock was provided by Jiangang Hao for the purposes of this comparison.

The Voronoi tessellation (see §1.4.5.2) method of Soares-Santos *et al.* (2011) is performed in 2+1 dimensions by dividing the input catalogue into non-overlapping slices of $1 - \sigma_z$, where σ_z is the redshift error. The tessellation is produced by dividing the galaxies in each slice into discrete cells. The size of the cells gives an indication of the density in the field. Cluster candidates are then chosen as those with a density above some threshold. A cluster catalogue found using v1.07 of the Stanford DES mock was provided by Marcelle Soares-Santos for the purposes of this comparison.

The FoF code was run on v2.14 of the Stanford DES mock, which has no significant differences to v2.10, with parameters $R_{\text{friend}}(0.5) = 0.077\ h^{-1}$Mpc and $K = 0.5$. These parameters were chosen solely based on the physical properties of the resulting groups and clusters.

Fig.5.14 shows the completeness as a function of mock halo mass (left panel) and the purity as

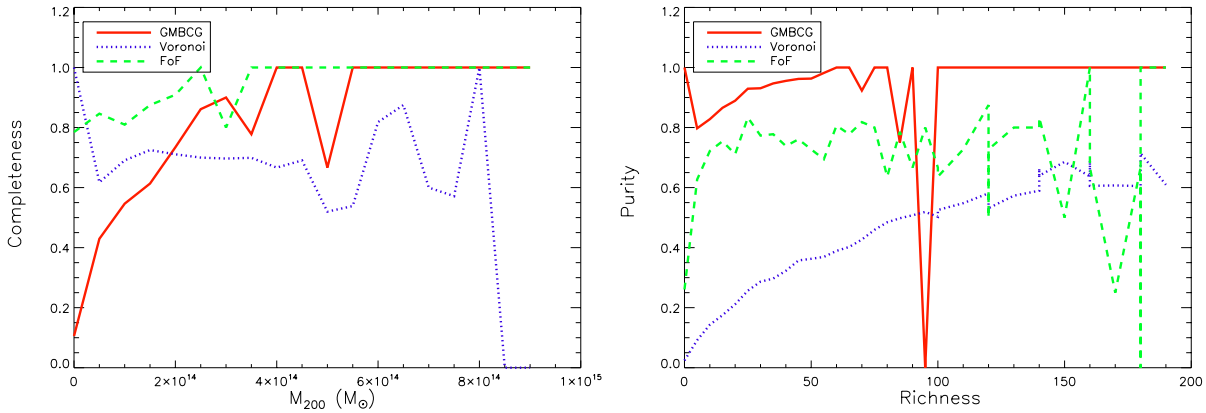


Figure 5.14: Completeness as a function of halo mass, M_{200}, for clusters detected using GMBCG (red solid line), Voronoi Tessellation (blue dotted line) and FoF (green dashed line). Purity as a function of cluster richness for clusters detected using GMBCG (red solid line), VT (blue dotted line) and FoF (green dashed line).

function of cluster richness (right panel) for GMBCG (red solid line), VT (blue dotted line) and FoF (green dashed line). This plot shows that the FoF does marginally better than GMBCG for the completeness, particularly for lower mass haloes, but generally worse for the purity. The VT method appears to perform the worst in both accounts. It should be clearly noted that the cylindrical matching technique is very simplistic and probably does not give the best account of the reliability of a cluster finder, but for the purposes of this chapter provides a consistent method to compare the three codes. It should also be noted that the FoF code was not fully optimised (as in chapters 2 and 3) for this comparison.

These results indicate that the FoF and GMBCG techniques are comparable in their ability to reproduce the mock halo catalogue. Part of the discrepancy with the VT technique can be explained by the fact that the catalogue used for this comparison only has a proxy for the cluster richness. A more accurate measure for the cluster richness can be obtained using a follow-up code. Soares-Santos *et al.* (2011) show much better values of completeness and purity obtained using a membership matching technique. In general membership matching techniques are a more reliable method for determining the efficiency of a cluster finder. This does, however, require that each cluster catalogue provides a list of the member galaxies that belong to each cluster or that some kind of afterburner be run to determine the membership of each cluster. Membership matching could not be implemented for the purposes of this chapter as cluster galaxy information was not available for all of the cluster catalogues. A more in-depth membership matching analysis will be done before the contents of this chapter are submitted for publication, this will be part of the DES

cluster comparison tests.

5.5 Conclusions and Future Work

The Dark Energy Survey is revolutionary new optical-near infrared survey of 5000 deg^2 of the southern galactic cap. DES will use four different methods to probe the properties of dark energy; baryon oscillations of the power spectrum, the abundance and spatial distribution of clusters, weak gravitational lensing and type Ia supernovae. The survey will carried out using the 4 m Blanco telescope based at the Cerro Tololo Inter-American Observatory (CTIO) with the 519 mexapixel optical CCD camera, DECam. To prepare for the abundance of data that will be made available, simulations of the data expected from DES have been prepared by several institutions.

Work has been done to obtain photometric redshifts for the simulated galaxies using the neural network code ANNz. Initial results showed that the scatter between spectroscopic and photometric redshifts was less than would be expected from real data. A comparison was then made between the different simulated galaxy colours in an attempt to explain this discrepancy. Furthermore, the friends-of-friend code described in chapter 3 was used to detect structure in the Stanford simulation.

All of the work described in this chapter is still ongoing. Since these results were presented, several improvements have been made to the DES simulations both at Stanford (now on DC5) and in Barcelona, and also to the friends-of-friends code. Work will continue to improve the galaxy colours in the simulations, which will in turn lead to more realistic photometric redshifts. Once the galaxies have reliable photometric redshifts, the FoF can be used to detect structure in the simulations.

CHAPTER 6

DISCUSSION AND CONCLUSIONS

6.1 Introduction

This thesis has presented a series of projects focused on the optical detection of clusters of galaxies. The principal aim has been to investigate techniques that can be applied to current and up coming galaxy surveys such as the Dark Energy Survey (DES), with particular interest in using estimated photometric redshifts to detect genuine structures. The following section (§6.2) summarises the results and conclusions from each chapter. The last section (§6.3) discusses some work that will be done in the future to further the work presented in this thesis.

6.2 Summary and Conclusions

6.2.1 Optical Cluster Detection I: Spectroscopic Friends-of-Friends

This chapter presents a catalogue of galaxy groups and clusters selected using a friends-of-friends algorithm, which is based on that of Huchra & Geller (1982) with a dynamic linking length, from the 2dF-SDSS and QSO (2SLAQ) luminous red galaxy survey. The linking parameters for the code are chosen through an analysis of simulated 2SLAQ haloes, produced from the Horizon 4π simulation and the HOD prescription of Blake *et al.* (2008). The resulting catalogue includes 313 clusters containing 1,152 galaxies. The galaxy groups and clusters have an average velocity dispersion of $\overline{\sigma}_v = 467.97$ kms^{-1} and an average size of $\overline{R}_{\rm clt} = 0.78\ h^{-1}$Mpc.

Galaxies from regions of one square degree and centred on the galaxy clusters were downloaded from the Sloan Digital Sky Survey Data Release 6 (SDSS DR6). Investigating the photometric

redshifts and cluster red-sequence of these galaxies shows that the galaxy clusters detected with the FoF algorithm are reliable out to z$\sim$0.6. The discrepancies beyond this range is partially owing to the small number of groups and clusters detected at higher redshifts. Also, there may be some contamination effects in the histograms from the photometric redshift errors in the SDSS data.

Mass estimates are made for the clusters using their velocity dispersions. These mass estimates are shown to be reliable by comparing them with the 2SLAQ mock halo masses. Some deviations are seen, which are the result of contaminating galaxies in the richness estimates and the small number of high mass clusters detected. In general the range of cluster masses is a good match to known masses of massive clusters.

Further analysis of the simulation haloes shows that clipping out low richness groups with large radii improves the purity of catalogue from 52$\%$ to 88$\%$, while retaining a completeness of 94$\%$. It is unknown to what accuracy this clipping procedure can be applied to the real 2SLAQ data, therefore the 2SLAQ clusters that survive the clipping are labelled as 'gold' sample objects, meaning that they have a higher probability of being genuine structures. The clusters that fail the clipping are labelled as 'silver' sample objects and may still be genuine, but with a lower probability. Some of these lower probability objects may be clusters that are still collapsing. Out of the 313 total 2SLAQ groups and clusters, 70 are gold and the remaining 243 are silver.

Finally, the two-point correlation function of the cluster catalogue is tested. The best-fitting power law model , $\xi(r) = (r/r_0)^{\gamma}$, with parameters $r_0 = 24 \pm 4$ Mpc h^{-1} and $\gamma = -2.1 \pm 0.2$ is found.

6.2.2 Optical Cluster Detection II: Photometric Friends-of-Friends

A slightly modified version of the Botzler *et al.* (2004) photometric redshift friends of friends algorithm, which used a dynamic linking length that changes with the surface number density of galaxies is presented. The code is applied to 2dF-SDSS and QSO (2SLAQ) luminous red galaxy survey using photometric redshifts obtained using the neural network code ANNz (Collister & Lahav 2004). A cluster catalogue is produced with 569 groups and clusters containing 2,180 member galaxies that matches the spectroscopic catalogue obtained in chapter 2 to 38% without any cuts and up to 80% if matching is only considered for 2SLAQ spectroscopic clusters with $N_{\rm mem} \geq 6$.

The code is also run on the Mega$z$-LRG DR7 catalogue with linking parameters chosen based on the physical properties of the clusters found and the completeness of the Mega$z$-LRG clusters found within the 2SLAQ area to the 2SLAQ spectroscopic clusters. Two catalogues are chosen for

further analysis as no obvious agreement can be found between these criteria. The first catalogue includes 21,255 groups and clusters (containing 73073 galaxies) that have physical properties that would be expected of clusters. The second catalogue includes 114,874 groups and clusters (containing 544,008 galaxies). This catalogue has a higher completeness relative to 2SLAQ. This difference occurs because as the linking parameters are increased, a large number of smaller structures are found contributing to the completeness and at the same time larger structures are merged together to form extremely large unphysical structures.

Galaxies from regions of one square degree and centred on the galaxy clusters were downloaded from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7). Investigating the photometric redshifts and cluster red-sequence of these galaxies shows that the galaxy clusters detected with the FoF algorithm are reliable out to $z \sim 0.6$ for the 2SLAQ photometric catalogue, consistent with the spectroscopic results, and out to $z \sim 0.65$ for the Mega$z$-LRG catalogues. These results appear to favour the more complete Mega$z$-LRG cluster catalogue, which shows a trend of ~ 0.3 in $g - i$ colour space consistent with the cluster red sequence. On dense galaxy catalogues, this cluster finding method in not capable of producing a catalogue that is both pure and complete. The code can, however, produce a cluster catalogue that is either pure or complete independently.

The radial profiles of all three cluster catalogues are examined, also indicating a better agreement between 2SLAQ and the more complete Mega$z$-LRG catalogue.

Finally, the two-point correlation function of the 2SLAQ photometric redshift cluster catalogue is tested. The best-fitting power law model , $\xi(r) = (r/r_0)^{\gamma}$, with parameters $r_0 = 9.55 \pm 4.23$ Mpc h^{-1} and $\gamma = -1.115 \pm 0.29$ is found. These results show less correlation that those found for the spectroscopic 2SLAQ cluster catalogue, which is the result of impurities diluting the clustering length. This implies that the two-point correlation function can be used to measure the impurity of the photometric 2SLAQ cluster catalogue.

6.2.3 AAT Spectroscopic Observations of a Selection of Photometric FoF Targets

Eight fields were targeted for observation based on cluster candidate positions found using a friends-of-friends (FoF) algorithm that incorporates photometric redshifts on the Mega$z$-LRG DR5 catalogue. The best candidates were chosen as those with seven or more LRGs and were labeled sample A clusters. The fields were supplemented with SDSS galaxies lying along the red-sequence of these clusters. 4116 spectra were obtained over four grey nights at the Anglo-Australian Observatory using AAOmega from the eight observing fields. From these 4116 spectra, 2379 galaxy redshifts were measured using the RUNZ package. Measuring $N(z)$ in regions 1

h^{-1}Mpc around a selection of the sample A clusters shows clear peaks in z around the cluster redshifts. Estimates are made for the cluster velocity dispersions and dynamical masses by taking only the galaxies within the $N(z)$ peak. The majority of these estimates are consistent with what would be expected for genuine structures, although these estimates are very approximate.

These results can potentially be improved by running a newer version of the FoF code on the Mega$z$-LRG catalogue. This may improve the accuracy of detecting the cluster centres. This code could also be run on a more recent release of Mega$z$-LRG. A more in depth spectroscopic analysis of this data could provide reliable velocity dispersion and dynamical mass estimates, which could be compared with lensing and/or X-ray data.

6.2.4 The Dark Energy Survey

The Dark Energy Survey is revolutionary new optical-near infrared survey of 5000 deg^2 of the southern galactic cap. DES will use four different methods to probe the properties of dark energy; baryon oscillations of the power spectrum, the abundance and spatial distribution of clusters, weak gravitational lensing and type Ia supernovae. The survey will carried out using the 4 m Blanco telescope based at the Cerro Tololo Inter-American Observatory (CTIO) with the 519 mexapixel optical CCD camera, DECam. To prepare for the abundance of data that will be made available, simulations of the data expected from DES have been prepared by several institutions.

Work has been done to obtain photometric redshifts for the simulated galaxies using the neural network code ANNz. Initial results have shown that the scatter between spectroscopic and photometric redshifts was less than would be expected from real data. This has highlighted problems with the colours in early versions of the Stanford DES mock, which have since be corrected. A comparison was then made between the different simulated galaxy colours in an attempt to explain this discrepancy. Furthermore, the friends-of-friend code described in chapter 3 was used to detect structure in the Stanford simulation and these results have been compared with other cluster finding techniques. This comparison process is still on going, but initial results indicate that this method is consistent with other methods currently used in DES.

All of the work described in this chapter is still ongoing. Since these results were presented, several improvements have been made to the DES simulations and the friends-of-friends code. Work will continue to improve the galaxy colours in the simulations, which will in turn lead to more realistic photometric redshifts. Once the galaxies have reliable photometric redshifts, the FoF code can be used to detect structure in the simulations.

6.3 Future Work

6.3.1 Improved Mocks

One of the problems encountered when working with the photometric friends-of-friends method was that a crude mock was unable to provide reliable estimates of the completeness and purity of a given set of linking parameters. The analysis done in §3.5 showed that taking a spectroscopic mock and adding photometric redshifts using the bias between 2SLAQ spectroscopic and photometric redshifts changed the number density of galaxies in certain ranges. To compensate for this effect it is necessary to remove a certain fraction of objects, however the way in which this is carried out has significant effects on the resulting levels of completeness and purity that can be obtained. Therefore, it is clear that a more sophisticated mock would be required to gain some understanding of the level of completeness of a given cluster catalogue. This mock would ideally have realistic colours assigned to each of the galaxies, which would allow photometric redshifts to be estimated for each of the objects.

The work mentioned in chapter 5 regarding the Barcelona DES mocks is the beginning of a project that should shed some light on the advantages and disadvantages of using simulated galaxy catalogues to explore the completeness and purity of photometric friends-of-friends cluster catalogues. It may turn out that using mock catalogues is not the most efficient way of producing a realistic cluster catalogue. Perhaps, as shown in chapter 4, it is better to follow up with a spectroscopic analysis on a selection of objects and identify the fraction of objects that constitute genuine bound structures. Most likely the most reliable catalogue will be produced by combining both approaches.

6.3.2 Voronoi Tessellation

Another major difficulty encountered when working with the photometric friends-of-friends technique is effect of cluster merging. In §3.5 it was shown that the completeness tends to peak $\sim 50\%$ because for low linking parameter combinations only larger structures are detected and if the parameters are increased smaller structures are found, however simultaneously lager structures merge together often forming highly unphysical objects. The results in chapter 3 indicate that it is better to use a higher parameter set that corresponds to higher completeness as it may be possible to reanalyse the more unphysical objects to look for substructure. It may, however, be possible to avoid the problem by removing some of the field galaxies that cause the clusters to merge when the linking parameters are increased.

One way of removing field galaxies from a catalogue before running the friend-of-friends would be to implement a Voronoi tessellation code (see 1.4.5.2). This technique can be used to identify voids in a galaxy catalogue and remove galaxies living in these regions before searching for the galaxy clusters. A three-dimensional Voronoi tessellation code has been developed by Walter Dos Santos, which could be implemented in the near future to test this approach.

6.3.3 Cluster Masses

In chapter 2 cluster masses were estimated from the cluster velocity dispersions. The masses obtained appear to be consistent with those expected for galaxy clusters, however there is large uncertainty associated with these estimates because of the small number of LRGs contributing to the velocity dispersions. It would be interesting to carry out a spectroscopic follow up on a selection of these clusters to examine all the member galaxies (not just the LRGs) to determine how reliable the velocity dispersions given solely by the LRGs are. Additionally, it would be interesting to compare the mass estimates with those provided by other methods such as gravitational lensing or X-rays.

6.3.4 DES Clusters

Continued involvement in the Dark Energy Survey is essential for keeping up to date with work regarding galaxy cluster detection. The Stanford simulations are useful for testing the photometric friends-of-friends algorithm. In addition, the DES clusters working group provides a platform to compare with other cluster finding techniques.

One task aimed for the near future is to incorporate the friends-of-friends code into the DES portal.

Bibliography

Abell, G. O., 1958, *ApJS*, **3**, 211

Abell, G. O., Corwin, Jr., H. G. & Olowin, R. P., 1989, *ApJS*, **70**, 1

Adams, W. S., 1908, *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, **22**, 1

Albrecht, A., Bernstein, G., Cahn, R., Freedman, W. L., Hewitt, J., Hu, W., Huth, J., Kamionkowski, M., Kolb, E. W., Knox, L., Mather, J. C., Staggs, S. & Suntzeff, N. B., 2006, *ArXiv Astrophysics e-prints*

Alcock, C., Akerlof, C. W., Allsman, R. A., Axelrod, T. S., Bennett, D. P., Chan, S., Cook, K. H., Freeman, K. C., Griest, K., Marshall, S. L., Park, H., Perlmutter, S., Peterson, B. A., Pratt, M. R., Quinn, P. J., Rodgers, A. W., Stubbs, C. W. & Sutherland, W., 1993*a*, *Nature*, **365**, 621

Alcock, C., Allsman, R. A., Axelrod, T. S., Bennett, D. P., Cook, K. H., Park, H. S., Marshall, S. L., Stubbs, C. W., Griest, K., Perlmutter, S., Sutherland, W., Freeman, K. C., Peterson, B. A., Quinn, P. J. & Rodgers, A. W., 1993*b*, in B. T. Soifer (ed.), *Sky Surveys. Protostars to Protogalaxies*, volume 43 of *Astronomical Society of the Pacific Conference Series*, pp. 291–+

Annis, J., Kent, S., Castander, F., Eisenstein, D., Gunn, J., Kim, R., Lupton, R., Nichol, R., Postman, M., Voges, W. & SDSS Collaboration, 1999, in *Bulletin of the American Astronomical Society*, volume 31 of *Bulletin of the American Astronomical Society*, pp. 1391–+

Bahcall, N. A. & Soneira, R. M., 1983, *ApJ*, **270**, 20

Banerji, M., Abdalla, F. B., Lahav, O. & Lin, H., 2008, *MNRAS*, **386**, 1219

Bartelmann, M. & Narayan, R., 1995, in S. S. Holt & C. L. Bennett (ed.), *Dark Matter*, volume 336 of *American Institute of Physics Conference Series*, pp. 307–319

Bartelmann, M. & Schneider, P., 2001, *Physics Reports*, **340**, 291

Battye, R. A. & Weller, J., 2003, *Physical Review D*, **68**(8), 083506

Baum, W. A., 1962, in G. C. McVittie (ed.), *Problems of Extra-Galactic Research*, volume 15 of *IAU Symposium*, pp. 390–+

Bennett, C. L., Banday, A. J., Gorski, K. M., Hinshaw, G., Jackson, P., Keegstra, P., Kogut, A., Smoot, G. F., Wilkinson, D. T. & Wright, E. L., 1996, *ApJL*, **464**, L1+

Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Spergel, D. N., Tucker, G. S., Wollack, E., Wright, E. L., Barnes, C., Greason, M. R., Hill, R. S., Komatsu, E., Nolta, M. R., Odegard, N., Peiris, H. V., Verde, L. & Weiland, J. L., 2003, *ApJS*, **148**, 1

Bertin, E. & Arnouts, S., 1996, *A&AS*, **117**, 393

Blake, C., Collister, A., Bridle, S. & Lahav, O., 2007, *MNRAS*, **374**, 1527

Blake, C., Collister, A. & Lahav, O., 2008, *MNRAS*, **385**, 1257

Blanton, M. R., Hogg, D. W., Bahcall, N. A., Brinkmann, J., Britton, M., Connolly, A. J., Csabai, I., Fukugita, M., Loveday, J., Meiksin, A., Munn, J. A., Nichol, R. C., Okamura, S., Quinn, T., Schneider, D. P., Shimasaku, K., Strauss, M. A., Tegmark, M., Vogeley, M. S. & Weinberg, D. H., 2003, *ApJ*, **592**, 819

Blindert, K., Yee, H. K. C., Gladders, M. D. & Ellingson, E., 2004, in A. Diaferio (ed.), *IAU Colloq. 195: Outskirts of Galaxy Clusters: Intense Life in the Suburbs*, pp. 215–219

Boggess, N. W., Mather, J. C., Weiss, R., Bennett, C. L., Cheng, E. S., Dwek, E., Gulkis, S., Hauser, M. G., Janssen, M. A., Kelsall, T., Meyer, S. S., Moseley, S. H., Murdock, T. L., Shafer, R. A., Silverberg, R. F., Smoot, G. F., Wilkinson, D. T. & Wright, E. L., 1992, *ApJ*, **397**, 420

Böhringer, H., 1995, in H. Böhringer, G. E. Morfill, & J. E. Trümper (ed.), *Seventeeth Texas Symposium on Relativistic Astrophysics and Cosmology*, volume 759 of *Annals of the New York Academy of Sciences*, pp. 67–+

Bolzonella, M., Miralles, J. & Pelló, R., 2000, *A&A*, **363**, 476

Borgani, S., 2006, *ArXiv Astrophysics e-prints*

Borgani, S., Plionis, M. & Kolokotronis, V., 1999, *MNRAS*, **305**, 866

Botzler, C. S., Snigula, J., Bender, R. & Hopp, U., 2004, *MNRAS*, **349**, 425

Bradač, M., Lombardi, M. & Schneider, P., 2004, *A&A*, **424**, 13

Bruzual, G. & Charlot, S., 2003, *MNRAS*, **344**, 1000

Bryan, G. L. & Norman, M. L., 1998, *ApJ*, **495**, 80

Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R. & Dekel, A., 2001, *MNRAS*, **321**, 559

Cannon, R., Drinkwater, M., Edge, A., Eisenstein, D., Nichol, R., Outram, P. & Pimbblet, K., 2006, *MNRAS*, **372**, 425

Capak, P., Cowie, L. L., Hu, E. M., Barger, A. J., Dickinson, M., Fernandez, E., Giavalisco, M., Komiyama, Y., Kretchmer, C., McNally, C., Miyazaki, S., Okamura, S. & Stern, D., 2004, *AJ*, **127**, 180

Carlberg, R. G., Yee, H. K. C., Ellingson, E., Abraham, R., Gravel, P., Morris, S. & Pritchet, C. J., 1996, *ApJ*, **462**, 32

Carroll, S. M., Duvvuri, V., Trodden, M. & Turner, M. S., 2004, *Physical Review D*, **70**(4), 043528

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

Colberg, J. M., White, S. D. M., Yoshida, N., MacFarland, T. J., Jenkins, A., Frenk, C. S., Pearce, F. R., Evrard, A. E., Couchman, H. M. P., Efstathiou, G., Peacock, J. A., Thomas, P. A. & The Virgo Consortium, 2000, *MNRAS*, **319**, 209

Coleman, G. D., Wu, C. & Weedman, D. W., 1980, *ApJS*, **43**, 393

Colín, P., Klypin, A. A., Kravtsov, A. V. & Khokhlov, A. M., 1999, *ApJ*, **523**, 32

Colless, M., 1989, *MNRAS*, **237**, 799

Colless, M., 1999, *Royal Society of London Philosophical Transactions Series A*, **357**, 105

Collins, C. A., Guzzo, L., Böhringer, H., Schuecker, P., Chincarini, G., Cruddace, R., De Grandi, S., MacGillivray, H. T., Neumann, D. M., Schindler, S., Shaver, P. & Voges, W., 2000, *MNRAS*, **319**, 939

Collister, A., Lahav, O., Blake, C., Cannon, R., Croom, S., Drinkwater, M., Edge, A., Eisenstein, D., Loveday, J., Nichol, R., Pimbblet, K., de Propris, R., Roseboom, I., Ross, N., Schneider, D. P., Shanks, T. & Wake, D., 2007, *MNRAS*, **375**, 68

Collister, A. A. & Lahav, O., 2004, *PASP*, **116**, 345

Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G. & Munn, J. A., 1995, *AJ*, **110**, 2655

Cooray, A. R., 1999, *MNRAS*, **307**, 841

Couch, W. J., Ellis, R. S., MacLaren, I. & Malin, D. F., 1991, *MNRAS*, **249**, 606

Cowie, L. L., Barger, A. J., Hu, E. M., Capak, P. & Songaila, A., 2004, *AJ*, **127**, 3137

Cypriano, E. S., Sodré, L. J., Kneib, J.-P. & Campusano, L. E., 2004, *ApJ*, **613**, 95

Dalton, G. B., Maddox, S. J., Sutherland, W. J. & Efstahiou, G., 1997, *VizieR Online Data Catalog*, **728**, 90263

Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M., 1985, *ApJ*, **292**, 371

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

Diaferio, A., Borgani, S., Moscardini, L., Murante, G., Dolag, K., Springel, V., Tormen, G., Tornatore, L. & Tozzi, P., 2005, *MNRAS*, **356**, 1477

Dodelson, S., 2003, *Modern cosmology*

Driver, S. P., Baldry, I. K., Bamford, S., Bland-Hawthorn, J., Bridges, T., Cameron, E., Conselice, C., J., C. W., Croom, S., Cross, N. J. G., Driver, S. P., Dunne, L., Eales, S., Edmondson, E., Ellis, S. C., Frenk, C. S., Graham, A. W. & Jones, H., 2008, *ArXiv e-prints*

Dyson, F. W., Eddington, A. S. & Davidson, C., 1920, *Royal Society of London Philosophical Transactions Series A*, **220**, 291

Efstathiou, G., 1990, in J. A. Peacock, A. F. Heavens, & A. T. Davies (ed.), *Physics of the Early Universe*, pp. 361–463

Efstathiou, G., Frenk, C. S., White, S. D. M. & Davis, M., 1988, *MNRAS*, **235**, 715

Einstein, A., 1905, *Annalen der Physik*, **323**, 639

Einstein, A., 1915*a*, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*, 844

Einstein, A., 1915*b*, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 778-786.*, 778

Einstein, A., 1917, *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 142-152.*, 142

Eisenstein, D. J., Hogg, D. W., Fukugita, M., Nakamura, O., Bernardi, M., Finkbeiner, D. P., Schlegel, D. J., Brinkmann, J., Connolly, A. J., Csabai, I., Gunn, J. E., Ivezić, Ž., Lamb, D. Q., Loveday, J., Munn, J. A., Nichol, R. C., Schneider, D. P., Strauss, M. A., Szalay, A. & York, D. G., 2003, *ApJ*, **585**, 694

Eisenstein, D. J. & Hut, P., 1998, *ApJ*, **498**, 137

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J., Connolly, A., Csabai, I., Doi, M., Fukugita, M., Frieman, J. A., Glazebrook, K., Gunn, J. E., Hendry, J. S., Hennessy, G., Ivezić, Z., Kent, S., Knapp, G. R., Lin, H., Loh, Y., Lupton, R. H., Margon, B., McKay, T. A., Meiksin, A., Munn, J. A., Pope, A., Richmond, M. W., Schlegel, D., Schneider, D. P., Shimasaku, K., Stoughton, C., Strauss, M. A., SubbaRao, M., Szalay, A. S., Szapudi, I., Tucker, D. L., Yanny, B. & York, D. G., 2005, *ApJ*, **633**, 560

Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., Colberg, J. M., Yoshida, N., White, S. D. M., Jenkins, A., Frenk, C. S., Pearce, F. R., Peacock, J. A. & Thomas, P. A., 2002, *ApJ*, **573**, 7

Falco, E. E., Gorenstein, M. V. & Shapiro, I. I., 1985, *ApJL*, **289**, L1

Fedeli, C., Carbone, C., Moscardini, L. & Cimatti, A., 2011, *MNRAS*, 498

Feldmann, R., Carollo, C. M., Porciani, C., Lilly, S. J., Capak, P., Taniguchi, Y., Le Fèvre, O., Renzini, A., Scoville, N., Ajiki, M., Aussel, H., Contini, T., McCracken, H., Mobasher, B., Murayama, T., Sanders, D., Sasaki, S., Scarlata, C., Scodeggio, M., Shioya, Y., Silverman, J., Takahashi, M., Thompson, D. & Zamorani, G., 2006, *MNRAS*, **372**, 565

Freedman, W. L., Madore, B. F., Gibson, B. K., Ferrarese, L., Kelson, D. D., Sakai, S., Mould, J. R., Kennicutt, Jr., R. C., Ford, H. C., Graham, J. A., Huchra, J. P., Hughes, S. M. G., Illingworth, G. D., Macri, L. M. & Stetson, P. B., 2001, *ApJ*, **553**, 47

Freese, K., 1986, *Physics Letters B*, **167**, 295

Friedmann, A., 1922, *Zeitschrift fr Physik A Hadrons and Nuclei*, **10**, 377

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K. & Schneider, D. P., 1996, *AJ*, **111**, 1748

Gal, R. R., 2006, *ArXiv Astrophysics e-prints*

Geller, M. J. & Huchra, J. P., 1983, *ApJS*, **52**, 61

Gladders, M. D. & Yee, H. K. C., 2000, *AJ*, **120**, 2148

Gladders, M. D. & Yee, H. K. C., 2005, *ApJS*, **157**, 1

Gladders, M. D., Yee, H. K. C., Majumdar, S., Barrientos, L. F., Hoekstra, H., Hall, P. B. & Infante, L., 2007, *ApJ*, **655**, 128

Glazebrook, K., Blake, C., Couch, W., Forbes, D., Drinkwater, M., Jurek, R., Pimbblet, K., Madore, B., Martin, C., Small, T., Forster, K., Colless, M., Sharp, R., Croom, S., Woods, D., Pracy, M., Gilbank, D., Yee, H. & Gladders, M., 2007, *ArXiv Astrophysics e-prints*

Goto, T., Sekiguchi, M., Nichol, R. C., Bahcall, N. A., Kim, R. S. J., Annis, J., Ivezić, Ž., Brinkmann, J., Hennessy, G. S., Szokoly, G. P. & Tucker, D. L., 2002, *AJ*, **123**, 1807

Guth, A. H., 1981, *Physical Review D (Particles and Fields)*, **23**, 347

Gwyn, S. D. J. & Hartwick, F. D. A., 1996, *ApJL*, **468**, L77+

Haiman, Z., Mohr, J. J. & Holder, G. P., 2001, *ApJ*, **553**, 545

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

Hansen, S. M., McKay, T. A., Wechsler, R. H., Annis, J., Sheldon, E. S. & Kimball, A., 2005, *ApJ*, **633**, 122

Hansen, S. M., Sheldon, E. S., Wechsler, R. H. & Koester, B. P., 2009, *ApJ*, **699**, 1333

Hao, J., Koester, B. P., Mckay, T. A., Rykoff, E. S., Rozo, E., Evrard, A., Annis, J., Becker, M., Busha, M., Gerdes, D., Johnston, D. E., Sheldon, E. & Wechsler, R. H., 2009, *ApJ*, **702**, 745

Hao, J., McKay, T. A., Koester, B. P., Rykoff, E. S., Rozo, E., Annis, J., Wechsler, R. H., Evrard, A., Siegel, S. R., Becker, M., Busha, M., Gerdes, D., Johnston, D. E. & Sheldon, E., 2010, *ApJS*, **191**, 254

Heath, D. J., 1977, *MNRAS*, **179**, 351

Heisler, J., Tremaine, S. & Bahcall, J. N., 1985, *ApJ*, **298**, 8

Hinshaw, G., Weiland, J. L., Hill, R. S., Odegard, N., Larson, D., Bennett, C. L., Dunkley, J., Gold, B., Greason, M. R., Jarosik, N., Komatsu, E., Nolta, M. R., Page, L., Spergel, D. N., Wollack, E., Halpern, M., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S. & Wright, E. L., 2009, *ApJS*, **180**, 225

Hogg, D. W., 1999, *ArXiv Astrophysics e-prints*

Holmberg, E., 1941, *ApJ*, **94**, 385

Hu, W. & Dodelson, S., 2002, *ArA&A*, **40**, 171

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

Huchra, J., Davis, M., Latham, D. & Tonry, J., 1983, *ApJS*, **52**, 89

Huchra, J. P. & Geller, M. J., 1982, *ApJ*, **257**, 423

Huggins, W., 1868, *Royal Society of London Philosophical Transactions Series I*, **158**, 529

Ilbert, O., Arnouts, S., McCracken, H. J., Bolzonella, M., Bertin, E., Le Fèvre, O., Mellier, Y., Zamorani, G., Pellò, R., Iovino, A., Tresse, L., Le Brun, V., Bottini, D., Garilli, B., Maccagni, D., Picat, J. P., Scaramella, R., Scodeggio, M., Vettolani, G., Zanichelli, A., Adami, C., Bardelli, S., Cappi, A., Charlot, S., Ciliegi, P., Contini, T., Cucciati, O., Foucaud, S., Franzetti, P., Gavignaud, I., Guzzo, L., Marano, B., Marinoni, C., Mazure, A., Meneux, B., Merighi, R., Paltani, S., Pollo, A., Pozzetti, L., Radovich, M., Zucca, E., Bondi, M., Bongiorno, A., Busarello, G., de La Torre, S., Gregorini, L., Lamareille, F., Mathez, G., Merluzzi, P., Ripepi, V., Rizzo, D. & Vergani, D., 2006, *A&A*, **457**, 841

Jackson, J. C., 1972, *MNRAS*, **156**, 1P

Jeans, J. H., 1902, *Royal Society of London Philosophical Transactions Series A*, **199**, 1

Jones, C. & Forman, W., 1984, *ApJ*, **276**, 38

Jones, D. H., Peterson, B. A., Colless, M. & Saunders, W., 2006, *MNRAS*, **369**, 25

Jones, M. H. & Lambourne, R. J. A., 2004, *An Introduction to Galaxies and Cosmology*

Kaiser, N., 1987, *MNRAS*, **227**, 1

Kaiser, N. & Squires, G., 1993, *ApJ*, **404**, 441

Kepner, J., Fan, X., Bahcall, N., Gunn, J., Lupton, R. & Xu, G., 1999, *ApJ*, **517**, 78

Kim, R. S. J., Kepner, J. V., Postman, M., Strauss, M. A., Bahcall, N. A., Gunn, J. E., Lupton, R. H., Annis, J., Nichol, R. C., Castander, F. J., Brinkmann, J., Brunner, R. J., Connolly, A., Csabai, I., Hindsley, R. B., Ivezić, Ž., Vogeley, M. S. & York, D. G., 2002, *AJ*, **123**, 20

Koester, B. P., McKay, T. A., Annis, J., Wechsler, R. H., Evrard, A. E., Rozo, E., Bleem, L., Sheldon, E. S. & Johnston, D., 2007, *ApJ*, **660**, 221

Kolb, E. W. & Turner, M. S., 1990, *The early universe*

Koo, D. C., 1985, *AJ*, **90**, 418

Krauss, L. M., Srednicki, M. & Wilczek, F., 1986, *Physical Review D*, **33**, 2079

Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottlöber, S., Allgood, B. & Primack, J. R., 2004, *ApJ*, **609**, 35

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

Liddle, A. R. & Lyth, D. H., 2000, *Cosmological Inflation and Large-Scale Structure*

Lidman, C. E. & Peterson, B. A., 1996, *AJ*, **112**, 2454

Lima, M. & Hu, W., 2004, *Physical Review D*, **70**(4), 043504

Lima, M. & Hu, W., 2005, *Physical Review D*, **72**(4), 043006

Lin, H., Cunha, C., Lima, M., Oyaizu, H., Frieman, J., Collister, A., Lahav, O. & Dark Energy Survey, 2004, in *Bulletin of the American Astronomical Society*, volume 36 of *Bulletin of the American Astronomical Society*, pp. 1462–+

Loh, E. D. & Spillar, E. J., 1986, *ApJ*, **303**, 154

Longair, M. S. (ed.), 1998, *Galaxy formation*

Longair, M. S. & Seldner, M., 1979, *MNRAS*, **189**, 433

Lopes, P. A. A., de Carvalho, R. R., Gal, R. R., Djorgovski, S. G., Odewahn, S. C., Mahabal, A. A. & Brunner, R. J., 2004, *AJ*, **128**, 1017

Loveday, J., Peterson, B. A., Efstathiou, G. & Maddox, S. J., 1992, *ApJ*, **390**, 338

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

Lynds, R. & Petrosian, V., 1986, in *Bulletin of the American Astronomical Society*, volume 18 of *Bulletin of the American Astronomical Society*, pp. 1014–+

Maddox, S. J., Efstathiou, G., Sutherland, W. J. & Loveday, J., 1990, *MNRAS*, **243**, 692

Mantz, A., Allen, S. W., Ebeling, H. & Rapetti, D., 2008, *MNRAS*, **387**, 1179

Martínez, V. J. & Saar, E., 2003, *Statistics of galaxy clustering*

Mather, J. C., Cheng, E. S., Eplee, Jr., R. E., Isaacman, R. B., Meyer, S. S., Shafer, R. A., Weiss, R., Wright, E. L., Bennett, C. L., Boggess, N. W., Dwek, E., Gulkis, S., Hauser, M. G., Janssen, M., Kelsall, T., Lubin, P. M., Moseley, Jr., S. H., Murdock, T. L., Silverberg, R. F., Smoot, G. F. & Wilkinson, D. T., 1990, *ApJL*, **354**, L37

McCarthy, I. G., Holder, G. P., Babul, A. & Balogh, M. L., 2003, *ApJ*, **591**, 526

Metzler, C. A., White, M., Norman, M. & Loken, C., 1999, *ApJL*, **520**, L9

Miller, C. J., Batuski, D. J., Slinglend, K. A. & Hill, J. M., 1999, *ApJ*, **523**, 492

Milne, E. A., 1934, *The Quarterly Journal of Mathematics*, **5**, 64

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

Navarro, J. F., Frenk, C. S. & White, S. D. M., 1996, *ApJ*, **462**, 563

Nichol, R. C., 2001, *ArXiv Astrophysics e-prints*

Oyaizu, H., Cunha, C., Lima, M., Lin, H. & Frieman, J., 2006, in *Bulletin of the American Astronomical Society*, volume 38 of *Bulletin of the American Astronomical Society*, pp. 140–+

Oyaizu, H., Lima, M., Cunha, C. E., Lin, H. & Frieman, J., 2008*a*, *ApJ*, **689**, 709

Oyaizu, H., Lima, M., Cunha, C. E., Lin, H., Frieman, J. & Sheldon, E. S., 2008*b*, *ApJ*, **674**, 768

Padmanabhan, T., 1996, *Cosmology and Astrophysics through Problems*

Pagels, H. & Primack, J. R., 1982, *Physical Review Letters*, **48**, 223

Peacock, J., 2008, in *A Decade of Dark Energy*

Peacock, J. A., 1999, *Cosmological Physics*

Peacock, J. A., Cole, S., Norberg, P., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R. D., Colless, M., Collins, C., Couch, W., Dalton, G., Deeley, K., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Percival, W. J., Peterson, B. A., Price, I., Sutherland, W. & Taylor, K., 2001, *Nature*, **410**, 169

Peacock, J. A., Schneider, P., Efstathiou, G., Ellis, J. R., Leibundgut, B., Lilly, S. J. & Mellier, Y., 2006, *ESA-ESO Working Group on "Fundamental Cosmology"*, Technical report

Peccei, R. D. & Quinn, H. R., 1977, *Physical Review D*, **16**, 1791

Peebles, P. J. E., 1982, *ApJL*, **263**, L1

Peebles, P. J. E., 1993, *Physics Today*, **46**, 87

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

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

Penzias, A. A. & Wilson, R. W., 1965, *ApJ*, **142**, 419

Percival, W. J., Nichol, R. C., Eisenstein, D. J., Frieman, J. A., Fukugita, M., Loveday, J., Pope, A. C., Schneider, D. P., Szalay, A. S., Tegmark, M., Vogeley, M. S., Weinberg, D. H., Zehavi, I., Bahcall, N. A., Brinkmann, J., Connolly, A. J. & Meiksin, A., 2007, *ApJ*, **657**, 645

Percival, W. J., Reid, B. A., Eisenstein, D. J., Bahcall, N. A., Budavari, T., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Knapp, G. R., Kron, R. G., Loveday, J., Lupton, R. H., McKay, T. A., Meiksin, A., Nichol, R. C., Pope, A. C., Schlegel, D. J., Schneider, D. P., Spergel, D. N., Stoughton, C., Strauss, M. A., Szalay, A. S., Tegmark, M., Vogeley, M. S., Weinberg, D. H., York, D. G. & Zehavi, I., 2010, *MNRAS*, **401**, 2148

Perlmutter, S., Aldering, G., della Valle, M., Deustua, S., Ellis, R. S., Fabbro, S., Fruchter, A., Goldhaber, G., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Knop, R. A., Lidman, C., McMahon, R. G., Nugent, P., Pain, R., Panagia, N., Pennypacker, C. R., Ruiz-Lapuente, P., Schaefer, B. & Walton, N., 1998, *Nature*, **391**, 51

Postman, M., Lubin, L. M., Gunn, J. E., Oke, J. B., Hoessel, J. G., Schneider, D. P. & Christensen, J. A., 1996, *AJ*, **111**, 615

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

Prunet, S., Pichon, C., Aubert, D., Pogosyan, D., Teyssier, R. & Gottloeber, S., 2008, *ApJS*, **178**, 179

Ramella, M., Boschin, W., Fadda, D. & Nonino, M., 2001, *A&A*, **368**, 776

Ramella, M., Geller, M. J., Pisani, A. & da Costa, L. N., 2002, *AJ*, **123**, 2976

Reid, B. A., Percival, W. J., Eisenstein, D. J., Verde, L., Spergel, D. N., Skibba, R. A., Bahcall, N. A., Budavari, T., Frieman, J. A., Fukugita, M., Gott, J. R., Gunn, J. E., Ivezić, Ž., Knapp, G. R., Kron, R. G., Lupton, R. H., McKay, T. A., Meiksin, A., Nichol, R. C., Pope, A. C., Schlegel, D. J., Schneider, D. P., Stoughton, C., Strauss, M. A., Szalay, A. S., Tegmark, M., Vogeley, M. S., Weinberg, D. H., York, D. G. & Zehavi, I., 2010, *MNRAS*, **404**, 60

Reiprich, T. H. & Böhringer, H., 2002, *ApJ*, **567**, 716

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B. & Tonry, J., 1998, *AJ*, **116**, 1009

Rines, K., Diaferio, A. & Natarajan, P., 2007, *ApJ*, **657**, 183

Ross, N. P., da Ângela, J., Shanks, T., Wake, D. A., Cannon, R. D., Edge, A. C., Nichol, R. C., Outram, P. J., Colless, M., Couch, W. J., Croom, S. M., de Propris, R., Drinkwater, M. J., Eisenstein, D. J., Loveday, J., Pimbblet, K. A., Roseboom, I. G., Schneider, D. P., Sharp, R. G. & Weilbacher, P. M., 2007, *MNRAS*, **381**, 573

Rozo, E., Wechsler, R. H., Rykoff, E. S., Annis, J. T., Becker, M. R., Evrard, A. E., Frieman, J. A., Hansen, S. M., Hao, J., Johnston, D. E., Koester, B. P., McKay, T. A., Sheldon, E. S. & Weinberg, D. H., 2010, *ApJ*, **708**, 645

Rubin, V. C. & Ford, Jr., W. K., 1970, *ApJ*, **159**, 379

Sachs, R. K. & Wolfe, A. M., 1967, *ApJ*, **147**, 73

Sawangwit, U., Shanks, T., Abdalla, F. B., Cannon, R. D., Croom, S. M., Edge, A. C., Ross, N. P. & Wake, D. A., 2009, *ArXiv e-prints*

Schechter, P., 1976, *ApJ*, **203**, 297

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., Riess, A. G., Filippenko, A. V., Hamuy, M., Smith, R. C., Hogan, C., Stubbs, C., Diercks, A., Reiss, D., Gilliland, R., Tonry, J., Maza, J., Dressler, A., Walsh, J. & Ciardullo, R., 1998, *ApJ*, **507**, 46

Schneider, P. & Seitz, C., 1995, *A&A*, **294**, 411

Schneider, S. E., 1996, in E. D. Skillman (ed.), *The Minnesota Lectures on Extragalactic Neutral Hydrogen*, volume 106 of *Astronomical Society of the Pacific Conference Series*, pp. 323–+

Scranton, R., Connolly, A. J., Nichol, R. C., Stebbins, A., Szapudi, I., Eisenstein, D. J., Afshordi, N., Budavari, T., Csabai, I., Frieman, J. A., Gunn, J. E., Johnson, D., Loh, Y., Lupton, R. H., Miller, C. J., Sheldon, E. S., Sheth, R. S., Szalay, A. S., Tegmark, M. & Xu, Y., 2003, *ArXiv Astrophysics e-prints*

Seljak, U., 2000, *MNRAS*, **318**, 203

Shapley, H. & Ames, A., 1932, *Annals of Harvard College Observatory*, **88**, 41

Shectman, S. A., 1985, *ApJS*, **57**, 77

Sheth, R. K. & Diaferio, A., 2001, *MNRAS*, **322**, 901

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

Silk, J., 1968, *ApJ*, **151**, 459

Silverman, B. W., 1986, *Density estimation for statistics and data analysis*

Skibba, R., Sheth, R. K., Connolly, A. J. & Scranton, R., 2006, *MNRAS*, **369**, 68

Smith, J. A., Tucker, D. L., Kent, S., Richmond, M. W., Fukugita, M., Ichikawa, T., Ichikawa, S., Jorgensen, A. M., Uomoto, A., Gunn, J. E., Hamabe, M., Watanabe, M., Tolea, A., Henden, A.,

Annis, J., Pier, J. R., McKay, T. A., Brinkmann, J., Chen, B., Holtzman, J., Shimasaku, K. & York, D. G., 2002, *AJ*, **123**, 2121

Smith, S., 1936, *ApJ*, **83**, 23

Soares-Santos, M., de Carvalho, R. R., Annis, J., Gal, R. R., La Barbera, F., Lopes, P. A. A., Wechsler, R. H., Busha, M. T. & Gerke, B. F., 2011, *ApJ*, **727**, 45

Soucail, G., Fort, B., Mellier, Y. & Picat, J. P., 1987*a*, *A&A*, **172**, L14

Soucail, G., Mellier, Y., Fort, B., Mathez, G. & Hammer, F., 1987*b*, *A&A*, **184**, L7

Springel, V. & Hernquist, L., 2003, *MNRAS*, **339**, 312

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J. & Pearce, F., 2005, *Nature*, **435**, 629

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

Sunyaev, R. A. & Zeldovich, Y. B., 1970, *ApSS*, **7**, 3

Suyu, S. H., Marshall, P. J., Auger, M. W., Hilbert, S., Blandford, R. D., Koopmans, L. V. E., Fassnacht, C. D. & Treu, T., 2010, *ApJ*, **711**, 201

Tang, J., Abdalla, F. B. & Weller, J., 2008, *ArXiv e-prints*

Teyssier, R., 2002, *A&A*, **385**, 337

Teyssier, R., Pires, S., Prunet, S., Aubert, D., Pichon, C., Amara, A., Benabed, K., Colombi, S., Refregier, A. & Starck, J., 2009, *A&A*, **497**, 335

The Dark Energy Survey Collaboration, 2005, *ArXiv Astrophysics e-prints*

Thomas, S. A., Abdalla, F. B. & Lahav, O., 2010, *ArXiv e-prints*

Tinker, J. L., 2007, *MNRAS*, **374**, 477

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

Valentinuzzi, T., Woods, D., Fasano, G., Riello, M., D'Onofrio, M., Varela, J., Bettoni, D., Cava, A., Couch, W. J., Dressler, A., Fritz, J., Moles, M., Omizzolo, A., Poggianti, B. M. & Kjærgaard, P., 2009, *A&A*, **501**, 851

van der Marel, R. P., Magorrian, J., Carlberg, R. G., Yee, H. K. C. & Ellingson, E., 2000, *AJ*, **119**, 2038

van Haarlem, M. P., Frenk, C. S. & White, S. D. M., 1997, *MNRAS*, **287**, 817

Vanderlinde, K., Crawford, T. M., de Haan, T., Dudley, J. P., Shaw, L., Ade, P. A. R., Aird, K. A., Benson, B. A., Bleem, L. E., Brodwin, M., Carlstrom, J. E., Chang, C. L., Crites, A. T., Desai, S., Dobbs, M. A., Foley, R. J., George, E. M., Gladders, M. D., Hall, N. R., Halverson, N. W., High, F. W., Holder, G. P., Holzapfel, W. L., Hrubes, J. D., Joy, M., Keisler, R., Knox, L., Lee, A. T., Leitch, E. M., Loehr, A., Lueker, M., Marrone, D. P., McMahon, J. J., Mehl, J., Meyer, S. S., Mohr, J. J., Montroy, T. E., Ngeow, C., Padin, S., Plagge, T., Pryke, C., Reichardt, C. L., Rest, A., Ruel, J., Ruhl, J. E., Schaffer, K. K., Shirokoff, E., Song, J., Spieler, H. G., Stalder, B., Staniszewski, Z., Stark, A. A., Stubbs, C. W., van Engelen, A., Vieira, J. D., Williamson, R., Yang, Y., Zahn, O. & Zenteno, A., 2010, *ArXiv e-prints*

Viana, P. T. P., Kay, S. T., Liddle, A. R., Muanwong, O. & Thomas, P. A., 2003, *MNRAS*, **346**, 319

von Hoerner, S., 1960, *Zeitschrift fur Astrophysik*, **50**, 184

Weller, J. & Battye, R. A., 2003, *New Astronomy Review*, **47**, 775

Weller, J., Battye, R. A. & Kneissl, R., 2002, *Physical Review Letters*, **88**(23), 231301

Wirth, G. D., Willmer, C. N. A., Amico, P., Chaffee, F. H., Goodrich, R. W., Kwok, S., Lyke, J. E., Mader, J. A., Tran, H. D., Barger, A. J., Cowie, L. L., Capak, P., Coil, A. L., Cooper, M. C., Conrad, A., Davis, M., Faber, S. M., Hu, E. M., Koo, D. C., Le Mignant, D., Newman, J. A. & Songaila, A., 2004, *AJ*, **127**, 3121

Wu, H., Rozo, E. & Wechsler, R. H., 2010, *ApJ*, **713**, 1207

Xue, Y. & Wu, X., 2000, *ApJ*, **538**, 65

Yee, H. K. C. & Ellingson, E., 2003, *ApJ*, **585**, 215

York, D. G., Adelman, J., Anderson, Jr., J. E., Anderson, S. F., Annis, J. & Bahcall, N. A., 2000, *AJ*, **120**, 1579

Younger, J. D., Bahcall, N. A. & Bode, P., 2005, *ApJ*, **622**, 1

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

Zeilik, M. & Gregory, S., 1998, *Introductory Astronomy and Astrophysics* (Harcourt College Publishers)

Zel'Dovich, Y. B., 1970, *A&A*, **5**, 84

Zuiderwijk, E. J., 1995, *A&AS*, **112**, 537

Zwicky, F., 1933, *Helvetica Physica Acta*, **6**, 110

Zwicky, F., 1937, *ApJ*, **86**, 217

Zwicky, F., Herzog, E. & Wild, P., 1961, *Catalogue of galaxies and of clusters of galaxies, Vol. I*

Zwicky, F., Herzog, E. & Wild, P., 1963, *Catalogue of galaxies and of clusters of galaxies, Vol. 2*

Zwicky, F., Herzog, E. & Wild, P., 1966, *Catalogue of galaxies and of clusters of galaxies, Vol. 3*

Zwicky, F., Herzog, E. & Wild, P., 1968, *Catalogue of galaxies and of clusters of galaxies*

Zwicky, F., Karpowicz, M. & Kowal, C. T., 1965, *"Catalogue of Galaxies and of Clusters of Galaxies", Vol. V*

Zwicky, F. & Kowal, C. T., 1968, *"Catalogue of Galaxies and of Clusters of Galaxies", Volume VI*

Zwicky, F. & Zwicky, M. A., 1971, *Catalogue of selected compact galaxies and of post-eruptive galaxies*

ACKNOWLEDGEMENTS

I wish to acknowledge first of all my supervisor Filipe Abdalla for all of his patience with my progress, generosity with his time and day to day help. It is without any doubt that any merit found within this thesis should reflect kindly on his integrity both as an educator and as a friend. I only hope that, in time, I can repay the belief he has invested in me with continued hard work, dedication and perhaps a big steak.

Secondly, I wish to acknowledge Eduardo Cypriano for all his guidance and friendship throughout my PhD. I am sure I would not be where I am today without his support. I can now safely say that I have awakened from the nightmare. I would also like to thank both him and his wife Lys Figuredo for sharing their home with me and making me feel like part of the family while I was in Brazil. I am sure they will be great parents for little Amanda.

Thirdly, I wish to thank Chris Blake for all his extremely helpful comments and ideas, which have contributed significantly to the work presented in this thesis. I would also like to thank him for his hospitality in Australia.

I would like also to thank Cristiano Sabiu for working hard to make mocks for me.

I wish to express my gratitude to Ofer Lahav for giving me the opportunity to prove myself and overseeing the "big picture" of my work.

I would like to acknowledge various members of group A who have helped me at various points during my PhD, namely John Deacon, Gert Hutsi, Anais Rassat, Fabrizio Sidoli, Molly Swanson, Jiayu Tang, Shaun Thomas, Jochen Weller and Dugan Witherick.

I would also like to express my gratitude to Sarah Bridle and Bob Nichol for taking the time to read my thesis and being kind enough to be my examiners.

I dedicate a special thanks to Raman Prinja for his support and enthusiasm as my undergraduate tutor and for my first lecture on cosmology.

Thanks to all of the members of the ICE in Barcelona who have had enough belief in me to offer me a job and allow me to continue to work hard and prove myself.

I would like also to thank all the past and present members of G18 for great conversations and memories. My desk next to the door was like a home away from home.

Thanks to Dave, Ingo, Japh and Matt for a fun trip to Glasgow. The farthest I have ever travelled just to play football.

Thanks to Adam, Cris, DK, Molly, Ole and Stephanie for a very amusing week in Madrid. I still need to go back to el Museo Nacional del Prado to see the second floor.

Thanks to Stephen for joining me for a feast of metal at the Mastodon concert. Hopefully I will get a chance to go see Municipal Waste at some point in the future.

Thanks to Roger Wesson for his company on our epic eastern european adventure to see Rammstein. I am sure neither of us will ever forget a somewhat disturbing hostel owner in Belgrade.

Thanks to Daniele Fantin for being a good friend and for the great days out in London and Nottingham. I look forward to visiting you in Venezuela.

Thanks to Iraklis Konstantopoulos for being a good friend and for a memorable trip to almost every museum in Milano. Hopefully I will get a chance to visit the Monroeville Mall with you someday.

Thanks to everyone who has ever come along to play football with me in Regents Park, ULU, the Sobell Centre, or the Power League (R.I.P.). Playing football was one of the things I looked forward to most every week.

Thanks to Lana Korunoska for being a good friend especially during our undergraduate years. UCL is not the same without a weekly trip to the printers.

Thanks to James Wilson for always being a really good friend and looking after all my comic book needs. I will always remember the hilarious nights playing Big 2 in random common rooms across UCL.

I wish to thank Kevin Smith and Scott Mosier for making everyone in my office think I was crazy laughing out loud in front of my computer listening to Smodcast. You guys kept me sane.

I wish also to acknowledge Metal for being awesome!

I owe a debt of gratitude to UCL, the STFC and Hackney Council for giving me the opportunity to be educated.

I would like to thank Vinicio Soto for looking after my mother and Victoria Walters for looking after my father. You are both good friends and welcome additions to the family.

I dedicate a special thanks to my brother Níall, my sister Abby and my mother for all their love, support and tolerance over the years. You have all made it possible for me to follow my dream and believe in myself. Regardless of what happens to me in this life, I am always comforted by the knowledge that I am really fortunate to have such a wonderful family. I love you guys.

I dedicate another very special thanks to the best friend anyone could ever ask for and certainly the best friend I have ever had, Magda Vasta. She has given me the confidence and self belief that have helped to make me a better person in all aspects of my life (*e.g.* I will never forget to put the toilet seat down ever again). I would never have survived the difficulties I have encountered during my PhD without her support. I doubt I will ever find a better flatmate with whom I can enjoy nice food and horror films or anyone who can make me laugh as much as this comico esperto. Pinguina, ti voglio bene per sempre.

My final acknowledgement is for my father to whom this entire thesis is dedicated. His love, support and care have helped to mold me into the person I am today. For as long as I can remember he has emphasised the importance of education and the beauty of knowledge. He is and will always be the best teacher I have ever had in life and what I have learnt from him allows me to appreciate the privilege and responsibility that comes with wisdom. I will never be able to repay all the sacrifices he has made in order for me to follow my dream, but I will endeavour to pass on everything I have learned from him to others. Thank you.