OBSERVATIONAL CONSTRAINTS ON THE
STATISTICAL PROPERTIES OF CLUSTERS
OF GALAXIES AND THE HALO
OCCUPATION DISTRIBUTION

This thesis is submitted in partial fulfilment of the requirements
for the award of the degree Doctor of Philosophy
of the University of Portsmouth

January 2010

By
Mehri Torki

Supervisor: Prof. Robert Nichol
Abstract

The Halo Occupation Distribution (HOD) is used to quantify the mean number of galaxies in a dark matter halo as a function of the halo mass ($\langle N \rangle \propto M^\beta$). The most direct method of measuring the HOD is by counting galaxies in known clusters and groups of galaxies, but present measurements using this method appear to disagree. In this work, we revisit this problem and empirically determine the number of galaxies per cluster as a function of cluster mass, using a sample of 1713 nearby clusters ($0.03 < z < 0.13$) selected from the Sloan Digital Sky Survey using the C4 cluster algorithm.

A key part of our analysis is the determination of the virial mass and radius of each cluster ($M_{200}$ and $R_{200}$). The former is computed using an empirical relationship between the summed cluster (r-band) luminosity and the mass, while the latter is estimated from stacked radial profiles of the clusters as a function of mass. We find excellent agreement between these relationships and those in the literature based on stacked weak lensing studies of clusters. We also demonstrate that our $R_{200}$ values are robust (to less than a 20% uncertainty) against systematic uncertainties associated with the cluster centroids, fiber collisions and color and luminosity cuts.

We find a strong relation between $M_{200}$ and $\langle N \rangle$ (within $R_{200}$) for cluster masses $> 10^{13} M_\odot$. The measured slope ($\beta$) of the HOD is dominated by systematic uncertainties and we conservatively estimate our data is fully consistent with $\beta = 1 \pm 0.1$ (in agreement with previous measurements). We also find that the statistical scatter about this relation is Poisson, as expected from simulations. We find no evidence for any environmental dependence of our HOD which is a key assumption of the halo
model. We do find significant dependence on the colour of cluster galaxies, greater than the expected systematic and statistical uncertainties. For “red” galaxies (galaxies on the red sequence), we find $\beta \simeq 0.6$ for all galaxy luminosities, including Luminous Red Galaxies (LRG). For “blue” galaxies (bluer than the red sequence), we find the HOD slope decreases with increasing galaxy luminosity from $\beta \simeq 0.5$ for the brightest galaxies to $\beta \simeq 0.9$ for the faintest blue galaxies studied here. We repeated the HOD analysis for a C4 mock catalogue of clusters containing 388 objects, and derived the best fit HOD $\beta$ values for “all” and “red” galaxies. One interesting result is that for the “red” mock galaxies, we see that the slope of the HOD appears to be the same ($\beta \simeq 0.6$) as the slope of the HOD for the real data, and this is robust.

We measure for the first time the cluster correlation function as a function of mass. We measured the best–fit to $\xi(r)$ over the range $2-40 h^{-1}$ Mpc, and found that the correlation function has the correlation length $r_0 = 22.5 \pm 6.0 h^{-1}$ Mpc and $\gamma = 1.7 \pm 0.2$ for the mass range $> 10^{14} M_\odot$, and $r_0 = 16.8 \pm 5.0 h^{-1}$ Mpc and $\gamma = 1.9 \pm 0.3$ for the mass range $< 10^{14} M_\odot$. We compare our cluster correlation function with that of the N–body simulations and found a close but not exact match between them. Moreover, we repeated the same analysis for a C4 mock catalogue of clusters containing 388 objects, and found the correlation length $r_0 = 16.27 \pm 5.5 h^{-1}$ Mpc and $\gamma = 2.1 \pm 0.2$. And the correlation function for the whole mass range of our cluster sample (real data), has the amplitude $r_0 = 16.8 \pm 4.9 h^{-1}$ Mpc and $\gamma = 1.9 \pm 0.3$. These values of $r_0$ and $\gamma$ are in good agreement. We carried out an analysis of the correlation function for both data and mock, and investigated how the correlation length ($r_0$) varies with cluster mass. We measure the $r_0$ values for four mass bins. We obtain the $r_0$ values for both the mock and the real data in agreement (within the error bars), and they are smaller for the smaller bins of mass and they increase for the larger bins of mass. Comparison of the results between the real and the mock data indicates that the C4 catalogue is a fair representation of the SDSS.
Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.
I thank my supervisor Prof. Robert Nichol for his guidance during my study and all the members of the ICG for their help, support, kindness and encouragement. I also acknowledge helpful discussions with Steven Bamford, Edd Edmondson, Ben Hoyle, Cristiano Sabiu, Ofer Lahav, Mat Smith, David Wake, Jim Cresswell, Will Percival during this work and specially Bjorn Schafer and Jussi Valiviita for their extensive help. I thank the ICG Portsmouth for financial assistance during this research and also for the people responsible for providing computer support.

A special thank to Prof. Roy Maartens for all his valuable and human help and support, the words have not been invented yet powerful enough to thank him.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago,
Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.
Contents

Abstract 2

Declaration 4

Acknowledgements 5

1 Introduction 14

1.1 Friedmann-Lemaitre-Robertson-Walker Cosmology . . . . . . . . 14

1.1.1 The Robertson-Walker metric . . . . . . . . . . . . . . . . 14

1.2 Hubble Law . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

1.3 Redshift . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

1.4 The era of precision cosmology . . . . . . . . . . . . . . . . . . . 18

1.5 Galaxy Clustering . . . . . . . . . . . . . . . . . . . . . . . . . . 19

1.5.1 Parametrisation of the HOD . . . . . . . . . . . . . . . . . . 23

1.6 The Sloan Digital Sky Survey . . . . . . . . . . . . . . . . . . . . 25

1.7 SDSS Data Releases . . . . . . . . . . . . . . . . . . . . . . . . . 27

1.8 SDSS results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 29

1.9 Content of the thesis . . . . . . . . . . . . . . . . . . . . . . . . 32

2 C4 - SDSS Galaxy Cluster Catalogue 35

2.1 C4 algorithm . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 35
CONTENTS

2.2 Measured parameters ................................. 41
2.3 Mass Estimation .................................. 43

3 Determination of $R_{200}$ .......................... 50
3.1 Group membership ................................. 51
3.1.1 Fibre collision correction ...................... 52
3.1.2 Colour cut .................................. 53
3.1.3 Mean space density of field galaxies ........... 54
3.2 Choosing cluster mass bins to compute the galaxy density profile .......................... 59
3.3 Testing $R_{200}$ .................................. 60
3.3.1 The effect of cluster centroids .................. 63
3.3.2 Checking the colour magnitude fit .............. 63
3.4 Colour and magnitude constraints ................ 69
3.5 Comparison with other results .................... 71
3.6 $R_{200}$ for X-ray detections ...................... 73
3.7 Comparison with Berlind catalogue ............... 74
3.8 Radial Distribution of Galaxies in Clusters ........ 82
3.8.1 Profile fitting ................................ 87
3.8.2 Mass dependence ............................. 90

4 Halo Occupation Distribution ....................... 94
4.1 HOD derivation ................................ 94
4.1.1 Calculating the best fit HOD parameters .......... 98
4.2 Scatter about the HOD ............................ 102
4.3 Environmental dependence of the HOD ............ 103
4.4 Discussion and Conclusions ....................... 105
4.4.1 The HOD for all galaxies ...................... 105
4.4.2 The HOD as a function of galaxy colours ....... 108
CONTENTS

4.4.3 The HOD for mock catalogue ........................................ 113

5 Cluster correlation function ........................................... 116
  5.1 The two point correlation function .................................. 116
    5.1.1 The estimators .................................................. 117
  5.2 Applying the cluster correlation function to the halo model .... 118
    5.2.1 Cluster correlation function .................................... 120
  5.3 Comparison with Numerical Simulation ............................. 123
    5.3.1 The cluster correlation function for mock catalogue ...... 131

6 Conclusion ........................................................................ 136

Bibliography ....................................................................... 140
List of Tables

1.1 SDSS photometric bands .............................................. 27
2.1 Mass estimation .......................................................... 49
3.1 Mass bins ................................................................. 59
3.2 Colour and magnitude constraints ................................. 70
3.3 $R_{200}$ for X-ray clusters ........................................... 74
3.4 NFW mass dependence ............................................... 92
4.1 Best fit HOD $\beta$ parameter ...................................... 99
4.2 Best fit HOD $\beta$ parameter, fibre collision effect ........ 99
4.3 Best fit HOD parameter $\beta$ in literature ........................ 108
4.4 Best fit HOD parameter for mock catalogue .................... 113
5.1 Mass bins for investigating the correlation length of mock data . 133
5.2 Mass bins for investigating the correlation length of real data . 133
List of Figures

1.1 Halo model ................................................. 24
1.2 SDSS imaging spectral sky coverage ...................... 30
1.3 SDSS redshift cone diagram ............................... 31
1.4 SDSS results ................................................ 33

2.1 Completeness check ......................................... 39
2.2 Completeness check ......................................... 39
2.3 The colour-magnitude relations ......................... 40
2.4 C4 purity and completeness ................................ 41
2.5 Luminosity and the halo mass relationship ............ 42
2.6 Distribution of C4 clusters on the sky ................. 43
2.7 Velocity dispersion profile ............................... 44
2.8 Velocity dispersion profile ............................... 45
2.9 Mass estimation ............................................. 48

3.1 Completeness limit ......................................... 55
3.2 Radial density profiles, 11 mass bins ................... 56
3.3 Radial density profiles, 4 mass bins .................... 61
3.4 Radial density profiles, 22 mass bins ................... 62
3.5 Colour-magnitude diagram ............................... 64
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>colour-magnitude diagram</td>
<td>65</td>
</tr>
<tr>
<td>3.7</td>
<td>colour-magnitude diagram</td>
<td>66</td>
</tr>
<tr>
<td>3.8</td>
<td>colour-magnitude diagram</td>
<td>67</td>
</tr>
<tr>
<td>3.9</td>
<td>colour-magnitude diagram</td>
<td>68</td>
</tr>
<tr>
<td>3.10</td>
<td>$R_{200}$ comparison</td>
<td>72</td>
</tr>
<tr>
<td>3.11</td>
<td>$R_{200}$ comparison</td>
<td>73</td>
</tr>
<tr>
<td>3.12</td>
<td>Comparing masses</td>
<td>75</td>
</tr>
<tr>
<td>3.13</td>
<td>Richness comparison, C4 and Berlind catalogue</td>
<td>78</td>
</tr>
<tr>
<td>3.14</td>
<td>Virial radii comparison, C4 and Berlind catalogue</td>
<td>79</td>
</tr>
<tr>
<td>3.15</td>
<td>Velocity dispersions comparison, C4 and Berlind catalogue</td>
<td>80</td>
</tr>
<tr>
<td>3.16</td>
<td>RA and DEC for C4 cluster</td>
<td>81</td>
</tr>
<tr>
<td>3.17</td>
<td>RA and DEC for C4 cluster with apertures</td>
<td>82</td>
</tr>
<tr>
<td>3.18</td>
<td>X-ray and optical luminosity relationship</td>
<td>83</td>
</tr>
<tr>
<td>3.19</td>
<td>X-ray luminosity and velocity dispersion for C4</td>
<td>84</td>
</tr>
<tr>
<td>3.20</td>
<td>X-ray luminosity and velocity dispersion for Berlind catalogue</td>
<td>85</td>
</tr>
<tr>
<td>3.21</td>
<td>X-ray luminosity and velocity dispersion from literature</td>
<td>86</td>
</tr>
<tr>
<td>3.22</td>
<td>Surface density profile</td>
<td>88</td>
</tr>
<tr>
<td>3.23</td>
<td>Likelihood contours</td>
<td>90</td>
</tr>
<tr>
<td>3.24</td>
<td>Concentration parameter–Mass relation</td>
<td>93</td>
</tr>
<tr>
<td>4.1</td>
<td>HOD for all galaxies</td>
<td>100</td>
</tr>
<tr>
<td>4.2</td>
<td>HOD for red galaxies</td>
<td>111</td>
</tr>
<tr>
<td>4.3</td>
<td>HOD for blue galaxies</td>
<td>112</td>
</tr>
<tr>
<td>4.4</td>
<td>HOD for all mock galaxies</td>
<td>114</td>
</tr>
<tr>
<td>4.5</td>
<td>HOD for red mock galaxies</td>
<td>115</td>
</tr>
<tr>
<td>5.1</td>
<td>Histogram of clusters with redshifts</td>
<td>123</td>
</tr>
<tr>
<td>5.2</td>
<td>Correlation Function</td>
<td>124</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>5.3</td>
<td>Correlation Function</td>
<td>125</td>
</tr>
<tr>
<td>5.4</td>
<td>Correlation Function</td>
<td>126</td>
</tr>
<tr>
<td>5.5</td>
<td>Correlation Function</td>
<td>127</td>
</tr>
<tr>
<td>5.6</td>
<td>Correlation Function</td>
<td>128</td>
</tr>
<tr>
<td>5.7</td>
<td>Correlation Function</td>
<td>131</td>
</tr>
<tr>
<td>5.8</td>
<td>Correlation Function for mock data</td>
<td>134</td>
</tr>
<tr>
<td>5.9</td>
<td>Correlation Function for mock and real data</td>
<td>135</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Recently, new technologies offer the possibility to study the evolutionary history of the universe. Cosmology in general, and the field of structure formation in particular, has been a ‘hot’ research topic for many years. Recent spectacular observational findings like the discovery by the COBE satellite in 1992 of fluctuations in the temperature of the cosmic microwave background (Smoot et al. 1992), have ushered in a ‘golden age of cosmology’.

The introduction presented in this thesis is mostly concentrated on the Halo Model (see Section 1.5) and the SDSS (see Section 1.5.1).

1.1 Friedmann-Lemaitre-Robertson-Walker Cosmology

1.1.1 The Robertson-Walker metric

As discussed in Lucchin (1998), the cosmological principle requires both the isotropy and homogeneity of the universe, averaged over sufficiently large scales and in this way, there are no preferred positions in the universe. In particular, the
important astronomical finding which supports the cosmological principle, is the
observation of the Cosmic Microwave Background (CMB) which shows that the
universe is isotropic on large scales, to the order of 0.01% (Smoot et al. 1991).
If the cosmological principle is valid, the metric that describes the space–time is
given by Robertson-Walker metric as,

$$ds^2 = dt^2 - a(t)^2 \left[ \frac{dr^2}{1 - Kr^2} + r^2(d\theta^2 + \sin^2 \theta d\phi^2) \right],$$  \hspace{1cm} (1.1)

where $r$, $\theta$ and $\phi$ are comoving coordinates ($r$ is dimensionless) and their value
will not change for comoving points (as defined by moving with the expansion of
the universe), and the time coordinate $t$ is cosmological time. The coefficient $a(t)$
is called the scale factor which is dimensionless (the current value of scale factor is
normalised to be one), and the conformal time $\tau$ is defined as $d\tau = dt/a(t)$. The
curvature parameter $K$ is a constant and does not change with the expansion of
the universe and its value can be 1, 0, $-1$ corresponding to a geometrically closed,
flat and open universe respectively.

The Friedmann equations are obtained by solving the Einstein field equations
for the Robertson-Walker metric and a perfect fluid,

$$\left(\frac{\dot{a}}{a}\right)^2 = H^2 = \frac{8\pi G}{3} \rho - \frac{K}{a^2},$$  \hspace{1cm} (1.2)

and,

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p),$$  \hspace{1cm} (1.3)

where $H$ is the Hubble parameter (see Section 1.2), $\rho$ is the mean energy density
of the universe and $p$ is the pressure.

The current energy density of the universe has various contributions, described
by the dimensionless parameters,
\[
\Omega_i = \frac{\rho_i}{\rho_{0c}},
\] (1.4)
where \(\rho_{0c}\) is the critical density today (the subscript 0 means the current value),
\[
\rho_{0c} = \frac{3H_0^2}{8\pi G} \simeq 1.9 \times 10^{-29} h^2 g \text{ cm}^{-3},
\] (1.5)
where \(G\) is Newton’s gravitational constant and \(H_0 = 100h \text{ km/s/Mpc}\) is the current Hubble constant.

These components consist of: radiation, baryonic matter (luminous matter), dark matter (non-luminous matter) and dark energy (vacuum energy).

It is possible to express the Friedmann equations in a form that contains the density parameters,
\[
\left(\frac{H}{H_0}\right)^2 = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda
\] (1.6)
where \(\Omega_r\) is the contribution of the radiation density, \(\Omega_m\) is what matter contributes including both baryonic and dark matter, \(\Omega_k\) is the curvature density parameter, \(\Omega_\Lambda\) is the cosmological constant (vacuum) density. In the study of structure formation, the radiation term can be neglected since the radiation density decreases with \(a\) rapidly after decoupling.

The WMAP satellite experiment (Komatsu et al. 2008) has provided accurate measurements of these density parameters: \(\Omega_r h^2 = 0.02267^{+0.00058}_{-0.00059}, \Omega_m h^2 = 0.1358^{+0.0037}_{-0.0036}, H_0 = 70.5 \pm 1.3 \text{ km s}^{-1} \text{ Mpc}^{-1}, -0.0179 < \Omega_k < 0.0081 (95 \% \text{ CL}), \Omega_\Lambda = 0.726 \pm 0.015.\)

\(^{1}\)WMAP (Wilkinson Microwave Anisotropy Probe), is a NASA Explorer mission which measured the fluctuations in the radiation from the early universe, see Komatsu et al. (2008).
1.2 Hubble Law

In 1929 Edwin Hubble discovered that galaxies are moving away from us with a recessional velocity \( v \) that is proportional to their distance \( d \) from us,

\[
v = H_0 d.
\]

Equation (1.7) is called Hubble’s Law. \( H_0 \) determines the time scale of the universe, and is an important quantity in astrophysics. The estimation of the current value of the Hubble constant derived from the WMAP is \( H_0 = 70.5 \pm 1.3 \text{ km s}^{-1} \text{ Mpc}^{-1} \) (Komatsu et al. 2008).

1.3 Redshift

When the light is emitted from a source moving away from the observer, the wavelength of radiation increases, shifting towards the redder end of the spectrum (Doppler effect). Redshift is defined as,

\[
z = \frac{\lambda_0 - \lambda_e}{\lambda_e} = \frac{\lambda_0}{\lambda_e} - 1
\]

(1.8)

where \( \lambda_0 \) is the wavelength of the observed radiation, and \( \lambda_e \) is the wavelength of the emitted radiation.

In cosmology, the redshift is caused by the expansion of the universe, and photons are redshifted during their travel from their source of emission to the observer. This therefore relates to the scale factor \( a \) at the time of emission and observation,

\[
1 + z = \frac{a_0}{a},
\]

(1.9)

The universe is expected to evolve with redshift. Redshift is determined through
spectroscopy. Study of star and galaxy spectra tells us about our evolving universe: what stars and galaxies are made of, how far away they are (about position and movements), how fast they are moving. By observing $z$ we can show that the universe is expanding, this redshift is basis of Hubbles’ Law (Eq. 1.7).

1.4 The era of precision cosmology

Recently, the cosmological scenario has changed remarkably. If the universe contains only matter and radiation, we expect that its expansion would slow down by the gravitational attraction exerted by the matter on itself. Observations of supernovae showed that in fact the expansion has been speeding up, leading to the need for a new form of energy called dark energy which can transform the decelerating to an accelerated expansion.

The standard inflationary models predict that the universe is practically flat today. Also the analysis of cosmic microwave background temperature anisotropies suggests that the universe is flat. This means that the total density should be equal to the critical one (Eq. 1.5). The ordinary matter and the radiation could only explain less than 5% of this value (WMAP: $\Omega_b \approx 0.046$). Now the question is whether dark matter gives the missing energy for a flat universe. In recent years, the new observations have confirmed that the total energy density in the form of matter (including dark matter) is less than half of the critical value.

To reconcile these indications, it is necessary to suppose that there exists a form of non luminous energy, dark energy, which dominates the present universe. It is not subject to gravitational collapse, at least not on very large scales, because we do not observe it in the matter halos associated with galaxies. Moreover, if this entity had always dominated the other components, it would have been impossible for the visible present structures to form. So, dark energy must have
been negligible in the past. Finally, we know from Eq. (1.3) that it must possess a negative pressure; this property can distinguish dark energy from the ordinary matter and the dark matter.

Type Ia supernovae, when treated as standard candles, are a direct test for the existence of dark energy and they suggest that the expansion of the universe is speeding up rather than slowing down (e.g. Riess et al. 1998, Perlmutter et al. 1999). An accelerating universe also receives independent support from CMB, large scale structure observations and Baryon Acoustic Oscillation (Dunkley et al. 2008, Tegmark et al. 2006, Percival et al. 2007). The simplest model that can explain these observations is a cosmological constant, $\Lambda$. From the theoretical point of view it can be justified by the energy associated with the quantum vacuum fluctuations (Bludman and Ruderman, 1977); but this cosmological scenario suffers from the serious problems of fine tuning (Zlatev et al. 1999). This has induced theorists to elaborate a large number of models.

### 1.5 Galaxy Clustering

Galaxy and cluster surveys give us a wealth of observational information and are therefore invaluable in our cosmological analysis. Clusters of galaxies are the most massive (up to $10^{15} M_\odot$) collapsed structures in the universe that are gravitationally bound and they contain large amounts of dark matter and also baryons. The three major mass components in clusters are: stellar mass, hot gas of the Inter Cluster Medium (ICM) and Dark Matter. Therefore clusters are used for cosmological analysis in particular to find the mass distribution of the universe. So, the precise determination of their masses is crucial for understanding the matter distribution and also for understanding of the formation and evolution of the structures. Historically cluster masses have been estimated in different ways.
Examples of these mass proxies are X-ray temperature and luminosities of the intra–cluster gas, gravitational lensing, optical luminosities, dynamical analysis of the observed velocity dispersion of cluster galaxies based on the virial theorem, and the number of galaxies in the cluster. In particular, the gravitational lensing cluster masses are obtained independently of the components of the cluster. Wu & Fang (1997) have compared these three different measurements of cluster mass; the cluster masses which are estimated from the velocity dispersion of galaxies, the X–ray cluster masses which are obtained from the X–ray emitted gas temperature, and the gravitational strong/weak lensing cluster masses. They have shown that the gravitational lensing cluster masses are in agreement with the dynamical masses obtained from the velocity dispersion of cluster galaxies and this means that the galaxies follow the gravitational potential of clusters, (and there is no bias between the velocities of the dark matter particles and the galaxies in clusters). But when these two methods are compared with the X–ray method, we see a discrepancy and underestimation of the cluster masses (by a factor of 2–3) which could be because of the simplification in the models for the X–ray gas distribution and dynamical evolution. In the analysis in this thesis, we have used the optical luminosity as a mass proxy, as discussed by Miller et al. (2005). This proved to be an excellent tracer of cluster mass (see Section 2.3).

Apart from the information that we obtain from the cluster surveys about the cluster masses, the surveys of cluster of galaxies give us another important information which is their redshift that tells us about the evolution of the universe (see Nichol 2001 for more details). Also the clusters can give us information about their formation history (Schindler 2002).

A detailed knowledge of how light traces the underlying dark matter is required to use galaxy surveys to constrain cosmology. Traditionally, this has been achieved using a bias factor ($b$) which relates the amount of observed light to
the unseen dark matter. A generic non–linear, non–local and stochastic model of
galaxy bias is given in Dekel & Lahav (1999) as,

$$\delta_{\text{gal}}(x) = b(\delta_{\text{dm}}, x)\delta_{\text{dm}}(x) + \epsilon,$$

where $\delta_{\text{gal}}$ is the galaxy overdensity field, $\delta_{\text{dm}}$ is the dark matter overdensity field, and $\epsilon$ is a stochastic term. The non–linear, non–local and deterministic model is,

$$\delta_{\text{gal}}(x) = b(\delta_{\text{dm}}, x)\delta_{\text{dm}}(x),$$

and a simple linear, local and deterministic model for bias is described by,

$$\delta_{\text{gal}}(x) = b\delta_{\text{dm}}(x).$$

Therefore, the relationship between the galaxy–galaxy and matter–matter correlation function (Kaiser 1984, Bardeen et al. 1986) is,

$$\xi_{gg}(r) = b^2\xi_{mm}(r),$$

or we can describe it in terms of galaxy and matter power spectra $P_g(k) = b^2P_m(k)$. This bias factor is assumed to be linear on large scales, but in general can be a function of both scale and galaxy properties. Cole et al. (2005) suggest an analytic model for the scale dependence of $b$ given by,

$$P_{gg}(k) = b^2P_{\text{Lin}}(k)\frac{1 + Qk^2}{1 + Ak},$$

where $A = 1.4$ and $Q$ varies. This model is referred to as the $Q$–model. Seljak (2001), Schulz & White (2006), Guzik et al. (2007) suggest another model, referred to as the $P$–model, given by,

$$P_{gg}(k) = b^2P_{\text{Lin}}(k) + P,$$
where the constant $P$ is a parametrisation of the small scale turn–up in power and this model does at least as well as the $Q$–model (Cresswell & Percival 2008). As discussed in Nichol (2008), there has also been some suggestion about expressing bias as a function of galaxy luminosity (Norberg et al. 2001, Tegmark et al. 2004) as,

$$\frac{b}{b_*} = 0.85 + 0.15 \frac{L}{L_*} + 0.04 (M_* - M_{0.1r}),$$

(1.15)

where $b_*$ is the bias of $L_*$ ($M_*$) galaxies, $L_*$ and $M_*$ are the characteristic luminosity and the turnover magnitude of the Schechter (1976) luminosity function and $M_{0.1r}$ is the absolute magnitude in $r$-band at redshift $z = 0.1$. Cresswell & Percival (2008) presented estimates of the large–scale bias parameter $b$ for the red and blue galaxies to be in the range of 1.17 to 1.44 for the red galaxies and 0.89 to 1.09 for the blue galaxies respectively. They show that the large–scale bias parameter $b$ changes as a function of luminosity.

In recent years, there has been a resurgence in the use of the halo model to describe how galaxies trace the dark matter. The halo model formalism was first considered by Neyman & Scott (1952), providing a theory of how galaxies are distributed in space under some postulates, which are: galaxies appear to be only inside the clusters, the number and the distribution of galaxies is not the same for each cluster and changes according to a probabilistic law, also the same assumption is valid for the distribution of the centres of clusters.

In recent years, the halo model has been investigated by many authors (e.g. Peacock & Smith 2000, Seljak 2000, Scoccimarro et al. 2001, Berlind & Weinberg 2002, Cooray & Sheth 2002, Berlind et al. 2003, Magliocchetti & Porciani 2003, van den Bosch et al. 2003). As an example, Berlind & Weinberg investigated the statistics of the clustering of galaxies and the effect of bias while motivated by the Halo Occupation Distribution, HOD (described below). Their method is to
apply the HOD models to N-body simulations, and then to examine and analyse the relation between the HOD parameters and the clustering statistics.

The halo model assumes that all galaxies in the Universe reside in virialized units of mass (see Figure 1.1), called halos, that can be understood through universal scaling relations (e.g., Navarro, Frenk & White 1997). On large scales, the clustering of galaxies reflects the clustering of galaxies between different halos (of different masses) and can be understood using the halo bias and clustering strength. On small scales, the clustering of galaxies is dominated by the details of how the galaxies occupy their host halos, as a function of halo mass and galaxy properties. This is included in the halo model, which gives the probability of a halo of mass \( M \) containing \( N \) galaxies of a specific type \( (P(N|M)) \). Therefore, to fully exploit the power and flexibility of the halo model, it is important to accurately determine the HOD over a range of different halo masses and galaxy types (see Zheng & Weinberg 2007). As discussed in Berlind and Weinberg (2002), the power of the HOD formalism is that it can give us physical information about the galaxy formation (the theory of galaxy formation tells us how the halos are populated with galaxies) and the statistics of clustering of galaxies. Therefore by deriving the HOD empirically, we can find out what galaxy clustering can tell us about the physics of galaxy formation.

### 1.5.1 Parametrisation of the HOD

With the HOD approach and through the halo model, we can study the clustering of galaxies analytically and the shape of the HOD has been widely studied by the theoretical methods. As an example, Berlind et al. (2003) have derived the HOD shape by using simulations and semi-analytical methods, which are in good agreement; the shape is a step, then a flat region and then a power law to describe
Figure 1.1: The halo model of clustering represents two terms, the distribution of matter inside the same halo, and the distribution of halos in space, $\xi_{DM}(r) = \xi_{1h}(r) + \xi_{2h}(r)$. All physics can be decomposed similarly: influences from within halo, versus from outside, Sheth 1996 (figure taken from lahmu.phyast.pitt.edu/~sheth/courses/allahabad/halomodel.ppt).

Two kinds of models have been demonstrated in Berlind and Weinberg (2002): the first is:

$$N_{\text{avg}} = \begin{cases} 
0 & \text{if } M < M_{\min} \\
(M/M_1)^\alpha & \text{otherwise}, 
\end{cases}$$

(1.16)

where $N_{\text{avg}}$ is the mean number of galaxies that populate dark matter halos of mass $M$, $M_{\min}$ is the cut-off mass of the halo, which means that halos of mass less than $M_{\min}$ do not have galaxies, $M_1$ is the amplitude which corresponds to the mass of the halos that have one galaxy on average and $\alpha$ is the power law index in this relation.

The second model is described as:

$$N_{\text{avg}} = \begin{cases} 
0 & \text{if } M < M_{\min} \\
(M/M_1')^\alpha & \text{if } M_{\min} \leq M \leq M_{\text{crit}} \\
(M/M_1'')^\beta & \text{otherwise}, 
\end{cases}$$

(1.17)
where $\alpha$ is the power law index for low mass and $\beta$ is the power law index for high mass, $M_{\text{crit}}$ is the mass of the halo for which the power law index changes from $\alpha$ to $\beta$, the value of $M_1$ is chosen to produce a galaxy population of the desired space density, $M'_1$ is not a free parameter and is defined as $M'_1 = M_1^{\alpha/\beta} M_{\text{ crit}}^{(1-\alpha/\beta)}$ (see Berlind and Weinberg, 2002, for more details).

Kravtsov et al. (2004) have also studied the HOD shape by using subhalos and have presented the shape of the HOD as a combination of central and satellite galaxies as:

$$N_c = \begin{cases} 
1 & \text{for } M \geq M_{\text{min}} \\
0 & \text{for } M < M_{\text{min}}
\end{cases} \quad (1.18)$$

and

$$\langle N_s \rangle \propto \mu^\beta \quad (1.19)$$

where $\mu \equiv M/M_{\text{min}}$, $N_c$ is the number of central galaxies, $N_s$ is the number of satellite galaxies.

Therefore we see that the HOD is a useful method and allows us to study galaxy clustering analytically. By using the HOD formalism, we can investigate and analyse different populations of galaxies and their properties (e.g, red and blue, colour and luminosity) and the bias effect related to them.

## 1.6 The Sloan Digital Sky Survey

The first major galaxy redshift survey was the CfA and illustrated the structure of galaxies (‘Great Wall’), which was swiftly followed by IRAS, LCRS, SSRS, 2dFGRS and SDSS surveys (Geller & Huchra 1989, Efstathiou et al. 1990, Huchra & Geller 1991, Shectman et al. 1996, da Costa et al. 1998, Colless et al. 2001,
Gunn & Weinberg 1995). They have helped to determined our governing cosmological model through precision measurements of the galaxy power spectrum, the redshift–space distortions, the galaxy bias (as a function of morphological type, luminosity, etc), the constraints on primordial non–gaussianity and the large–scale filamentary or web–like inter–cluster structure. These are all low redshift surveys. The DEEP (Davis et al. 2003) and the VIRMOS (Le Fèvre et al. 2004) are the surveys at high redshift which allow a more detailed study of the evolution of clustering over comic time.

Motivated by the Halo Model, we apply it to the SDSS–C4 cluster catalogue (Miller et al. 2005). The Sloan Digital Sky Survey (SDSS) is a large spectroscopic survey with imaging in five bands of u, g, r, i and z (Table 1.1), which provides optical images and covers more than a quarter of the sky. This 2.5-meter optical telescope is situated at Apache Point Observatory in New Mexico (Gunn et al. 2006). The photometry is calibrated by using data from three different telescopes and three different pipelines (see Tucker et al. 2006).

The survey started in 2000. Galaxies are targeted for spectroscopy in two samples: the Main Galaxy sample (Strauss et al. 2002), which is flux-limited to $r = 17.7$, and the Luminous Red Galaxy (LRG) sample (Eisenstein et al. 2001). LRGs can be targeted at greater distances than typical Main survey galaxies because of the intrinsic luminosity of the LRGs. The finite size of the fibres means that objects separated by more than 55" can be uniquely targeted on a given plate. This restriction results in $\sim 10\%$ incompleteness in galaxy spectroscopy (Blanton et al. 2003a). However this incompleteness is characterized well and it is possible to correct for it (e.g., Zehavi et al. 2002).

In June 2005, the SDSS achieved its first phase of operations known as SDSS-I. In the first five years, 8,000 square degrees of the sky were imaged in five bandpasses described above. It detected nearly 200 million celestial objects, and
Table 1.1: SDSS photometric bands with magnitude limits (see http://www.sdss.org/).

<table>
<thead>
<tr>
<th>Band</th>
<th>Average wavelength (Å)</th>
<th>Magnitude limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>3551</td>
<td>22.0</td>
</tr>
<tr>
<td>g</td>
<td>4686</td>
<td>22.2</td>
</tr>
<tr>
<td>r</td>
<td>6165</td>
<td>22.2</td>
</tr>
<tr>
<td>i</td>
<td>7481</td>
<td>21.3</td>
</tr>
<tr>
<td>z</td>
<td>8931</td>
<td>20.5</td>
</tr>
</tbody>
</table>

it measured spectra of more than 675,000 galaxies, 90,000 quasars, and 185,000 stars. These data have enabled studies from solar system scales up to the large scale structure of the Universe.

1.7 SDSS Data Releases

The regular Data Releases of SDSS have been DR1-DR4 (Abazajian et al. 2003, 2004, 2005; Adelman-McCarthy et al. 2006). In our analysis, we use Data Release 5 (DR5; Adelman-McCarthy et al. 2007). There have been no substantive changes to the imaging or spectroscopic software since DR2, so DR5 includes data identical to DR2-DR4 in the overlapping regions. DR5 contains 8,000 square degrees of imaging from which 215 million objects have been found, including stars, galaxies, nebulae etc. In addition, spectroscopy has been obtained for 1,048,960 objects, including 675,000 galaxies (see SDSS web page). The DR5 spectroscopic data covers 5740 square degrees (see Figure 1.2). We have added a redshift cone diagram in Figure 1.3 to illustrate better the extent of the volume coverage of the SDSS.

We have used a flexible web site (http://www.sdss.org/dr5) access to the SDSS database, CAS (the Catalogue Archive Server), which is a useful resource
for gathering data found in the SDSS based on our selection criteria: their location, photometric parameters, and spectroscopic parameters (if they were observed spectroscopically).

Another web site used in our work is http://nvodevs1.ctio.noao.edu/dr5_vac/, Value Added Catalogue (VAC) database (by K. S. Krughoff and C. J. Miller) containing 663442 SDSS-DR5 galaxies, which allows us to do scientific analyses on the SDSS spectroscopic galaxy data, to query and to use spectral information (i.e. emission line-widths). It is also possible to upload any information that belongs to specific galaxies.

The SDSS has various magnitude expressions (Lupton, Gunn, Szalay 1999). We use VAC (see the last paragraph) to download the model magnitudes as the preferred ones which are hereby described. In order to create optimal measures of the fluxes of a galaxy, two galaxy models are fit to the two-dimensional image of each object in each band. A pure deVaucouleurs profile (de Vaucouleurs, 1948) has surface brightness,

\[ I(r) = I_0 \exp \left(-7.67(r/r_e)^{1/4}\right), \]

where \( r_e \) is the radius which contains half of the total luminosity and \( I_0 \) is the surface brightness at \( r_0 \). It is truncated beyond \( 7r_e \) to smoothly go to zero at \( 8r_0 \) with some softening within \( r = r_e/50 \). This is fitted to the light profile of galaxies with a pure exponential profile with surface brightness profile,

\[ I(r) = I_0 \exp \left(-1.68(r/r_e)\right), \]

in such a way that it is truncated beyond \( 3r_e \) to smoothly go to zero at \( 4r_e \).
CHAPTER 1. INTRODUCTION

The other magnitude expression in SDSS is the Petrosian magnitude (Petrosian 1976) which measures the galaxy fluxes within a circular aperture and is set by the profile of the galaxy, and measures a constant fraction of a galaxy’s total light regardless of the amplitude of its surface brightness profile (see Blanton et al. 2001). The Petrosian magnitude is better to use for extended sources and the model magnitude for point sources. However the model magnitude is now a good estimator for a universal magnitude of all types of objects and has advantages over Petrosian magnitude, in particular, it may be less affected by crowding than Petrosian magnitudes. The model magnitude has smaller errors than the Petrosian magnitude, is close to optimal noise properties, and is the most useful for galaxies (information available at http://www.sdss.org).

The SDSS photometry is on the AB system (Oke & Gunn 1983). As described in Blanton & Roweis (2007), the best estimate for the AB offset is \( [u, g, r, i, z] = [-0.036, +0.012, +0.010, +0.028, +0.040] \). Systematic errors of \( [\Delta u, \Delta g, \Delta r, \Delta i, \Delta z] = [0.05, 0.02, 0.02, 0.02, 0.03] \) should be added to the SDSS magnitude errors in quadrature.

1.8 SDSS results

One of the key goals of the SDSS has been a precision measurement of how galaxies cluster under the influence of gravity. The measurement of the cosmic clustering of galaxies and dark matter gives us the possibility of better understanding the structure and evolution of the universe.

The three-dimensional SDSS map of the universe has over 200,000 galaxies up to two billion light years away over 6% of the sky. By combining these measurements with those from WMAP (see Komatsu et al. 2008), we find that the cosmic matter consists of \( \Omega_{\Lambda} = 0.726 \pm 0.015, \Omega_m h^2 = 0.1358^{+0.0037}_{-0.0036} \)
Figure 1.2: Imaging (Top) and Spectral (Bottom) Sky coverage for the SDSS Data Release 5, shown in J2000 equatorial coordinates. Shaded areas represent regions for which imaging or spectroscopy have been obtained (the lightly shaded regions indicate those newly released in DR5). Aitoff projection of equatorial coordinates. Figures are from Adelman-McCarthy et al. (2007).
Figure 1.3: In this figure (from Blanton et al. 2003b), equatorial distribution of right ascension and redshift for Main Sample galaxies from SDSS–DR1 is shown. The galaxies are within 6 degree of the Celestial Equator. In the analysis of our work, we use SDSS–DR5 data which has the same redshift limit and therefore it shows the correct extent in redshift.
(H_0 = 70.5 \pm 1.3 \text{ km s}^{-1} \text{ Mpc}^{-1}), \text{ and } \Omega_b h^2 = 0.02267_{-0.00059}^{+0.00058} \text{ (see Figure 1.4).}

SDSS data helps to improve the accuracy of WMAP data on the cosmic matter density and the Hubble parameter. The WMAP–SDSS measurements agree well with the previous results that combined WMAP with the Anglo-Australian 2dF galaxy redshift survey (2dFGRS; Colless et al. 2001).

DR5 has been used in our analysis, but the Sixth Data Release (DR6) of the Sloan Digital Sky Survey is now available. The spatial coverage of DR6 is about 7% larger than that of DR5. The photometric data in DR6 are based on five-band imaging observations of 8520 square degrees of sky, and include measures of 300 million unique objects. SDSS-II contains three surveys: the Sloan Legacy Survey (a continuation of SDSS-I), the Sloan Extension for Galactic Understanding and Exploration (SEGUE), and the Sloan Supernova Survey (see SDSS web page for the information provided).

1.9 Content of the thesis

This chapter contains a review of the elements of modern cosmology and how the galaxy and the cluster surveys are used for the cosmological analysis and the information that they can give us. In particular, it describes how with the halo model approach, we aim to study the clustering of the galaxies while using the SDSS dataset. Chapter 2 presents the C4–SDSS galaxy cluster catalogue and the algorithm and the measured parameters which has been used. This Chapter also includes the mass estimation method for our clusters. In Chapter 3, we define and measure the virial sizes of our clusters called as R_{200}. The measurements of mass and size which have been described in Chapters 2 and 3 are an important

\footnote{At the time of submission, DR7 was available (Abazajian et al. 2008).}
Figure 1.4: The SDSS–DR3 results (black dots) give an accurate measurement of how the density of the Universe fluctuates on scales of millions of light-years. These and other cosmological measurements agree with the theoretical prediction (blue curve) for a Universe composed of 5 percent baryons, 25% dark matter and 70% dark energy. The figure is taken from SDSS web site.
and main part of this thesis. Chapter 4 presents the Halo Occupation Distribution derivation and revisits the problem of disagreement between the present measurements of HOD, and we discuss and interpret our results of the HOD derivation. In Chapter 5, we derive the cluster correlation function of our sample and compare it with the simulations and the mock data. Finally, we conclude in Chapter 6 with a brief summary of our work and some future plan. For the cosmological model, throughout our work, we assumed a flat, ΛCDM cosmology with $H_0 = 100 \, h \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$ ($h = 1$), $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$. 
Chapter 2

C4 - SDSS Galaxy Cluster Catalogue

2.1 C4 algorithm

For the analysis presented in this thesis, we use the SDSS-C4 cluster catalogue, which is described in detail in Miller et al. (2005). Briefly, the C4 catalogue uses an algorithm that identifies clusters from the SDSS main galaxy spectroscopic sample (Strauss et al. 2002) as over-densities in a seven-dimensional position and colour space: 4 colours are u-g, g-r, r-i, i-z and 3 spatial coordinates which are the Right Ascension (RA), Declination (DEC) and redshift, based on the fact that galaxies within clusters are evolving together (see Figure 2.3). Thus this is a way of minimising projection effects (or false-positive detections) that have caused problems for previous attempts to identify optical clusters. As discussed by many authors (e.g. Miller 2000, Postman et al. 1992, Nichol et al. 1992, Sutherland 1988, Lucey et al. 1983), previous catalogues have suffered from strong projection effects causing difficulties in the identification of real clusters. The problem is that many apparent clusters are only a projection effect and not a physical association.
of galaxies. The term contamination is usually used to mark the non-cluster members projected on the clusters. As discussed by Collins et al. (2000), the projection effects are important to be looked at and there is an evidence for this fact which reveals when we compare the amplitude of the correlation function in the two decomposed directions of $\pi$ and $\sigma$ where $\sigma$ is the distance measured on the plane of sky and $\pi$ is the distance measured along the line-of-sight (the redshift direction). This is the isotropy test. When a cluster sample like Abell (1958) cluster catalogue, fails this test, we see the line-of-sight elongations in the contours of $\xi(\sigma, \pi)$. In this way, the correlation function increases artificially (Sutherland 1988). However, when the cluster correlation function is measured by using the the X-ray cluster samples, we can overcome this problem (Collins et al. 2000). The other importance of the C4 catalogue algorithm and the key difference between it and the previous cluster-finding algorithms is that it requires that the colours of nearby galaxies are the same as the target galaxy. In this way, the C4 algorithm is sensitive to the range of clusters like a cluster dominated by a “blue” population of galaxies, compared to the colours of the field galaxies (see Nichol 2004).

Figure 2.4 (right) shows the completeness of the C4 algorithm as a function of halo mass ($M_{200}$) selected from the mock SDSS catalogue (see Miller et al. 2005). As we see in this figure, the C4 catalogue remains more than 90% complete for the systems with $M_{200} \geq 2 \times 10^{14} M_\odot$ and up to a redshift of $z \sim 0.12$. For the lower mass systems, there may be some incompleteness in the original dark matter halo catalogue because of the mass resolution of the Hubble Volume simulations. Solving this problem needs more work to be done in order to determine accurately the completeness for lower mass systems of the C4 catalogue. Note that for the analysis in this thesis we have used C4–DR5 sample and we have checked that the completeness of the sample of C4–DR5 clusters is the same as
Miller et al. (2005) sample. The selection function is measured for the clustering algorithm itself on SDSS mocks. Since the SDSS photometry and spectra did not change between DR2 and DR5, the selection function is the same by definition and only the size of the real dataset has changed. Figures 2.1 and 2.2 show the comoving number density of clusters as a function of redshift, this confirms that our sample is virtually complete up to a redshift of $z \sim 0.12$. Completeness is important since maximising completeness that is controlling the selection function is a challenge in building a robust optically selected cluster catalogue, and the C4 catalogue has the great advantage of having such a high completeness. Understanding the selection function is important for all the statistical analyses that use the catalogue and this has not been traditionally considered for optical cluster catalogues (see Bramel, Nichol, & Pope 2000, Kochanek et al. 2003).

The purity is the percentage detected in the mock catalogue by using the C4 algorithm which can match a dark matter halo in the Hubble Volume simulations. We can see from Figure 2.1 (left) that the purity of the C4 catalogue remains at 100% for the most massive systems and it drops to $\sim 90\%$ for the remaining. Also the measure of the purity as a function of velocity dispersion (by using only the clusters that contain ten or more galaxies), finds that the C4 catalogue is 100% pure for these systems. This high purity in C4 has great importance since the algorithm aims to produce real clusters in a high-dimensional space. The catalogue contains a wide range of systems including groups of galaxies with a minimum of only 10 members, up to rich clusters with several hundreds of galaxy members.

Figure 2.5 shows the relation between the halo mass and the summed cluster $r$–band luminosity. The total summed $r$–band optical luminosity for each cluster is obtained by converting the apparent magnitudes of all cluster members into optical luminosities, using the conversions in Fukugita et al. 1996, and summing
 CHAPTER 2. C4 - SDSS GALAXY CLUSTER CATALOGUE

them. All magnitudes are $k$-corrected according to Blanton et al. 2003a and extinction corrected according to Schlegel, Finkbeiner, and Davis 1998. The figure is separated by the value of the Structure Contamination Flag (SCF), which is defined to measure the degree of isolation in redshift-space. As discussed in Miller et al. 2005, the radial variations of the velocity dispersion have been examined for each of the clusters and then an SCF is determined which is based on the ratio of the standard deviation of the dispersions over the mean of the velocity dispersions. The C4 catalogue has provided the velocity dispersion profiles for each cluster. We see that SCF increases when the standard deviations increase (see Figures 2.7 and 2.8 for some examples). Note the important fact that the summed luminosity, which is a cluster mass proxy, is less contaminated by nearby large scale structure than velocity dispersion. The scatter is smaller (reduced by a factor of two when we use SCF=0 clusters) for mass versus luminosity than it is for the mass versus richness or mass versus velocity dispersion. This gives the conclusion that in the mock catalogues, the summed $r$–band luminosity is the best measure of the cluster dark matter halo mass (Miller et al. 2005).

We use an updated version of the C4 catalogue presented in Miller et al. (2005), which includes clusters selected from the SDSS Data Release Five (DR5; Adelman-McCarthy et al. 2007). This new catalogue contains 1713 clusters and groups (Figure 2.6 shows the C4,DR5 distribution on the sky), compared to 748 systems published by Miller et al. (2005), but uses the same algorithm as in Miller et al. This new catalogue also contains the same measured cluster parameters as outlined in Miller et al. (2005), e.g., cluster velocity dispersion, summed total luminosity, companion flag, etc.
Figure 2.1: The comoving number density of clusters as a function of redshift. The feature at $z = 0.08$ is the Sloan Great Wall of galaxies (see Gott et al. 2005).

Figure 2.2: The comoving number density of clusters as a function of redshift. This figure is the same as Figure 2.1 but for a larger range of redshift.
Figure 2.3: The colour-magnitude relations for a previously undiscovered cluster with redshift 0.06 in the Early Data Release of the SDSS detected by the C4 algorithm. The black dots are galaxies within an aperture of $1h^{-1}$ Mpc around the cluster centre. Red and green dots are cluster members. The red dots are defined to be galaxies with $H\alpha$ equivalent widths $<4$ angstroms and are within 1000kpc of the cluster centre. The green dots are defined to be galaxies with $H\alpha$ equivalent widths $>4$ angstroms and are within 1000kpc of the cluster centre, representing starforming galaxies. Note the tight correlation in colour of the points with error bars on the colours of the red galaxies. This correlation is the E/S0 ridge-line which is co-evolving population of galaxies having a similar spectral energy distribution, otherwise known as “red envelope” (Baum 1959; McClure & van den Bergh 1968; Lasker 1970; Visvanathan & Sandage 1977). This is what the C4 algorithm has used in order to find the clusters. Figure from Miller et al. (2005).
2.2 Measured parameters

There is a large number of measured parameters in the C4 catalogue, including positions, velocity dispersions, number counts, colour-magnitude diagram, radii, summed $r$-band optical luminosity ($L_r$). In our analysis, as we will discuss later, we also use the colour-magnitude diagram, which provides the slope and y-intercept of the [r-i vs. r] and [g-r vs. r], together with the 1 sigma error on the y-intercept. As discussed by Miller et al. (2005), the centres of the clusters of galaxies are computed with three different methods. In this way the cluster centroid measurements are called BCG, MEAN and GEOM cluster centres. The BCG cluster centres means the cluster centroid measured is the position of the brightest galaxy in the cluster. The MEAN are the coordinates of the galaxy with the highest density. The GEOM are the luminosity weighted mean centroids which means the cluster centres are calculated using all galaxies within $1h^{-1}$Mpc;
Figure 2.5: The relationship between total $r$-band cluster luminosity and the halo mass (top) or velocity dispersion (bottom), as a function of the Structure Contamination Flag (SCF). (Figure from Miller et al. 2005).
an average $r$-band weighted luminosity is then calculated. We think the BCG centroid is the best one to use, because this method relies on the observations that clusters have a massive elliptical at the centre. The algorithm used to find the clusters identifies the BCG as the galaxy that maximises the likelihood (the statistic depends on luminosity, colour, and the number of fainter neighbours of similar colour). So, the positions measured in the C4 Catalogue with these three methods are called: RA_MEAN, DEC_MEAN, RA_GEOM, DEC_GEOM, RA_BCG, DEC_BCG. Later we will discuss the effect of different cluster centres on our result in measuring the virial radius.

2.3 Mass Estimation

A key issue in measuring the HOD is an accurate determination of the halo mass. Even though mass is not directly observable, by identifying a mass tracer and its
Figure 2.7: The velocity histogram (top) and the velocity dispersion profile (bottom). The errors on the spectroscopic redshifts are of the order 0.0001 (Strauss et al. 2002), and the errors on velocity dispersion are of the order 100 km/s. Red means passive, Yellow means C4 galaxy. Black filled circles are every spectroscopic galaxy projected onto the sky. The galaxy velocity dispersion profile has a high level of scatter in the outer radii, causing the SCF flag to be set to 3 (figures are from C4 webpage). The radius plotted here in the velocity dispersion profile is in $h^{-1}$kpc, and the errors are $30h^{-1}$kpc.
This cluster has SCF=0 (figures are from C4 webpage). The errors on the spectroscopic redshifts are of the order 0.0001 (Strauss et al. 2002), and the errors on velocity dispersion are $\sim 100\,\text{km/s}$. The radius plotted here in the velocity dispersion profile is in $h^{-1}\text{kpc}$, and the errors are $30h^{-1}\text{kpc}$. 
relation to halo mass, we hope to constrain cosmology. Combining measurements of the cluster power spectrum with theory could place constraints on $\sigma_8$ and $\Omega_m$ (see Section 5.2 for more details). Some examples of mass tracers in clusters are X-ray temperatures and luminosities of the intra-cluster gas, optical luminosities, and the number of galaxies found in the cluster. We use the summed r–band luminosities ($L_r$) of all galaxies in the cluster as a tracer of the halo mass. Miller et al. (2005) have shown that this cluster parameter is an excellent tracer of $M_{200}$ (by a factor of two better than the velocity dispersion and the richness), especially for systems not embedded in complex large-scale structures (i.e., with their SCF set to zero). This is in agreement with the findings of Popesso et al. (2006) and Lin et al. (2004), who studied samples of nearby X–ray clusters and concluded that the summed optical/IR luminosity of all the cluster members is tightly correlated with the X-ray luminosity of the cluster and the underlying dark mass. Likewise, Johnston et al. (2007) studied the weak lensing mass of clusters as a function of both richness and luminosity, and concluded that luminosity was more closely correlated with mass than galaxy counts (see also Reyes et al. 2008). Therefore, we use the measured $L_r$ values from the C4 catalogue and the non–parametric $M_{200}–L_r$ scaling relation given in Miller et al. (see Figure 2.9 made by Miller et al. 2005) to compute the $M_{200}$ of each system. We note this is the $M_{200}$ as measured in the Hubble Volume simulation and therefore, consistent with the definition of mass used in the halo model (Cooray & Sheth 2002). The error on the halo mass is determined from the one sigma irreducible scatter about this fit as given in Miller et al. (2005).

These results on $M_{200}$ were used in Bamford et al. (2008) as part of their analysis of the Galaxy Zoo data.

We provide the numerical values of our $M_{200}–L_r$ scaling relation in Table 2.1. We have compared this relation to the $M_{200}–L_{200}$ from Johnston et al. (2007)
and find broadly consistent results (within 50%). A detailed comparison is not possible because of differences in the definition of the summed luminosities: Miller et al. (2005) use a fixed metric radius (of $1h^{-1}\text{Mpc}$), while Johnston et al. used $R_{200}$. Also, Miller et al. use the SDSS r-band compared to Johnston et al. who used the i-band.
Figure 2.9: This figure shows the correlation between the summed $r$-band cluster luminosities and the halo mass for the simulations. The $M_{200}$ for the C4 clusters used in this analysis was obtained using the non-parametric $M_{200}$ to $(L_r$ relationship derived from Miller et al. 2005) and given in Table 2.1. This relationship was derived using mock catalogues of galaxies built on top of the Hubble Volume simulations and the $M_{200}$ used came directly from the halo catalogue of this simulation. This was achieved by matching clusters found in the mock galaxy catalogue (based on their colours, as in the real data) with the detected (underlying) dark matter halos. In this way, a $M_{200}$ to luminosity relationship was empirically constructed with no ambiguity in the definition of mass between the theoretical models and the observables i.e., our $M_{200}$ is the same as used by simulators. The grey solid and dashed lines are fits to mass versus luminosity and luminosity versus mass respectively. The solid black line is a non-parametric fit to the data and the dashed black lines were obtained by plotting the one sigma confidence band around this best fit. The shaded grey region contains the one sigma irreducible scatter of the data about the best non-parametric fit and includes effects from matching, measurement error, and any intrinsic scatter in this relationship (Figure is from Miller et al. 2005).
Table 2.1: The relationship between luminosity and mass for C4 clusters taken originally from Miller et al. (2005). We provide the mean relation, as well as the upper and lower error on the irreducible scatter in the cluster mass which includes effects from matching (halos to the C4 clusters, see caption in Figure 2.9), measurement error, and any intrinsic scatter in this relationship.

<table>
<thead>
<tr>
<th>Summed ( L_r ) ( (10^{11} L_\odot) )</th>
<th>Lower mass ( (10^{13} M_\odot) )</th>
<th>Upper mass ( (10^{13} M_\odot) )</th>
<th>Mean mass ( (10^{13} M_\odot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.01</td>
<td>5.01</td>
<td>5.33</td>
<td>5.17</td>
</tr>
<tr>
<td>2.11</td>
<td>5.58</td>
<td>5.91</td>
<td>5.74</td>
</tr>
<tr>
<td>2.215</td>
<td>6.20</td>
<td>6.55</td>
<td>6.37</td>
</tr>
<tr>
<td>2.325</td>
<td>6.90</td>
<td>7.255</td>
<td>7.08</td>
</tr>
<tr>
<td>2.44</td>
<td>7.68</td>
<td>8.04</td>
<td>7.86</td>
</tr>
<tr>
<td>2.56</td>
<td>8.54</td>
<td>8.91</td>
<td>8.72</td>
</tr>
<tr>
<td>2.69</td>
<td>9.50</td>
<td>9.88</td>
<td>9.69</td>
</tr>
<tr>
<td>2.825</td>
<td>10.56</td>
<td>10.95</td>
<td>10.75</td>
</tr>
<tr>
<td>2.97</td>
<td>11.74</td>
<td>12.14</td>
<td>11.94</td>
</tr>
<tr>
<td>3.12</td>
<td>13.05</td>
<td>13.47</td>
<td>13.26</td>
</tr>
<tr>
<td>3.285</td>
<td>14.50</td>
<td>14.95</td>
<td>14.72</td>
</tr>
<tr>
<td>3.46</td>
<td>16.11</td>
<td>16.59</td>
<td>16.35</td>
</tr>
<tr>
<td>3.65</td>
<td>17.89</td>
<td>18.42</td>
<td>18.15</td>
</tr>
<tr>
<td>3.85</td>
<td>19.85</td>
<td>20.46</td>
<td>20.15</td>
</tr>
<tr>
<td>4.07</td>
<td>22.03</td>
<td>22.73</td>
<td>22.38</td>
</tr>
<tr>
<td>4.31</td>
<td>24.43</td>
<td>25.26</td>
<td>24.845</td>
</tr>
<tr>
<td>4.565</td>
<td>27.09</td>
<td>28.09</td>
<td>27.59</td>
</tr>
<tr>
<td>4.85</td>
<td>30.03</td>
<td>31.24</td>
<td>30.63</td>
</tr>
<tr>
<td>5.15</td>
<td>33.28</td>
<td>34.75</td>
<td>34.01</td>
</tr>
<tr>
<td>5.49</td>
<td>36.88</td>
<td>38.665</td>
<td>37.76</td>
</tr>
<tr>
<td>5.855</td>
<td>40.85</td>
<td>43.03</td>
<td>41.93</td>
</tr>
<tr>
<td>6.27</td>
<td>45.24</td>
<td>47.90</td>
<td>46.55</td>
</tr>
<tr>
<td>6.73</td>
<td>50.10</td>
<td>53.32</td>
<td>51.69</td>
</tr>
<tr>
<td>7.24</td>
<td>55.48</td>
<td>59.37</td>
<td>57.39</td>
</tr>
<tr>
<td>7.81</td>
<td>61.425</td>
<td>66.10</td>
<td>63.72</td>
</tr>
<tr>
<td>8.42</td>
<td>68.00</td>
<td>73.61</td>
<td>70.75</td>
</tr>
<tr>
<td>9.08</td>
<td>75.28</td>
<td>81.975</td>
<td>78.56</td>
</tr>
<tr>
<td>9.79</td>
<td>83.335</td>
<td>91.295</td>
<td>87.22</td>
</tr>
<tr>
<td>10.55</td>
<td>92.25</td>
<td>101.68</td>
<td>96.85</td>
</tr>
<tr>
<td>11.34</td>
<td>102.11</td>
<td>113.25</td>
<td>107.53</td>
</tr>
<tr>
<td>12.18</td>
<td>113.00</td>
<td>126.15</td>
<td>119.40</td>
</tr>
<tr>
<td>13.11</td>
<td>125.02</td>
<td>140.57</td>
<td>132.565</td>
</tr>
<tr>
<td>14.15</td>
<td>138.28</td>
<td>156.68</td>
<td>147.19</td>
</tr>
<tr>
<td>15.29</td>
<td>152.91</td>
<td>174.68</td>
<td>163.43</td>
</tr>
<tr>
<td>16.54</td>
<td>169.06</td>
<td>194.76</td>
<td>181.46</td>
</tr>
<tr>
<td>17.90</td>
<td>186.92</td>
<td>217.18</td>
<td>201.48</td>
</tr>
<tr>
<td>19.36</td>
<td>206.66</td>
<td>242.16</td>
<td>223.71</td>
</tr>
<tr>
<td>20.93</td>
<td>228.51</td>
<td>269.99</td>
<td>248.39</td>
</tr>
<tr>
<td>22.59</td>
<td>252.72</td>
<td>300.97</td>
<td>275.79</td>
</tr>
<tr>
<td>24.33</td>
<td>279.56</td>
<td>335.41</td>
<td>306.22</td>
</tr>
</tbody>
</table>
Chapter 3

Determination of $R_{200}$

In addition to knowing the mass of each halo, we require the number of galaxies in each cluster to construct the HOD. For this purpose, we must first determine the radius of the clusters. This is achieved using a methodology similar to Hansen et al. (2005) to determine $R_{200}$, which is the radius interior to which the mean number density of galaxies is 200 times the critical density, as a function of mass. Simulations have shown that this is equivalent to the virial radius of massive clusters of galaxies.

The difference between our work and Hansen et al. is that in their analyses, they rely on projected photometric data taken in multiple band-passes, and then correct for the foreground and background galaxies that contaminate the line-of-sight to each cluster. In their investigation, systems of galaxies come in a range of masses, from single galaxies in larger halos to the most massive clusters, and they use the full range of their cluster finder to develop their catalogue of systems. So, they refer to any system with two or more galaxies as a “cluster”. Clusters used in their study are detected by the maxBCG algorithm, which relies on the observation that early-type galaxies with a small dispersion in colour are hosted by clusters; these cluster members lie on the red sequence on a colour-magnitude
CHAPTER 3. DETERMINATION OF $R_{200}$

The Hansen et al. methodology is based on using the population subtraction methods, i.e., applying the background subtraction techniques to the SDSS data to find the galaxies associated with the clusters, given a set of cluster centres with three-dimensional positions, and then determining the radius at which the cluster galaxy number density is $200\Omega^{-1}_{m}$ times the mean galaxy density, without assuming a model for the radial distribution of galaxies in clusters. In our analysis, we use the SDSS–C4 spectroscopic data and a different algorithm (see Section 2.1 for details of the C4 cluster catalogue algorithm).

Unfortunately we cannot directly measure $R_{200}$ for each individual C4 cluster, as many systems only possess a handful of bright galaxies. Therefore, we must compute $R_{200}$ from stacked cluster profiles as a function of the cluster mass. In detail, we compute the galaxy density profile from adding together all cluster galaxies in a series of eleven cluster mass bins. These bins were chosen to contain approximately equal numbers of galaxies and were empirically determined to provide a series of density profiles with the same signal–to–noise, e.g., the lower mass bins have more clusters each, but each cluster has fewer galaxies to contribute to the overall stacked profile. We provide details of the eleven mass bins in Table 3.1 which includes the mass range for each bin, the number of clusters and galaxies respectively in the bin and the computed $R_{200}$.

The results of this chapter are partly based on Torki et al. (2008).

3.1 Group membership

To construct the stacked radial profiles, we must determine which SDSS galaxies to include as members of C4 clusters. First, we restricted our analysis to the redshift range $0.03 < z < 0.13$ (both clusters and galaxies) and a luminosity range $-24 < M_r < -21.2$. The lower luminosity limit (shown in Figure 3.1)
corresponds to the apparent magnitude limit of $r = 17.77$ (Strauss et al. 2002) at $z = 0.13$ (all absolute magnitudes were k-corrected using the the kcorrect code of Blanton & Roweis 2007). Next, we only used galaxies within a projected radius of $4 \, h^{-1}\text{Mpc}$ and $\pm 4\sigma_v$ of the cluster redshift, where $\sigma_v$ is the measured velocity dispersion of the C4 clusters, to be consistent with Miller et al. (2005). We apply the constraint that members should be within $\pm 4\sigma_v$ to reject obvious foreground and background galaxies. (Note that we do not let the redshift cylinders go beyond the redshift limits set above; so the cluster redshift plus $4\sigma_v$ must be less than 0.13 and the cluster redshift minus $4\sigma_v$ must be greater than 0.03). These cuts produce a pseudo volume–limited sample of 154997 galaxies located near 1624 C4 clusters in the SDSS DR5.

### 3.1.1 Fibre collision correction

One problem that needs to be addressed is the issue of fibre collision, i.e., the SDSS can not place two spectroscopic fibres closer than 55 arcsec on a single plate. This corresponds to $R < 0.09 \, R_{200}$ for the majority of our clusters (assuming a mean redshift of $z = 0.1$, where most of our clusters are located). Overall, this problem has been mitigated by allowing overlapping spectroscopic plates, but there is still some incompleteness in the spectroscopic survey especially in dense regions like the cores of clusters (see Blanton et al. 2003a for details of this issue). Therefore, to correct for fibre collisions we search the SDSS photometric database for galaxies in the directions of our clusters that should have been targeted for spectroscopy (i.e., satisfy the criteria of Strauss et al. 2002) but do not have a measured redshift. In total, we found 46027 photometric galaxies, compared to the 154997 galaxies with a redshift. We assigned them a redshift equal to the redshift of their host C4 cluster. After applying the cuts to the redshift range...
(0.03 < z < 0.13) and the luminosity range ($-24 < M_r < -21.2$) as mentioned in the Section 3.1, we found 11887 galaxies and then these galaxies were added to the spectroscopic sample.

### 3.1.2 Colour cut

As a final step, we also apply a colour cut to the SDSS galaxies to ensure they are legitimate cluster members and further reduce the contamination from interlopers, especially at large cluster radii where the cluster members are swamped by bluer field galaxies. This cut follows the principle of the C4 algorithm that clusters are seen as over-densities in both space and colour, i.e., clusters display a tight red sequence or ridge line in their colour–magnitude diagrams (Ostrander et al. 1998, Gladders & Yee 2000). This is achieved using a linear fit to the red sequence in the $r$ versus $(r - i)$ colour–magnitude plane, as supplied as part of the C4 cluster catalogue (see Miller et al. 2005). In particular, we use the C4 measurements of the red sequence for each cluster, namely the slope and intercept of a fit to the $(r - i)$ versus $r$-band colour–magnitude diagram, as well as the one sigma error ($\sigma_{\text{width}}$) on the fitted intercept. Using these values, we only keep galaxies that possess $(r - i)$ colours (given their $r$ magnitude) within $\pm 2\sigma_{\text{width}}$ of the host cluster red sequence.

These criteria reduce the number of SDSS galaxies to 83928 available for our stacked radial density profiles, of which 6603 of the galaxies are the ones that have been added after the fibre collision correction. (Note that the 2 sigma range above was empirically determined and we did not add the errors on the slope because the C4 catalogue does not provide them. Also we do not look at the substructure flag and thus included clusters with and without a companion).

---

1We use parameters $RI_{\text{SLOPE}}$ and $RI_{\text{INTER}}$ from the C4 webpage.
In Figure 3.2, we show the cumulative stacked radial density profiles for the eleven cluster mass bins discussed above (see also Table 3.1). We have divided the number of galaxies in each radial bin by the physical volume within that radial shell, assuming the clusters are spherical. To determine $R_{200}$, we then find the radius at which the number density of cluster galaxies is $200 \Omega^{-1} m^3$ times the field density. This involved dividing the radial profiles in Figure 3.2 by the mean space density of field galaxies to get the fractional excess above the field, and interpolating to gain an accurate estimate of $R_{200}$ for each radial profile. These results on $R_{200}$ were used in Bamford et al. (2008) as part of their analysis of the Galaxy Zoo data.

3.1.3 Mean space density of field galaxies

A key component of our $R_{200}$ determination is the estimation of the mean field space density. We estimated this in two ways; firstly with the use of the luminosity function. The luminosity function $LF$ relates the number density of galaxies to absolute magnitude and it enables comparison between the sample of galaxies observed and the general population of galaxies. The form of this function can be fitted by a universal Schechter function (Schechter 1976). We integrated the Blanton et al. (2003b) luminosity function over the absolute magnitude limits applied to our galaxy sample ($M_r < -21.2$),

$$
\phi(M) dM = 0.4 \ln(10) \phi_* 10^{-0.4(M - M_*)} (\alpha + 1) \\
\times \exp(-10^{-0.4(M - M_*)}) dM,
$$

where $\alpha$ is the faint-end slope and $M_*$ is the turnover magnitude. In the rest-frame luminosities in $0.1 r$ band (i.e., the band has been k-corrected for $z = 0.1$ instead of $z = 0.0$), the best fit Schechter function to their results has $\phi_* =$
Figure 3.1: Absolute magnitude in r band versus redshift for the SDSS database (from http://nvodevs1.ctio.noao.edu/dr5_vac/querypage.html). The lower horizontal line shows $M_r = -21.2$ as our limit of completeness for the redshift range of $0.03 < z < 0.13$ and the upper shows $M_r = -20.5$ as another possible limit of completeness for the redshift range of $0.03 < z < 0.10$ (i.e., the completeness is defined to be the fraction of galaxies satisfying the selection criteria that are identified as spectroscopic targets). We have chosen the first range since it contains more galaxies.
Figure 3.2: (Top) The stacked radial density profiles of galaxies in C4 clusters of the eleven different bins of mass given in Table 3.1. In the analysis, we have excluded C4 clusters with less than 10 galaxies. These profiles were used to determine $R_{200}$ as discussed in the text and given in Table 3.1. The dashed line shows the radial profile for X-ray detected clusters discussed in Section 3.6. Poisson error bars are only given on the most and least massive profile to avoid overcrowding. (Bottom) The eleven radial profiles scaled by the derived $R_{200}$ values given in Table 3.1.
(1.49 ± 0.04) \times 10^{-2} h^3 \text{Mpc}^{-3}, M_* - 5 \log_{10} h = -20.44 \pm 0.01, \text{and } \alpha = -1.05 \pm 0.01 \text{ (Blanton et al. 2003b).}

We take into account differences in the \( h \) values and must convert our limit of completeness, which is \( M_r = -21.2 \) from \( h = 0.7 \) as it is in VAC (see Section 1.7), to \( h = 1 \), in order to be consistent. We use the equation,

\[
M^{h=1} + 5 \log_{10}(h) = M^{h=0.7}. \tag{3.2}
\]

In this way, we obtain \(-21.2 + 0.8 = -20.42\) as the limit of completeness for our sample. Also we take into account differences in the passbands, i.e., Blanton et al. use rest-frame luminosities in the \( 0.1r \) band, while we use \( 0.0r \) here (i.e., the band has been k-corrected for \( z = 0.0 \)). The transformation between \( 0.0r \) and \( 0.1r \) is given by,

\[
0.0r = 0.1r - 0.2217 - 0.3943[0.1(r - i) - 0.4084], \tag{3.3}
\]

providing a corrected \( M_* \) of \(-20.66\) in the Schechter function of Blanton et al. This gives a mean space density of field galaxies in the \( 0.0r \) band of \( 0.0045 h^3 \text{Mpc}^{-3} \), with a 2% error. In order to derive the error on the value of the mean space density of field galaxies, we generate random numbers for 10000 cases of \( \alpha, M_* \) and \( \phi_* \) and use the Blanton et al. (2003b) luminosity function to calculate the mean space density of field galaxies in these 10000 sets. Then by calculating the sigma of the Gaussian distribution obtained, we find a 2% error on the value of the mean space density of field galaxies. So to compute the error on the values of \( R_{200} \) obtained, we derive the radial profiles with the mean space density of field galaxies equal to \( 0.0045 h^3 \text{Mpc}^{-3} \pm 2\% \), and in the same way as we measured the \( R_{200} \) described in the introduction to Section 3, we compute the errors. We
estimate an error of 0.01 in the \( R_{200} \) due to this 2% error in the mean field space density. As another source of error, we estimate the uncertainties due to the fiber collision effect. To measure this error, we derive the radial profiles by removing the galaxies that had been originally added due to the fiber collision correction, and in the same way as we measured the \( R_{200} \) described in the introduction to Section 3, we compute the errors. In this way, we estimate an error of 0.03 in the \( R_{200} \) values. Adding the errors in quadrature, we find a total error of 0.03 for our \( R_{200} \) values.

Secondly, we also determined the mean space density of field galaxies directly from the DR5 spectroscopic data by simply counting the number of galaxies in the redshift range \( z = 0.03 \) to \( z = 0.13 \) (as used for our cluster sample), and dividing by the cosmological volume (with an areal coverage of 5740 deg\(^2\) for DR5):

\[
N = 154997 \quad \text{is the number of galaxies,}
\]

\[
S = \text{total sky area} = (180/\pi)^2 4\pi,
\]

\[
F = \text{the fraction of sky covered} = 5740/S,
\]

\[
V_1 = \text{volume of the sphere in redshift 0.13},
\]

\[
V_2 = \text{volume of the sphere in redshift 0.03},
\]

\[
V = (V_1 - V_2) \times F,
\]

and finally \( N/V = 0.0048 h^3 \text{Mpc}^{-3} \).

This gave 0.0048 \( h^3 \text{Mpc}^{-3} \), in excellent agreement with the value derived from integrating the luminosity function (0.0045 \( h^3 \text{Mpc}^{-3} \pm 2\% \)) and consistent with our hypothesis that we have a volume–limited sample. We use the first of these estimates for the mean space density of the field when estimating \( R_{200} \).

In Figure 3.2, we show the same eleven radial profiles, but now scaled by their derived \( R_{200} \) in Table 3.1. As expected, the profiles are in excellent agreement and demonstrate that such scaling to a “universal” profile does work and is more
Table 3.1: The bins used to determine the radial profiles of C4 clusters as a function of mass and thus to determine $R_{200}$ for these stacked clusters. We estimate an error of 0.01 in the $R_{200}$ due to the 2% error in the mean field space density (see text), and also we estimate an error of 0.03 in the $R_{200}$ values due to the fiber collision effect (see Section 3.1.3). Adding the errors in quadrature, we find a total error of 0.03 for our $R_{200}$ values.

<table>
<thead>
<tr>
<th>Mass range $(10^{13} M_\odot)$</th>
<th>No. clus</th>
<th>No. gal</th>
<th>$R_{200}$ (Mpc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01 - 1.7</td>
<td>182</td>
<td>6330</td>
<td>0.41</td>
</tr>
<tr>
<td>1.7 - 2.9</td>
<td>189</td>
<td>9311</td>
<td>0.54</td>
</tr>
<tr>
<td>2.9 - 4.1</td>
<td>189</td>
<td>9800</td>
<td>0.64</td>
</tr>
<tr>
<td>4.1 - 6.2</td>
<td>180</td>
<td>9732</td>
<td>0.68</td>
</tr>
<tr>
<td>6.2 - 7.75</td>
<td>118</td>
<td>6589</td>
<td>0.74</td>
</tr>
<tr>
<td>7.75 - 9.7</td>
<td>120</td>
<td>7761</td>
<td>0.79</td>
</tr>
<tr>
<td>9.7 - 12.5</td>
<td>120</td>
<td>6604</td>
<td>0.84</td>
</tr>
<tr>
<td>12.5 - 15.4</td>
<td>89</td>
<td>5738</td>
<td>0.91</td>
</tr>
<tr>
<td>15.4 - 20.8</td>
<td>91</td>
<td>7241</td>
<td>0.95</td>
</tr>
<tr>
<td>20.8 - 29.5</td>
<td>87</td>
<td>6565</td>
<td>1.04</td>
</tr>
<tr>
<td>29.5 - 200</td>
<td>87</td>
<td>8257</td>
<td>1.16</td>
</tr>
</tbody>
</table>

physically meaningful than using fixed metric scaling relationships.

### 3.2 Choosing cluster mass bins to compute the galaxy density profile

As explained in the introduction to Section 3 for the C4 clusters we compute $R_{200}$, as a function of cluster mass, from stacked cluster profiles. We wish to see the sensitivity of the results to the binning. In detail, firstly we computed the galaxy density profile from adding together all cluster galaxies in four different mass bins for the host clusters, $10^{11} < M_{200} < 2.9 \times 10^{13} M_\odot$, $2.9 \times 10^{13} < M_{200} < 6.1 \times 10^{13} M_\odot$, $6.1 \times 10^{13} < M_{200} < 1.2 \times 10^{14} M_\odot$, $1.2 \times 10^{14} < M_{200} < 2 \times 10^{15} M_\odot$. These bins were chosen to contain approximately equal numbers of clusters ($\sim 400$ each). Note that the formalism is the same for the stacked radial density profiles.
of galaxies in C4 clusters of the eleven different bins of mass, which has been described in the introduction to Section 3; the only difference is the number of mass bins, which was four instead of eleven. Figure 3.3 shows these stacked radial density profiles of galaxies of the four different bins of mass. Then we decided to measure the galaxy density profile from adding all cluster galaxies in a series of eleven cluster mass bins instead of four. As explained in the introduction to Section 3, these bins were chosen to contain approximately equal numbers of galaxies and to obtain the density profiles with the same signal-to-noise (see Table 3.1). Then we split the bins again into 22 bins of mass, since we were interested to see how this distribution compares to the case where we have 11 bins of mass, and at what point the noise would dominate. We took each of the eleven bins and split it by two again, and we constructed the stacked radial density profiles of galaxies of the 22 different bins of mass (shown in Figure 3.4). However we decided to do our analysis with the eleven bins of mass described in the introduction to Section 3 and not to consider the 22 mass bins, since 22 is a large number of bins and the difference in mass of each bin is small. The errors on the mass of the cluster are comparable to the size of the bin, and the errors between bins are large and therefore we are not gaining anything by splitting.

3.3 Testing $R_{200}$

The derivation of mass and size is the main part of our work and their accuracy is important. In this chapter we test extensively the values of $R_{200}$ derived with our method, to make sure that they are correct, since this is crucial for our work and all the rest of the analysis is based on the value of $R_{200}$. Our results for the virial radii were different and smaller than those in the literature. We examine and investigate the possible sources of error.
Figure 3.3: The stacked radial density profiles of galaxies in C4 clusters of the four different bins of mass given in Section 3.2. In the analysis, we have excluded C4 clusters with less than 10 galaxies. These profiles were used to determine $R_{200}$ as discussed in the text (see Section 3.2) and the value of $R_{200}$ for each of the four bins of mass is shown in this figure. The red line shows the radial profile for X-ray detected clusters discussed in Section 3.6. Poisson error bars are given as well.
Figure 3.4: The same as Figures 3.3 and 3.4, but for the 22 different bins of mass given in Section 3.2. The Poisson error bars are only given on the most and least massive profile to avoid overcrowding.
3.3.1 The effect of cluster centroids

In this section, we check the effect of different cluster centroids on our radial profiles and thus our measured $R_{200}$ values. First, we checked the effect of using different cluster centroids provided by the C4 catalogue, namely BCG, MEAN and GEOM (described in Section 2.2). We find no significant differences in our $R_{200}$ values calculated using these different cluster centroids, and have therefore used the BCG centroid throughout. We have also checked our results by restricting our cluster sample to only clusters with an improved BCG assignment, as determined by von der Linden et al. (2007). Again, we found no significant difference in our measured $R_{200}$ values as a function of cluster mass.

3.3.2 Checking the colour magnitude fit

As explained in Section 3.1.2, we applied a colour cut to the SDSS galaxies to ensure they are authorised cluster members and reduce the contamination from interlopers. This is an important constraint which we apply since this cut follows the principle of the C4 algorithm that early–type galaxies within clusters are co-evolving. We visually checked the fits to all the red sequences of all C4 clusters used herein and find, at most, 4% of the fits are potentially in error (see Figure 3.7 as an example of this case). Figures 3.8 and 3.9 show good examples of a tight “red sequence”. A few of them have no passive galaxies in the red sequence and this must be an error with the line fitting, $(r - i)$ versus $r$ (see Figure 3.6 as an example of this case). By looking at the $(g - r)$ vs. $r$ colour-magnitude diagrams, we see that usually the one sigma error on the $(g - r)$ is too large (see Figure 3.5 as an example of this case). All these 5 figures are taken from the C4 webpage. We make no correction for these errors as they are small and, at worst, have the same impact as if no colour cut was applied to these clusters.
Figure 3.5: This figure shows the $(g - r)$ vs. $r$ colour-magnitude diagram for a cluster (ID: 3624) at redshift 0.126 of C4–DR5 catalogue (taken from C4 webpage http://www.ctio.noao.edu/~chrism/current/research/C4/dr5/view_table.html). Red represents passive, green denotes starforming, yellow means C4 galaxy, blue is the BCG (Brightest Cluster Galaxy). Black filled circles are every spectroscopic galaxy projected onto the sky. Yellow filled circles are C4 galaxies within 1000kpc of the cluster centre. Passive galaxies are defined to be galaxies with $H\alpha$ equivalent widths $< 4$ angstroms and are within 1000kpc of the cluster centre. Starforming galaxies are defined to be galaxies with $H\alpha$ equivalent widths $> 4$ angstroms and are within 1000kpc of the cluster centre. Blue Cross is the Brightest Cluster Galaxy. The dotted lines are the upper and lower 1 sigma limits to the fit to the galaxies with $M(u) - M(r) \geq 1.8$ (k-corrected absolute magnitudes with $h = 1$, see Blanton et al. 2003a). Therefore a “good” cluster is the one with a good fit and a “bad” cluster is the one with a bad fit.
Figure 3.6: The panel and the symbols are the same as in Figure 3.5. This figure shows the \((r - i)\) vs. \(r\) colour-magnitude diagram for a cluster (ID: 2038) at redshift 0.129 of C4–DR5 catalogue; which has no passive galaxy in the red sequence, an error in line fitting.
Figure 3.7: The panel and the symbols are the same as in Figure 3.5. This figure shows the \((r - i)\) vs. \(r\) colour-magnitude diagram for a cluster (ID: 3909) at redshift 0.074 of C4–DR5 catalogue; we see an error in line fitting.
Figure 3.8: The panel and the symbols are the same as in Figure 3.5. This figure shows the $(r - i)$ vs. $r$ colour-magnitude diagram for a cluster (ID: 3010) at redshift 0.080 of C4–DR5 catalogue. This is a “good” cluster; we see a tight “red sequence”.
Figure 3.9: The panel and the symbols are the same as in Figure 3.5. This figure shows the $(r - i)$ vs. $r$ colour-magnitude diagram for a cluster (ID: 1033) at redshift 0.085 of C4–DR5 catalogue. This is a “good” cluster; we see a tight “red sequence”.
3.4 Colour and magnitude constraints

As discussed in Section 3.1.2 we have applied a tight colour constraint on the selection of cluster galaxies to help decrease the effects of interlopers on our radial profiles, especially far from the cluster cores. To evaluate the impact of this colour constraint on the radial profiles, we relaxed the present $\pm 2\sigma$ constraint on the $(r-i)$ colour (or $\pm 2\sigma_{width}$ above) to $\pm 4\sigma$ and $\pm 6\sigma$ (the latter is equivalent to having no colour constraint), and found that the $R_{200}$ values increased by between 10% and 20%. By looking at the results (see Table 3.4), we notice that relaxing this colour constraint leads to much larger $R_{200}$ values. As expected, the weaker colour constraints have larger $R_{200}$ values as they have allowed in more (bluer) field galaxies, especially at large radii from the cluster cores.

We also investigated the effect of changing the limiting absolute magnitude on our $R_{200}$ values. For example, we increased the limit to $M_{0,r} = -21.7$ (i.e., the absolute magnitude in r–band which has been k–corrected for $z = 0.0$ instead of $z = 0.1$), known as $L^u$ below, and decreased it to $M_{0,r} = -20.7$ (known as $L^d$ below). In both cases, we also changed our estimate of the mean space density of field galaxies appropriately, i.e., $0.0084$ and $0.0021 h^3 \text{Mpc}^{-3}$ for the lower and upper limit respectively. We also changed the redshift ranges appropriately (in order to be complete, see Figure 3.1), which is $0.03 \leq z \leq 0.15$ (121,023 galaxies) for $L^u$, and $0.03 \leq z \leq 0.10$ (123,955 galaxies) for $L^d$. The differences in $R_{200}$ were small and less severe than the effect of changing the colour cuts. At most, the $R_{200}$ changed by between 5% and 10%, and in many cases, the changes were consistent with the observed statistical errors.
Table 3.2: The values of $R_{200}$ in Mpc (assuming $H_0 = 100\ h\ km\ s^{-1}\ Mpc^{-1}$, $h = 1$) for 11 bins of mass (see Table 3.1) and 3 different limiting absolute magnitude in r-band ($M_r$) relaxing the colour constraint. We estimate an error of 0.01 in the $R_{200}$ due to the 2% error in the mean field space density (see text), and also we estimate an error of 0.03 in the $R_{200}$ values due to the fiber collision effect (see Section 3.1.3). Adding the errors in quadrature, we find a total error of 0.03 for our $R_{200}$ values.

<table>
<thead>
<tr>
<th>Mass range $\left(10^{13} M_\odot\right)$</th>
<th>$R_{200}$ (Mpc) for $M_r = -20.7$ and $\pm 2\sigma$</th>
<th>$R_{200}$ (Mpc) for $M_r = -21.2$ and $\pm 2\sigma$</th>
<th>$R_{200}$ (Mpc) for $M_r = -21.7$ and $\pm 2\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01 - 1.7</td>
<td>0.38 $\pm 2\sigma$ 0.41 $\pm 4\sigma$ 0.43 $\pm 6\sigma$</td>
<td>0.41 $\pm 2\sigma$ 0.44 $\pm 4\sigma$ 0.46 $\pm 6\sigma$</td>
<td>0.42 $\pm 2\sigma$ 0.44 $\pm 4\sigma$ 0.46 $\pm 6\sigma$</td>
</tr>
<tr>
<td>1.7 - 2.9</td>
<td>0.46 $\pm 2\sigma$ 0.51 $\pm 4\sigma$ 0.53 $\pm 6\sigma$</td>
<td>0.54 $\pm 2\sigma$ 0.58 $\pm 4\sigma$ 0.61 $\pm 6\sigma$</td>
<td>0.57 $\pm 2\sigma$ 0.60 $\pm 4\sigma$ 0.61 $\pm 6\sigma$</td>
</tr>
<tr>
<td>2.9 - 4.1</td>
<td>0.56 $\pm 2\sigma$ 0.63 $\pm 4\sigma$ 0.65 $\pm 6\sigma$</td>
<td>0.64 $\pm 2\sigma$ 0.72 $\pm 4\sigma$ 0.74 $\pm 6\sigma$</td>
<td>0.70 $\pm 2\sigma$ 0.76 $\pm 4\sigma$ 0.78 $\pm 6\sigma$</td>
</tr>
<tr>
<td>4.1 - 6.2</td>
<td>0.58 $\pm 2\sigma$ 0.69 $\pm 4\sigma$ 0.72 $\pm 6\sigma$</td>
<td>0.68 $\pm 2\sigma$ 0.79 $\pm 4\sigma$ 0.82 $\pm 6\sigma$</td>
<td>0.74 $\pm 2\sigma$ 0.85 $\pm 4\sigma$ 0.88 $\pm 6\sigma$</td>
</tr>
<tr>
<td>6.2 - 7.75</td>
<td>0.62 $\pm 2\sigma$ 0.75 $\pm 4\sigma$ 0.77 $\pm 6\sigma$</td>
<td>0.74 $\pm 2\sigma$ 0.87 $\pm 4\sigma$ 0.89 $\pm 6\sigma$</td>
<td>0.82 $\pm 2\sigma$ 0.95 $\pm 4\sigma$ 0.97 $\pm 6\sigma$</td>
</tr>
<tr>
<td>7.75 - 9.7</td>
<td>0.65 $\pm 2\sigma$ 0.81 $\pm 4\sigma$ 0.84 $\pm 6\sigma$</td>
<td>0.79 $\pm 2\sigma$ 0.95 $\pm 4\sigma$ 0.98 $\pm 6\sigma$</td>
<td>0.87 $\pm 2\sigma$ 1.02 $\pm 4\sigma$ 1.05 $\pm 6\sigma$</td>
</tr>
<tr>
<td>9.7 - 12.5</td>
<td>0.70 $\pm 2\sigma$ 0.88 $\pm 4\sigma$ 0.91 $\pm 6\sigma$</td>
<td>0.84 $\pm 2\sigma$ 1.03 $\pm 4\sigma$ 1.08 $\pm 6\sigma$</td>
<td>0.94 $\pm 2\sigma$ 1.12 $\pm 4\sigma$ 1.14 $\pm 6\sigma$</td>
</tr>
<tr>
<td>12.5 - 15.4</td>
<td>0.75 $\pm 2\sigma$ 0.98 $\pm 4\sigma$ 1.03 $\pm 6\sigma$</td>
<td>0.91 $\pm 2\sigma$ 1.14 $\pm 4\sigma$ 1.18 $\pm 6\sigma$</td>
<td>1.02 $\pm 2\sigma$ 1.23 $\pm 4\sigma$ 1.24 $\pm 6\sigma$</td>
</tr>
<tr>
<td>15.4 - 20.8</td>
<td>0.76 $\pm 2\sigma$ 1.02 $\pm 4\sigma$ 1.04 $\pm 6\sigma$</td>
<td>0.95 $\pm 2\sigma$ 1.22 $\pm 4\sigma$ 1.25 $\pm 6\sigma$</td>
<td>1.07 $\pm 2\sigma$ 1.32 $\pm 4\sigma$ 1.35 $\pm 6\sigma$</td>
</tr>
<tr>
<td>20.8 - 29.5</td>
<td>0.83 $\pm 2\sigma$ 1.14 $\pm 4\sigma$ 1.18 $\pm 6\sigma$</td>
<td>1.04 $\pm 2\sigma$ 1.34 $\pm 4\sigma$ 1.36 $\pm 6\sigma$</td>
<td>1.16 $\pm 2\sigma$ 1.45 $\pm 4\sigma$ 1.47 $\pm 6\sigma$</td>
</tr>
<tr>
<td>29.5 - 200</td>
<td>0.91 $\pm 2\sigma$ 1.26 $\pm 4\sigma$ 1.29 $\pm 6\sigma$</td>
<td>1.16 $\pm 2\sigma$ 1.51 $\pm 4\sigma$ 1.55 $\pm 6\sigma$</td>
<td>1.29 $\pm 2\sigma$ 1.62 $\pm 4\sigma$ 1.66 $\pm 6\sigma$</td>
</tr>
</tbody>
</table>
3.5 Comparison with other results

We compare our $R_{200}$ values against similar measurements in the literature. Johnston et al. (2007) have recently derived $R_{200}$, as a function of cluster mass, using the stacked weak lensing signal around clusters of galaxies in the SDSS maxBCG sample (Koester et al. 2007). In Figure 3.10 we compare our measurements with those of Johnston et al., and find excellent agreement between the two measurements of $R_{200}$ over nearly two orders of magnitude in halo mass. At low masses ($< 10^{13} M_\odot$), we do find larger $R_{200}$, which may indicate some interloper contamination on our part. We have calculated the HOD for C4 clusters (see Chapter 4) using the Johnston et al. relationship between $R_{200}$ and mass, and find very similar HOD results to those using our own relation. These agreements are encouraging, given the different methodologies and catalogues used to measure both the mass and radius of these clusters, and this gives weight to the concept of using integrated optical profiles to measure these cluster properties.

We have also compared our $R_{200}$ measurements to a commonly used analytical relation (Finn et al. 2004),

$$R_{200} = 2.47 \frac{\sigma_x}{1000\text{km/s}} \frac{1}{\sqrt{\Omega_\Lambda + \Omega_m (1 + z)^3}} h_{70}^{-1}\text{Mpc},$$

(3.4)

where $\sigma_x$ is the line-of-sight velocity dispersion and $h_{70} = H_0/(70\text{km/s/Mpc})$. Figure 3.11 shows our values for $R_{200}$ compared to the $R_{200}$ values obtained from Eq. (3.4). There is a good correlation between the two $R_{200}$ values for a majority of clusters, but there is considerable scatter off the expected one-to-one line for a significant fraction of clusters, with Eq. (3.4) giving much larger values for $R_{200}$ than we measure. In Figure 3.11, we show (star symbols) all C4 clusters that have a non-zero SCF which Miller et al. (2005) uses to signify clusters embedded
To compute the error on the values of $R_{200}$ obtained, we derive the radial profiles with the mean space density of field galaxies equal to $0.0045 \, h^3 \, \text{Mpc}^{-3} \pm 2\%$ and in the same way as we measured the $R_{200}$ described in the introduction to Section 3, we compute the errors. We estimate an error of 0.01 in the $R_{200}$ due to 2% error in the mean field space density (see Section 3.1.3 for the derivation of the error in the mean field space density). As another source of error, we estimate the uncertainties due to the fiber collision effect. To measure this error, we derive the radial profiles by removing the galaxies that had been originally added due to the fiber collision correction, and in the same way as we measured the $R_{200}$ described in the introduction to Section 3, we compute the errors. In this way, we estimate an error of 0.03 in the $R_{200}$ values. Adding the errors in quadrature, we find a total error of 0.03 for our $R_{200}$ values.
CHAPTER 3. DETERMINATION OF $R_{200}$

Figure 3.11: Comparison of our $R_{200}$ values with those estimated using Eq. (3.4). The star symbols show C4–DR3 clusters labeled as having a non-zero SCF and therefore have a high velocity dispersion (see text).

in complex large-scale structures or exhibiting signs of substructure. Therefore, the high velocity dispersion of these clusters should be used with caution and it appears that a majority of the differences between Eq. (3.4) and our $R_{200}$ values is due to this issue. It is reassuring that our $R_{200}$ measurements do not suffer from this problem and can be trusted even for clusters with complex velocity structure.

3.6 $R_{200}$ for X-ray detections

As a final confirmation of our $R_{200}$ values, we have determined the stacked radial profile for X-ray detected clusters in the C4 cluster catalogue. We matched the NORAS (The Northern ROSAT All-sky) Galaxy Cluster Survey (Bohringer et al. 2000) to the C4 catalogue using a matching radius of 0.2 $h^{-1}$ Mpc (between
Table 3.3: The values of $R_{200}$ for X-ray detected C4 clusters. The measurements are for different colour constraints. Due to the 2% error on the value of mean space density of field, there is $\sim 0.01$ error on the values of $R_{200}$.

<table>
<thead>
<tr>
<th>±2σ</th>
<th>±4σ</th>
<th>±6σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.94</td>
<td>1.10</td>
<td>1.16</td>
</tr>
</tbody>
</table>

the optical and X-ray centroids) and a constraint of ±2$\sigma_v$ on the difference in redshift between the NORAS and C4 clusters ($\sigma_v$ is the velocity dispersion for the C4 detected cluster). In this way, we find 85 matching clusters between C4 and NORAS and have stacked the radial profiles of these clusters to measure their $R_{200}$ value (using 4953 galaxies). The radial profile for these X-ray detected C4 clusters is shown in Figure 3.2. As expected, the radial profile for these X-ray detected C4 clusters is similar to the profile for the more massive C4 clusters; we find $R_{200} = 0.94$, with the colour cut, or $R_{200} = 1.16$ without. This again confirms that our optical mass estimates are sensible and can be used to sub-divide the catalogue as a function of mass. In Figure 3.12, we have also checked the mass proxy, X-ray luminosity, $L_X$ using the conversion from Popesso et al. (2005),

$$\log(L_X/(10^{44}\text{ergs}^{-1})) = \alpha \times \log(M_{200}) + \beta,$$  

(3.5)

(where $\alpha$ is equal to 0.58±0.03 and $\beta$ is equal to 0.45±0.02) and compared with our $M_{200}$ values derived from the optical luminosity. We found a good correlation with a log–slope of 0.92±0.34 indicating a moderate level of scatter.

3.7 Comparison with Berlind catalogue

As a test of the C4 catalogue, we compare our measurements of $R_{200}$ and some other quantities with those from the Berlind et al. (2006) catalogue, which has
Figure 3.12: In this figure, we have checked the mass proxy, X-ray luminosity, $L_X$ using the conversion from Popesso et al. (2005) and compared with our $M_{200}$ values derived from the optical luminosity. The solid line shows the best fit to the data. The linear fit gives a log–slope of $0.92\pm0.34$. 
identified clusters in volume-limited samples of the SDSS and used a different friends-of-friends algorithm. By downloading all the catalogue of clusters identified by Berlind et al.\(^2\), we try to find all the clusters in common with the ones in the C4 catalogue and compare our estimate of \(N\) (richness), velocity dispersion and especially \(R_{200}\) with their estimates, to find and investigate any correlation between them. The criteria that we use to determine the clusters in common between the C4 and Berlind catalogues is obtained by using a matching radius of 0.1 \(h^{-1}\) Mpc (between the C4 and Berlind centroids) and a constraint of \(\pm 1\sigma_v\) on the difference in redshift between the Berlind and C4 clusters (\(\sigma_v\) is the velocity dispersion). We found 20 clusters in common. After plotting our \(N, \sigma_v, R_{200}\) against those from Berlind et al. we see that there is not a good correlation between them especially between the velocity dispersions (see Figures 3.13 and 3.14 and 3.15). Since they have used the friends-of-friends method, it is possible that they include the other clusters. For all the clusters found in common we put the C4 cluster in the centre and other C4 and Berlind clusters around it within 2 degrees and plot RA versus DEC for all these clusters in common (see Figure 3.16).

From Figure 3.16 and associated plots we see a range of problems. In many cases, it is uncertain which Berlind cluster corresponds to a C4 cluster. By looking at redshifts and velocity dispersions of all these clusters, we investigate how close they are to each other, to determine the best fit (see Figure 3.16 as an example for a C4 cluster with redshift 0.0897 and velocity dispersion equal to 846.00 km/s which is in common with a Berlind cluster with redshift 0.0891 and velocity dispersion equal to 387.8 km/s). We notice that the velocity dispersions in the C4 catalogue are all systematically higher than those in the Berlind catalogue. We need to investigate how the velocity dispersions have been calculated in both

\(^2\)Available from http://cosmo.nyu.edu/aberlind/Groups
catalogues and to see within which radii they have been computed and what aperture has been considered. This is important because the velocity dispersions are used to derive the mass and we need to make sure that they are correct.

For this purpose, we first look at the redshift distribution of the clusters to see if they are correlated and also to find out if they have substructures, by looking at the SCF flag, and we find that most of them have substructures. We found that the C4 catalogue has always considered the aperture to be 1.5 Mpc as a constant radius, while in Berlind’s catalogue the aperture varies for each cluster. We also see that the apertures in the Berlind catalogue are all much smaller than 1.5 Mpc (see Figure 3.14 which corresponds to Figure 3.10). We need another variable and therefore we look at the X-ray detection. For this purpose, we match the NORAS Galaxy Cluster Survey (Bohringer et al. 2000) to both C4 and Berlind catalogues by using the same matching criteria explained in Section 3.6, in order to see which one has X-ray detection. We find $\sim 40$ common clusters. We plot X-ray luminosity versus velocity dispersion to find which catalogue has a better correlation, since more massive clusters give bigger X-ray luminosity and velocity dispersion (see Figures 3.19 and 3.20). We see that the correlation between the X-ray luminosity and velocity dispersion for the C4 catalogue is 2.7 times better than for the Berlind catalogue. This is derived by fitting a line to the data and then calculating the root mean square. The linear fit in Figure 3.19 gives a log–slope of 1.92$\pm$0.46 (for the C4 catalogue), and the linear fit in Figure 3.20 gives a log–slope of -0.16$\pm$0.42 indicating no correlation with X–ray luminosity and velocity dispersion (for the Berlind catalogue). Figure 3.18 shows the relation between X-ray luminosity and optical luminosity for C4 to see the correlation and the scatter around it. Moreover, Figure 3.21 shows the X-ray luminosity versus velocity dispersion derived by Ortiz–Gil et al. (2004), indicating a closer match (a log–slope of $4.1 \pm 0.3$) to the C4 catalogue than to the Berlind catalogue.
Figure 3.13: This figure shows the richness (number of galaxies in the cluster) for Berlind clusters versus those in common with the C4. The Poisson error bars are given.
Figure 3.14: In this figure, which corresponds to Figure 3.13 for the clusters in common between Berlind and C4, we present our derived virial radius versus the radius derived in the Berlind catalogue.
Figure 3.15: In this figure, which corresponds to Figure 3.13 and Figure 3.14 for the clusters in common between Berlind and C4, we present the velocity dispersions of the Berlind catalogue and the C4 catalogue.
Figure 3.16: RA and DEC for the C4 cluster at redshift 0.0897 in the middle, and all the C4 and Berlind surrounding clusters within 2 degrees. The diamond symbols show C4 and the square symbols show the Berlind clusters. The redshifts and velocity dispersions of the clusters in common and very nearby are shown.
3.8 Radial Distribution of Galaxies in Clusters

We study here the galaxy distribution in clusters by their surface density profile. This is important because the galaxy distribution within dark matter halos is an important aspect of the HOD formalism. Different clustering statistics may be the result of different distributions at small scales (Seljak 2000; Peacock & Smith 2000; Berlind & Weinberg 2002).

In Figure 3.22, we present the stacked (projected) surface density profile ($\Sigma(R)$) for all galaxies in C4 clusters, where $R$ is the projected distance from the cluster centre. This was achieved by re-scaling the projected distances ($R$) of cluster galaxies from the cluster centroids by the individual $R_{200}$ values given the mass of the cluster (from the summed optical luminosity, see Section 2.3).

First we calculated the distance from the cluster centre to each galaxy. Secondly,
Figure 3.18: The scaling relationship between X-ray and optical luminosity for the clusters in common between the NORAS and C4 catalogues. The linear fit gives a log–slope of $1.83 \pm 0.62$. 
Figure 3.19: The scaling relationship between X-ray luminosity and velocity dispersion for the clusters in common between the NORAS and C4 catalogues. The linear fit gives a log-slope of 1.92±0.46.
Figure 3.20: The scaling relationship between X-ray luminosity and velocity dispersion for the clusters in common between the NORAS and Berlind catalogues. The red linear fit gives a log-slope of $-0.16 \pm 0.42$ indicating no correlation with X-ray luminosity and velocity dispersion.
Figure 3.21: The scaling relationship between bolometric X-ray luminosity and velocity dispersion for 171 clusters from REFLEX sample. The line has a log-slope of $4.1 \pm 0.3$ and the dashed lines show the 1-sigma errors. Filled circles correspond to the clusters with $z \leq 0.05$, and the open triangles represent the clusters at $0.05 < z \leq 0.1$ and open circles are clusters at the redshift $z > 0.1$. The X-ray luminosities data are in units of $L_x/10^{45}$ erg/s ($L_{45}$) and the velocity dispersions are in units of $\sigma_v/500$ km/s ($\sigma_{500}$). Figure is from Ortiz-Gil et al. (2004).
we divided these distances by the virial radius for each cluster in order to express them in units of \( R_{200} \). Then we stacked our clusters once in four groups of mass (see Table 3.4) and then for the whole sample. The total surface density of galaxies is finally obtained by calculating the number of galaxies in radial bins divided by the surface of each bin. (Note that spherical symmetry has been assumed.) We stress that we follow all the same constraints on the data that have been described in Section 3.1 to build our projected surface density profile (e.g., the redshift range \( 0.03 < z < 0.13 \), both clusters and galaxies and a luminosity range \( -24 < M_r < -21.2 \)).

In Section 3.1.1 we described one problem which is the issue of fibre collision and corresponds to \( R < 0.09 r_{vir} \) for the majority of our clusters for \( z = 0.1 \) (this is redshift dependent). As explained in Section 3.1.1 to correct for fibre collisions we added the missed galaxies (photometric) to the rest of them (spectroscopic sample), and for their redshifts we considered the redshifts of the cluster that they belong to.

### 3.8.1 Profile fitting

We present the best fit NFW profile (Navarro, Frenk and White 1997), one of the most commonly used model profiles for dark matter halos, to the stacked (projected) surface density profile of galaxies in C4 clusters. In clusters, the dark matter is dynamically dominant and therefore it is important to have a good model of the dark-matter mass distribution. The NFW profile is a spatial distribution of dark matter predicted from N-body simulations and has been described as the universal density profile:

\[
\rho(r) = \frac{\rho_s}{r/r_s (1 + r/r_s)^2},
\]
where $r_s$ is a characteristic scale radius, and $\rho_s$ is the amplitude of the profile. However, this formula is not valid for large radii since it predicts unbounded mass, but it may hold out to the virial radius $r_{200}$, which is the radius enclosing a mean overdensity of 200 (Cooray & Sheth 2002). Moreover, Thomas et al. (1998) found that the NFW formula was a good approximation to the mass distribution of an average cluster, but that there was a wide dispersion in the rate at which the density deviated at the virial radius. However, this profile is achieved as the equilibrium configuration of dark matter in simulations of collisionless dark matter particles by numerous groups of scientists (Jing, 2000). Note the important point that before the dark matter virializes, the distribution of dark matter is not expected to match an NFW profile, and significant substructure is observed in simulations while halos collapse. This profile is expressed in terms of the concentration parameter, $c = r_{\text{vir}}/r_s$, the ratio of the virial radius to the
characteristic radius in the NFW formula. The plane-projected surface density of the NFW profile is given by

\[ \Sigma(x) = \frac{2\rho_s r_s}{x^2 - 1} f(x), \quad (3.7) \]

where

\[ f(x) = \begin{cases} 
1 - \frac{2}{\sqrt{x^2-1}} \arctan \sqrt{\frac{x-1}{x+1}} & (x > 1) \\
1 - \frac{2}{\sqrt{1-x^2}} \tanh \sqrt{\frac{1-x}{1+x}} & (x < 1) \\
0 & (x = 1) 
\end{cases} \quad (3.8) \]

and \( x = R/r_s = cR/r_{\text{vir}} \) (Bartelmann 1996). In order to avoid the problem of potential degeneracy between the concentration parameter and the amplitude of the NFW profile discussed by Nagai & Kravtsov (2005), we use an alternative formulation

\[ \Sigma_{\text{gal}}(x) = \frac{c^2 N_{\text{vir}}}{2\pi g(c)(x^2 - 1)} f(x), \quad (3.9) \]

where \( N_{\text{vir}} = \int_0^{r_{\text{vir}}} \rho_{\text{gal}}(r) 4\pi r^2 dr \) is the number of galaxies within the virial radius of the group and \( g(c) = \ln(1 + c) - c/(1 + c) \).

We have fitted Eq. (3.9) to the data in Figure 3.22 but restricted ourselves to the range \( 0.05 < R/r_{\text{vir}} < 0.95 \). The lower limit arises because towards the centre the galaxy surface density declines. The upper limit is because above \( R/r_{\text{vir}} = 0.95 \Sigma_{\text{gal}} \) we are in the field which becomes noise dominated. The best fit is shown in Figure 3.22 with \( c = 3.8 \pm 0.1 \) (one–sigma error marginalised over the probability distribution of the NFW amplitude). We have also fitted the surface density profile of galaxies just in the red sequence (as defined in Section 3.1.2), and find a similar result of \( c = 3.4 \pm 0.1 \). In Figure 3.23, we show the likelihood contours for our joint fit to the NFW concentration parameter and the amplitude for galaxies in the red–sequence, and we see that the two parameters
are nearly independent. We have investigated the sensitivity of the values of the concentration parameter $c$ on the range of $R/r_{\text{vir}}$ over which the fits are made, by determining the value of $c$ when different ranges are considered. A variation of $\sim 10\%$ was observed. This is due to the NFW profile being weakly dependent on the value of $c$.

### 3.8.2 Mass dependence

Finally, we have investigated the mass dependence of the NFW profile fits using four bins of mass (Table 3.4), since the distribution of the cluster concentration is extended. The four mass bins have been chosen to have approximately equal numbers of clusters in each and thus similar errors on the NFW fitted parameters, resulting in a more effective comparison. We calculate the concentration parameter which is the representative of the halo’s central density and depends on the
halo’s time of formation and history. This is important since the concentrations of the massive galaxy clusters which are the most recent bound systems to form, are a probe of the mean density of the universe at late time. We find that $c$ is almost constant across these bins (Table 3.4 and Figure 3.24), which is different from that expected from simulations, where the mean concentration parameter of dark matter halos should decrease with mass as $c = c_0 (M/M_*)^{-\lambda}$, with $\lambda \sim 0.1$ and $c_0 \sim 10$. For the mass range probed here, we expect $c$ to decrease from $\sim 8$ to $\sim 5$ (or nearly 50%), which is clearly not seen. Our results are consistent with Collister & Lahav (2005) and Hansen et al. (2005), who find $c$ is a decreasing function of galaxy occupation number. Figure 3.24 shows the concentration parameter values obtained by Collister & Lahav (2005) together with the concentration parameter values obtained in this thesis. We see that the galaxy radial distribution is significantly less concentrated with respect to the average dark matter halos in simulations performed by Bullock et al. (2001). As mentioned in Section 3.8.1 the effect of close-pair incompleteness, and probably the uncertainty in the position of the cluster centres, and also the deblending effect (when two galaxies are in one image on top of each other and we can not separate them), could cause the measured galaxy light profile to decrease on small scales and thus cause an underestimation of the concentration parameter. Moreover, we expect a number of effects to add to the difficulties of constraining concentration parameters. As discussed by Collister & Lahav (2005), when building the galaxy radial distribution and stacking profiles with a wide range of concentrations, it is difficult to constrain the concentration parameters for the individual groups because of not having enough galaxy members for all the groups. Therefore, we can not know exactly what we expect from the distribution of the concentration parameters for the groups since it is difficult to measure the profile of the individual groups. Therefore, the values obtained, may represent the ‘average’
Table 3.4: The mass dependence of our NFW profile fits using four bins of mass.

<table>
<thead>
<tr>
<th>Mass ($M_{200}$)</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$3.1 \times 10^{13} M_\odot$</td>
<td>$3.4^{+0.25}_{-0.30}$</td>
</tr>
<tr>
<td>$7.3 \times 10^{13} M_\odot$</td>
<td>$3.0^{+0.16}_{-0.17}$</td>
</tr>
<tr>
<td>$1.4 \times 10^{14} M_\odot$</td>
<td>$2.7^{+0.15}_{-0.18}$</td>
</tr>
<tr>
<td>$3.5 \times 10^{14} M_\odot$</td>
<td>$2.9^{+0.17}_{-0.18}$</td>
</tr>
</tbody>
</table>

concentration parameters of the groups. Another effect could be that because the determination of the mass has a large scatter for small groups, this impacts the stacked profiles, and therefore the radial scaling causing the underestimation of the measurement of the concentration parameters.

There have been various measurements of the concentration parameter in the literature, using different methods of deriving the dark matter halo’s mass distribution, such as strong lensing, weak lensing, combined weak lensing and strong lensing, X-ray temperature, line–of–sight velocity dispersion. The concentration which has been determined from lensing methods (weak, strong, combined weak and strong lensing) is systematically higher ($c > 10$) than the concentration which has been determined from other methods like the X-ray concentration parameters ($c < 10$, see Comerford et al. 2007). This could be because in the X-ray method there is an assumption of hydrostatic equilibrium, which is not valid for unrelaxed clusters. This tends to underestimate the total mass of the cluster and therefore to result in a lower value for the concentration parameter.

Also a variety of definitions is used in the literature for the virial radius. The difference of our work resides in both the mass estimation (derived from the optical luminosity, see Section 2.3 for more detail) and the virial radius estimation, $R_{200}$ (see the introduction to Section 3).
Figure 3.24: Variation of group concentration parameter with mass ($M_{200}$) for red galaxies of our C4-SDSS cluster sample (diamond symbols). In addition, we have plotted the concentration parameter values obtained by Collister & Lahav, 2005 (square symbols). The line shows the trend in the mean concentration parameter of dark matter halos in simulations (due to Bullock et al., 2001).
Chapter 4

Halo Occupation Distribution

4.1 HOD derivation

There have been many measurements of the HOD using a variety of statistical properties of galaxies. The most widely used methods are based on the luminosity function, spatial clustering (as a function of galaxy luminosity and colour) and the abundance of clusters of galaxies (Peacock & Smith 2000, Magliocchetti & Porciani 2003, Scranton 2003, van den Bosch et al. 2003, Zehavi et al. 2004, 2005, Yang et al. 2005a,b,c, Blake, Collister & Lahav 2008, Kulkarni et al. 2007, Rozo et al. 2007, Yang et al. 2008). As an example, Magliocchetti & Porciani have used the correlation function and the luminosity function of the galaxies (taken from 2dFGRS, Colless et al. 2001, and in the redshift range $0.01 < z < 0.15$) in order to examine the HOD.

However, the most direct method of determining the HOD is by counting galaxies within known dark matter halos, e.g., by counting the number of galaxies in known clusters of galaxies as a function of the cluster mass. This direct approach has been followed by several authors (Kochanek et al. 2003, Lin, Mohr & Stanford 2004, Collister & Lahav 2005, Popesso et al. 2006, Ho et al. 2007),
using a variety of different cluster catalogues at both high and low redshift. Unfortunately, there is significant disagreement between these results, especially in the slope of the relationship \((\beta)\) between mass and number of galaxies \((N \propto M^\beta)\), see Section 1.5.1. For example, Kochanek et al. (2003) find \(\beta = 1.1 \pm 0.09\), based on a sample of nearby optical/IR clusters detected objectively in the Two Micron All-Sky Survey (2MASS), while Lin et al. (2004) report \(\beta = 0.84 \pm 0.04\), also using data from 2MASS, but focusing on only X-ray detected clusters (see also Popesso et al. 2006, who find a consistent result, within 2 sigma).

It is surprising that different results arise for the slope of the HOD, given how straightforward the counting of galaxies in clusters should be. Clearly there are several systematic problems that plague this measurement, including projection effects (including both spurious clusters and groups of galaxies as well as galaxy interlopers), mass estimates and the definition of the size of the clusters and groups. As stressed by Kochaneck et al., the systematic errors are dominant and probably account for an error of 0.05 in the slope of the HOD, which would bring the different results above into closer agreement. Theoretically, the latest simulations suggest that the slope of the HOD, in the high mass end, should be close to unity and be steepest for the oldest galaxies in the most massive clusters (Kravtsov et al. 2004, Zheng et al. 2005), and that the scatter around this best fit relation should be Poisson.

We provide here a brief overview of the halo model (see Cooray & Sheth 2002 for a more detailed explanation of the halo model formalism). In the halo model, the distribution of mass in the Universe can be understood as the combination of the mass within every dark matter halo and the spatial distribution of the halos. This formalism thus provides a way to describe the spatial statistics of the dark matter density field from the linear to non-linear regimes. If we also know the probability distribution for the number of galaxies present in each halo, then we
can write the galaxy power spectra as:

$$P_{\text{gal}}(k) = P_{\text{gal}}^{(1h)}(k) + P_{\text{gal}}^{(2h)}(k),$$  \hspace{1cm} (4.1)

where $P_{\text{gal}}^{(1h)}(k)$ is the power spectrum of the (non-linear) intra-halo galaxy clustering, known as the one–halo term (1$h$), and $P_{\text{gal}}^{(2h)}(k)$ is the (linear) power spectrum for the inter-halo clustering, or the two–halo term (2$h$). These two spectra can be written in terms of the probability distribution of galaxies in halos, or $P(N|M)$ the Halo Occupation Distribution (HOD), as

$$P_{\text{gal}}^{(1h)}(k) = \int dM \, n(M) \frac{\langle N(N - 1)|M \rangle}{\bar{n}_{\text{gal}}^2} |\hat{u}_{\text{gal}}(k|M)|^2,$$  \hspace{1cm} (4.2)

and

$$P_{\text{gal}}^{(2h)}(k) = P_{\text{dm}}^{(\text{lin})}(k) \left[ \int dM n(M) b(M) \frac{\langle N|M \rangle}{\bar{n}_{\text{gal}}} \hat{u}_{\text{gal}}(k|M) \right]^2,$$  \hspace{1cm} (4.3)

where $n(M)$ is the halo mass function, $b(M)$ is the halo biasing factor, $\langle N|M \rangle$ and $\langle N(N - 1)|M \rangle$ are the first and second factorial moments of the HOD, $\bar{n}_{\text{gal}}$ is the average number density of galaxies, and $\hat{u}_{\text{gal}}(k|M)$ is the Fourier transform of the normalised radial distribution of galaxies within halos. The integrals are over the halo mass, $M$ (see Collister & Lahav 2005 for more detail). Therefore, an accurate understanding of the HOD is required to fully exploit the power of the halo model. At present, this can only be achieved through observations and the most direct method of constraining the HOD is through counting galaxies in known clusters and groups of galaxies. We revisit the computation of the HOD of galaxies using clusters and groups of galaxies from the SDSS–C4 cluster catalogue (Miller et al. 2005). The main difference in our work, compared to these other direct HOD estimates, is that we have attempted to minimise the number of assumptions and extrapolations required, thus providing as clean a
measurement of the HOD as possible. This is possible because of the size of the C4 catalogue and its purity, as we are confident that $\geq 95\%$ of all clusters in the catalogue are true over-densities (see Miller et al. 2005). Therefore, we have attempted to minimise the systematic uncertainties (see Kochanek et al. 2003) in such an estimate, while still having a large enough sample to control the statistical errors. In Figure 4.1 we present the Halo Occupation Distribution (HOD), or the measured number of galaxies per cluster as a function of the cluster mass, i.e., $N(M)$. Cluster membership was defined in Section 3.1. In summary, we used the $R_{200}$-mass relation determined for our eleven bins of cluster mass to interpolate individual values of $R_{200}$ for each cluster, based on the mass of that cluster (from the summed optical luminosity, Section 2.3). We counted the galaxies within $R_{200}$ and brighter than the chosen limiting magnitude range ($-24 \leq 0.0r \leq -21.2$), initially without any colour cuts.

As discussed in Section 1.5.1 at high masses, in the HOD model, the mean number of galaxies per cluster as a function of mass is predicted to be a power-law of the form, $\langle N|M \rangle \sim (M/M_0)^\beta$, with $M_0$ and $\beta$ as free parameters (Berlind & Weinberg 2002). Therefore, we fit this model to our $N(M)$ data for the individual clusters to determine the best fit values of $M_0$ and $\beta$. The details of this fitting are given in Section 4.1.1, with our results summarised in Table 4.1 where the errors are just statistical (as defined in Section 4.1.1) and do not account for uncertainties in the mass measurements.

In Table 4.1, we present a series of our best–fit $\beta$ values as a function of the limiting luminosity ($L^u$, $L^d$), halo mass range and colour of galaxies (all, “red”, “blue”). First, it is clear that the quoted statistical errors on these $\beta$ fits are significantly smaller than the systematic uncertainties discussed in Section 3.3.

\footnote{Also interactive data language (IDL) routines, such as CURVEFIT, LINFIT, and ROBUST-LINEFIT, give essentially the same results as the Monte Carlo technique described in Section 4.1.1}
Even though the Poisson error on any individual measurement of \(N(M)\) is large, the errors on \(\beta\) are greatly reduced because we are averaging over thousands of clusters. In the literature (e.g., Collister & Lahav 2005), the errors calculated on these \(\beta\) fits are larger than the errors obtained by us. This is because they have not fitted the HOD model described above to the data for the individual clusters. Instead, they have binned the data in 10 different bins of masses and then fitted the model to these 10 average points. To illustrate this fact, we present in Table 4.2 the effect on the HOD \(\beta\) value from our approach of assigning to photometric galaxies, with no available SDSS redshift, the redshift of their host cluster (see Section 3.1.1). In this table we show the best-fit \(\beta\) values for two subsets of C4 clusters split by the percentage of photometric galaxies added to these clusters; namely clusters above and below the median percentage of added galaxies of 40%. As can be seen, when considering all galaxies (regardless of their color), the slope is systematically larger (by \(\sim 0.05\)) for clusters with the biggest correction. If we now add a color cut (only “red” galaxies), the disagreement disappears, as the number of interlopers is reduced because the galaxies must now have the same color as the cluster red sequence. In summary, the systematic errors are significantly larger than the quoted statistical errors on our \(\beta\) values, and we would conservatively suggest assuming a systematic error of 0.1 on any of our quoted \(\beta\) values (this systematic uncertainty is similar to Kochanek et al. 2003, who estimated the systematic errors on the slope of the HOD were between 0.05 and 0.1). The results of this chapter are partly based on Torki et al. (2008).

### 4.1.1 Calculating the best fit HOD parameters

We present here the details of our HOD parameter fitting and the errors on those parameters (\(\beta\) and \(\log(M_0/h^{-1}M_\odot)\)). We scanned this two dimensional parameter...
Table 4.1: Our best fit HOD $\beta$ parameter. The label “A” refers to “All” which describes our selection without a colour cut within the virial radius of the clusters, while “R” refers to “red” and “B” refers to “blue” which means galaxies within the $\pm 2\sigma (r - i)$ limit on the host cluster red sequence (see text) and bluer than this cut, respectively. We also consider three luminosity bins. The first is the standard luminosity limit of $M_{\text{0r}} = -21.2$ used throughout this work (labeled L in the table), and $L^u$ and $L^d$ are as discussed in Section 3.4. We also investigate two lower limits of the mass range used in fitting the HOD parameters, namely $M_1 = 5 \times 10^{13} h^{-1} M_\odot$ and $M_2 = 10^{14} h^{-1} M_\odot$.

<table>
<thead>
<tr>
<th></th>
<th>$L^u$</th>
<th>$L$</th>
<th>$L^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M &gt; M_1$</td>
<td>$M &gt; M_2$</td>
<td>$M &gt; M_1$</td>
<td>$M &gt; M_2$</td>
</tr>
<tr>
<td>A</td>
<td>0.79 ± 0.01</td>
<td>0.79 ± 0.02</td>
<td>0.92 ± 0.01</td>
</tr>
<tr>
<td>R</td>
<td>0.61 ± 0.02</td>
<td>0.58^{+0.02}_{-0.03}</td>
<td>0.65 ± 0.01</td>
</tr>
<tr>
<td>B</td>
<td>0.48^{+0.06}_{-0.07}</td>
<td>0.50^{+0.09}_{-0.08}</td>
<td>0.76 ± 0.04</td>
</tr>
</tbody>
</table>

Table 4.2: The effect of adding photometric galaxies (to correct for fibre collisions) on our fitted HOD $\beta$ parameter. We have split the cluster sample above and below the median percentage correction to our clusters. We present fits for two different lower limits on the halo mass range used: $M_1 = 5 \times 10^{13} h^{-1} M_\odot$ and $M_2 = 10^{14} h^{-1} M_\odot$.

<table>
<thead>
<tr>
<th></th>
<th>Above Median</th>
<th>Below Median</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M &gt; M_1$</td>
<td>$M &gt; M_2$</td>
</tr>
<tr>
<td>All galaxies</td>
<td>0.93^{+0.04}_{-0.02}</td>
<td>0.96 ± 0.03</td>
</tr>
<tr>
<td>Red galaxies</td>
<td>0.62 ± 0.02</td>
<td>0.58 ± 0.02</td>
</tr>
</tbody>
</table>
Figure 4.1: The number of galaxies (no colour constraints) within $R_{200}$ in C4 clusters as a function of the C4 cluster mass ($M_{200}$). Galaxy counts shown here are for absolute magnitudes of $M_r < -21.2$. The solid line is the best fit HOD model (see Section 4.1.1) for masses $\geq 10^{14} h^{-1} M_\odot$, while the dashed line is for $\geq 5 \times 10^{13} h^{-1} M_\odot$. 
space using the Markov chain Monte Carlo (MCMC) method with the Metropolis algorithm. For each pair of $\beta$ and $\log(M_0/h^{-1}M_\odot)$ values, we assume a theoretical model of $\mu(M) = \langle N|M \rangle = (M/M_0)^\beta$, and assume that the observed data are Poisson scattered around the mean. Therefore, the likelihood of observing the data ($\{N_{i\text{obs}}\}$) for each given value of $\beta$ and $\log(M_0/h^{-1}M_\odot)$ is

$$\mathcal{L} \left( \{x_i\} \middle| \beta, \log(M_0/h^{-1}M_\odot) \right) = \sum_i \left( \mu_i x_i / x_i! \right) e^{-\mu_i},$$

(4.4)

where $\mu_i = \mu(M_i)$ and $x_i = N_{i\text{obs}}$. In terms of $\chi^2$, this is

$$\chi^2 = -2 \ln(\mathcal{L}) = -2 \sum_i \left( x_i \ln(\mu_i) - \ln(x_i!) - \mu_i \right).$$

(4.6)

To create the one dimensional marginalised posterior likelihoods, or $\mathcal{L}_1(\beta)$, and find the 68% confidence level (CL) interval of $\beta$, we divide $\beta$ into 50 bins, and count the total multiplicity of models in our MCMC chain falling into each bin. Then $\mathcal{L}_1(\beta)$ is proportional to the multiplicity of each bin. We then find the peak of $\mathcal{L}_1(\beta)$, and iteratively integrate the likelihood function about this maximum, e.g.,

$$\int_{\beta_{\text{low}}}^{\beta_{\text{up}}} \mathcal{L}_1(\beta) d\beta / \int \mathcal{L}_1(\beta) d\beta,$$

(4.7)

such that $\mathcal{L}_1(\beta_{\text{low}}) = \mathcal{L}_1(\beta_{\text{up}})$ are symmetrical about the peak, and the integral becomes equal to 68%. Once this is achieved, we report the result as

$$\beta = \beta_p + (\beta_{\text{up}} - \beta_p) \left[ \frac{1}{\beta_{\text{p}} - (\beta_{\text{up}} - \beta_{\text{low}})} \right],$$

(4.8)
where $\beta_p$ is the peak likelihood presented as the best fits in Table 4.1. We treat log($M_0/h^{-1}M_\odot$) in the same way. Our 1-D posterior likelihoods are almost Gaussian, so the method described above agrees well with the 68% intervals around the median or mean.

### 4.2 Scatter about the HOD

Finally, we test the scatter about the best-fit power–law in Figure 4.1 as it is expected to be Poisson from theoretical predictions. For this purpose, we use the methodology outlined in Gehrels (1986), who provides Gaussian–like confidence intervals for Poisson statistics. In this case, Gehrels demonstrates that the one–sigma error for any count of $n$ is just $1 + \sqrt{n + 0.75}$ to an accuracy of 1.5%. Therefore, for each cluster, we compute the ratio ($\alpha$) of the observed (y–axis) distance of that cluster from the best–fit power law (given in Figure 4.1) and the estimated one-sigma error given above. For clusters of mass greater than $10^{14}h^{-1}M_\odot$, we find $\langle \alpha \rangle = 0.98$ (the mean value of $\alpha$), which confirms that the scatter about the best fit is Poisson, as expected (see also Yang et al. 2008). At lower masses, $\langle \alpha \rangle$ is much less than unity ($\sim 0.6$), and therefore the scatter is sub–Poisson (see also the next paragraph). However, this is where the C4 starts becoming incomplete, as it is hard to find clusters/groups with only a few bright galaxies (with redshifts). Also, the mass estimates for these lower mass systems have larger errors.

The second moment $\langle N(N - 1)|M \rangle$ of $P(N|M)$ which describes the width of $P(N|M)$, is usually expressed in terms of $\alpha$, where

$$\langle N(N - 1)|M \rangle = \alpha^2(M)\langle N|M \rangle^2,$$  \hspace{1cm} (4.9)
while \( \alpha(M) = 1 \) for a pure Poisson distribution, and \( \alpha(M) < 1 \) for a sub–Poisson distribution. We measured the second moment \( \langle N(N-1) | M \rangle \) of \( P(N|M) \) from our data for clusters of mass greater than \( 10^{14} h^{-1} M_\odot \) and obtained 789.6. Then we used the formula above and obtained \( \alpha = 1.12 \pm 0.07 \). This is broadly consistent with the previous result where \( \langle \alpha \rangle = 0.98 \) (see Section 4.2). Thus, we infer a Poisson scatter about the best-fit power–law in Figure 4.1. At lower masses, we measure \( \langle \alpha \rangle = 0.68 \pm 0.05 \) (consistent with the previous result in Section 4.2), and therefore the scatter is sub–Poisson (\( \alpha < 1 \)).

### 4.3 Environmental dependence of the HOD

The standard picture of large scale structure formation shows galaxies condensing to the centre of hierarchically merging dark matter halos (White & Rees, 1978). As discussed in Berlind et al. (2003), the importance of the halo model approach is the ability to give information on the galaxy bias. Lemson & Kauffmann (1999), using N-body simulations, showed that the HOD is effectively independent of environmental factors. But Sheth & Tormen (2004) found that “halos in dense regions form at slightly earlier times than halos of the same mass in less dense regions”. Recent numerical simulations have indicated that the formation time of dark matter halos is linked to the local environment of halos, in that old halos are preferentially in denser environments (Gao, Springel & White 2005). Such a correlation may have a noticeable impact on the galaxy content of halos as a function of environment. In a recent paper, Tinker et al. (2007) looked for such an environmental dependence of the HOD, using the projected galaxy correlation function and galaxy void statistics, but found no evidence for a HOD dependence on environment. As this is a central assumption in the formalism given in Section 1.5, we revisit this here because of the size and completeness of the C4 catalogue.
CHAPTER 4. HALO OCCUPATION DISTRIBUTION

First, we make a rudimentary cut in the clusters based on their distance to the nearest companion cluster, i.e., we study pairs of clusters and isolated clusters. A pair of clusters is defined to be any two clusters whose centroids are separated (in redshift-space) by less than the combined $R_{200}$ values of the two clusters. Using this definition, we found that 656 C4 clusters were in pairs out of the 1713 clusters in the DR5 C4 catalogue ($\simeq 38\%$). We found no statistical difference between the HODs of these two samples of clusters and both were consistent with the HOD parameters as given in Table 4.1. This test also indicates that the C4 catalogue, and our HOD analysis, confidently handles the issue of overlapping clusters, while also suggesting that the HOD is universal with halo environment.

To test this further, we also studied the HOD as a function of the SCF given by Miller et al. (2005), which provides information on the local environment of the C4 clusters (based on the redshift histogram). We also split the C4 catalogue in area and redshift to isolate the sample of clusters associated with the “Sloan Great Wall” (see Gott et al. 2005), which is known to dominate the higher-order correlation function of SDSS galaxies (see Nichol et al. 2006). We calculate the mean galaxy density of the Sloan Great Wall in the region $0.065 \leq z \leq 0.09$, $150^\circ \leq RA \leq 210^\circ$, $-3^\circ \leq DEC \leq 6^\circ$. The Sloan Great Wall contains 7482 galaxies in this region, about 84% of the total galaxy number of the Sloan Great Wall (Deng et al. 2006). In all cases, we find no difference between the fitted HOD parameters (within the errors) for these different subsamples of the clusters and the HOD for all C4 clusters. Therefore, we conclude there are no strong dependences of the HOD on the large-scale environment of the halos.
4.4 Discussion and Conclusions

4.4.1 The HOD for all galaxies

We present here an empirical estimate of the Halo Occupation Distribution (HOD) in nearby C4 clusters of galaxies. Our work differs from many of the other HOD measurements as we directly count the number of galaxies in each cluster as a function of their halo mass. In this way, we do not infer the HOD parameters from multi-parameter fits to statistical measures of the galaxy population, like the luminosity function, two–point correlation function or cluster abundance (e.g., van den Bosch et al. 2003, Zehavi et al. 2004, 2005, Blake et al. 2008, Rozo et al. 2007). Furthermore, we have attempted to minimise the number of assumptions required in such an estimate, which is complementary to the recent analysis of Collister & Lahav (2005), while still using a large sample of clusters and galaxies.

We stress here that the C4 catalogue used here is a factor of 8 larger than the sample of clusters used by Popesso et al. (2006) in their direct measurement of the HOD. Furthermore, the C4 catalogue used here has the same completeness, but higher quality (purity) than the 2PIGG catalogue used by Collister & Lahav (2005) in their analysis (see Figure 2 in Eke et al. 2004, where the mean quality of groups is less than 50% above a mass of $M_{200} = 10^{14} h^{-1} M_\odot$). Our catalogue is also larger (by a factor of 4) than the sample used by Kochanek et al. (2003). These advantages are an aid in the determination of the HOD, allowing control at the number of spurious clusters used in computing the mean number of galaxies per halo, as well as allowing a split of the sample as a function of cluster properties, while maintaining sufficient statistics.

To measure the HOD, we must first estimate the mass and size of each C4
cluster. This was achieved using the summed optical $r-$band luminosity of the clusters (Miller et al. 2005), as a proxy for mass (see Sheldon et al. 2007), and the stacked radial profile of cluster galaxies to determine $R_{200}$ (the radius interior to which the mean number density of galaxies is 200 times the critical density). We find that our values of $R_{200}$ are in excellent agreement with recent measurements from the stacked weak lensing profiles of maxBCG clusters (see Johnston et al. 2007). This agreement is remarkable given the differences in the methodology and cluster samples used. We also find our $R_{200}$ values are insensitive (at the $\sim 10\%$ level) to the details of the clusters centroid and the color and luminosity cuts on galaxies used to compute $R_{200}$. This indicates that our C4 cluster masses and radii are robust over nearly two orders of magnitude in mass ($10^{13} M_\odot$ to $10^{15} M_\odot$).

In Figure 4.1, we present the HOD, which is the number of galaxies within $R_{200}$ as a function of the halo mass of C4 clusters. The best fit power–law ($\langle N|M \rangle = (M/M_0)^\beta$) to halo masses $> 10^{14} M_\odot$ is $\beta = 0.99 \pm 0.01$, but we stress that the statistical error on this fit is significantly smaller than the systematic uncertainties associated with our analysis. We conservatively estimate these to be 0.1 in the slope, and therefore conclude that our HOD is consistent with a slope of one. We find the HOD slope does not depend on the environment of the halos, and supports the idea that the HOD is a universal function not strongly dependent on details of the halo formation (see Tinker et al. 2007, who find the same result looking at the SDSS and 2dFGRS correlation functions and void statistics). We should note that the link between halo formation time and environment is strongest for the less massive halos (Gao & White 2006), which are not probed here (i.e., below $10^{14} M_\odot$).

Our best fit slope ($\beta$) for all galaxies is in excellent agreement with several results in the literature using similar methodology to that outlined herein (e.g.,
Kochanek et al. 2003, Collister & Lahav 2005). Our work is closest to that of Collister & Lahav (2005; CL05) who used the 2PIGGS cluster catalogue from the 2dF Galaxy Redshift Survey (2dFGRS; Colless et al. 2001). This agreement is re-assuring, but somewhat surprising given the detailed differences between the two analyses. For example, CL05 used the velocity dispersion to estimate the masses of their clusters and groups, which can be unreliable when small numbers of clusters or group members are available, e.g., CL05 used groups with as few as four observed galaxy members per system. CL05 used 3147 groups with a total of 25118 galaxies, while we have, on average, 3 times more galaxies located in half the number of groups and clusters. Furthermore, CL05 simply scaled their cluster radii by the mass of the system, assuming \( r_{\text{vir}}^3 \propto M \). Another significant difference between the two analyses is in the definition of cluster membership. CL05 used a luminosity function to correct for the flux limit and incompleteness of the original 2dFGRS, while we have simply limited our data to a volume–limited sample of galaxies and corrected for fibre collisions using the real data. Given these differences, it suggests the high–mass end of the HOD is robust and insensitive to the details of how the clusters were formed and how one counts cluster members (at the 20% level). Our findings are also consistent with \( \beta \) values determined from the HOD modelling of galaxy correlation functions (see recent analysis of Zheng et al. 2007).

Our best fit HOD slope could be seen as being in disagreement with Popesso et al. (2006), Lin et al. (2004) and Rozo et al. (2007), who find 0.89 ± 0.05\(^2\), 0.84 ± 0.04 and 0.83 ± 0.06 respectively for the slope of the HOD. However, we feel we are consistent with these measurements given the size of the systematic errors on our measurements (see the range of results in Table 4.1), as well as the

\(^2\)We quote here the result of Popesso et al. for their spectroscopically-confirmed cluster members as this is closest to our measurement.
Table 4.3: The best fit power–law ($\langle N|M \rangle = (M/M_0)^\beta$). The value of $\beta$ as determined by this work and previous studies.

<table>
<thead>
<tr>
<th>The Literature</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>This work</td>
<td>$0.99 \pm 0.01$</td>
</tr>
<tr>
<td>Kochanek et al. (2003)</td>
<td>$0.98 \pm 0.05$</td>
</tr>
<tr>
<td>Collister &amp; Lahav (2005)</td>
<td>$0.99^{+0.15}_{-0.17}$</td>
</tr>
<tr>
<td>Zheng et al. (2007)</td>
<td>$1.04^{+0.55}_{-0.59}$</td>
</tr>
<tr>
<td>Popesso et al. (2006)</td>
<td>$0.89 \pm 0.05$</td>
</tr>
<tr>
<td>Lin et al. (2004)</td>
<td>$0.84 \pm 0.04$</td>
</tr>
<tr>
<td>Rozo et al. (2007)</td>
<td>$0.83 \pm 0.06$</td>
</tr>
</tbody>
</table>

degeneracy between the parameters in the HOD. It is unclear if the errors quoted by Popesso et al., Lin et al. and Rozo et al. include estimates for their systematic uncertainties. In summary, we feel all these results are likely consistent with a slope of unity, given the systematic errors involved (see Kochanek et al. 2003 for the same argument). Such findings are consistent with $\beta$ values determined from the HOD modelling of galaxy correlation functions (Zheng et al. 2007). We provide estimates of $\beta$ from the literature and this work in Table 4.3.

As a final test of our HOD parameters, we re-fit the data in Figure 4.1 using a modified version of the HOD model, namely we assumed each cluster possessed a single central galaxy and satellite galaxies with the form $(M/M_0)^\beta$, i.e., we assumed $\langle N|M \rangle = 1 + (M/M_0)^\beta$. This change in parametrisation made little difference to the fitted $\beta$ value of $1.02^{+0.15}_{-0.01}$, compared to $0.99 \pm 0.01$ in Table 4.1.

### 4.4.2 The HOD as a function of galaxy colours

In Table 4.1, we have split our HOD by the colour of the galaxies in C4 clusters. We define “red” galaxies to be galaxies with colours consistent with the red sequence of the cluster, while “blue” galaxies are bluishward of the cluster red
sequence. In Figures 4.2 and 4.3, we present the HOD for “red” and “blue” galaxies (HOD parameters in Table 4.1). For the red galaxies, we see a noticeable decrease in the slope of the HOD, which appears to be robust against changes in the fitting range of the halo mass and the luminosity limit of the galaxies used. We clearly measure $\beta \simeq 0.6$, which is significantly different from the HOD slope for all galaxies, even taking into account the likely systematic uncertainties at the level of 0.1 (see also Table 4.2). We have investigated the sensitivity of our HOD fit to outlying points, by studying the variation of $\beta$ when they are removed. We found that the fitted value of $\beta$ is still between 0.6 and 0.7, which remains inconsistent with the HOD slope for all galaxies (Fig. 4.1).

The most interesting consequence of this result is the fact that even the most luminous red galaxies in our sample ($M_{0.0r} < -21.7$ fitting over the halo mass range of $M_2 \geq 10^{14}h^{-1}M_\odot$) have $\beta = 0.58^{+0.02}_{-0.03}$ (0.1 systematic), which is significantly different from most other HOD slope measurements for “Luminous Red Galaxies” (LRG; see Eisenstein et al. 2001). To test this further, we re-measure the HOD slope for the “red” galaxies, but with $M_{0.0r} < -22.5$, which is close to the LRG luminosity limit at low redshift (we further constrain the data to $0.03 < z < 0.1$ for both clusters and galaxies, giving us 427 C4 clusters with 874 LRGs). For these LRGs, we find $\beta = 0.65 \pm 0.07$ for clusters masses of $M \geq 10^{14}h^{-1}M_\odot$ (181 of the 427 clusters are used).

Our LRG HOD measurements are different from those in the literature based on the analyses of the LRG correlation functions (Kulkarni et al. 2007; Blake et al. 2008; Wake et al. 2008). Such analyses find $\beta > 1$ to high significance, which is contrary to the findings of this thesis. Interestingly, our measurement for LRGs is in good agreement with the measurement of Ho et al. (2007), who uses a similar methodology to that outlined here, by counting LRG galaxies within high redshift X-ray clusters. They find $\beta = 0.56 \pm 0.25$, using a similar model.
fit to their data. These differences need to be further investigated and could be a result of different LRG definitions and differences in the methodology, e.g., we do not demand that every cluster contains an LRG in our analysis as clusters without LRGs are simply not included in the fits.

The results show for a sample of “red” galaxies, the distribution of mass is proportionally higher for small galaxies when compared to a sample of “all” galaxies. The converse is true for more massive galaxies (this is only valid for clusters of large masses, $M \geq 10^{14} h^{-1} M_\odot$).

Collins et al. (2009) have investigated the brightest cluster galaxies in the X–ray emitting clusters of galaxies at redshifts 1.2–1.5 (BCGs, located at the centres of clusters of galaxies and are the most luminous objects), and found that their stellar masses remain unchanged. As discussed by Simard et al. (2009) and Poggianti et al. (2009), who studied high–redshift cluster samples (SDSS and ESO Distant Cluster Survey, Gonzalez et al. 2002) to examine how the fractions of early–type objects change as a function of cluster redshift, velocity dispersion, and star–formation activity. They found no changes in the early–type fractions as a function of velocity dispersion. They also found that towards low redshift, the fraction of spirals decreases in favour of those of lenticulars (S0). This is probably due to the physical processes within clusters which strip off the gas from the spiral and makes a compact object, and therefore we observe changes in the mass distribution. As possible and interesting future work, we can compare our mass distribution with other cluster samples and simulations with the same mass and different redshift and to investigate and compare how our mass distribution changes with redshift.

For blue galaxies, we also see (Figure 4.3) that it is hard to fit a line to the data because we have few objects and we see a shallower slope than for all galaxies. It is also clear in Table 4.1 that the statistical errors on our fits have
significantly increased because of the paucity of blue galaxies in clusters (we note that the C4 cluster–finding algorithm is not biased against clusters of blue galaxies, which is just the well-known morphology–colour–density relationship). It is also interesting to note that the slope decreases with increasing luminosity of galaxies considered, starting at $\beta = 0.89^{+0.05}_{-0.04}$ for $L^d$, decreasing to $\beta = 0.50^{+0.09}_{-0.08}$ for $L^n$. This appears significant even given the larger statistical errors and the known systematic uncertainties (0.1 in $\beta$). Such an effect is probably expected, given the decrease in blue cluster galaxies with increasing luminosity, i.e., the bright end of the colour–magnitude relationship is dominated by red luminous galaxies.
Figure 4.3: The same as Figure 4.1, but for “blue” galaxies (bluer than the cut mentioned in Figure 4.2) within $R_{200}$ in C4 clusters.
Table 4.4: Our best fit HOD $\beta$ parameter for mock data. The “All galaxies” means our selection without a colour cut within the virial radius of the clusters, while “Red galaxies” means galaxies within the $\pm 2\sigma (r-i)$ limit on the host cluster red sequence (see text). We consider luminosity limit of $M_{0.0r} = -21.2$ used throughout this work. We also investigate two lower limits of the mass range used in fitting the HOD parameters, namely $M_1 = 5 \times 10^{13} h^{-1} M_\odot$ and $M_2 = 10^{14} h^{-1} M_\odot$.

<table>
<thead>
<tr>
<th></th>
<th>$M &gt; M_1$</th>
<th>$M &gt; M_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All galaxies</td>
<td>$0.71 \pm 0.06$</td>
<td>$0.79 \pm 0.05$</td>
</tr>
<tr>
<td>Red galaxies</td>
<td>$0.67 \pm 0.07$</td>
<td>$0.65 \pm 0.08$</td>
</tr>
</tbody>
</table>

4.4.3 The HOD for mock catalogue

In Figures 4.4 and 4.5, we present the results of the HOD analysis repeated for a C4 mock catalogue of clusters containing 388 objects. We measured the HOD in the mock galaxies using the same techniques that we applied for the real data. Our best fit HOD $\beta$ values are presented in Table 4.4 for “all” and “red” galaxies as a function of halo mass range. (Note that the “all” galaxies describes our selection without a colour cut, and we define “red” galaxies to be galaxies with colours consistent with the red sequence.) The interesting consequence of this result is the fact that for the red mock galaxies, we see that the slope of the HOD appears to be the same ($\beta \simeq 0.6$) as the slope of the HOD for the real data, and this is robust. We measure a value of $\beta \simeq 0.6$ for red galaxies, which is different from the HOD slope for all galaxies. This difference matches the result obtained from the real data (see Section 4.1). We did not measure the HOD for “blue” galaxies because of the low number of blue galaxies in clusters.

---

3This catalogue was kindly provided by Chris Miller (2000), private communication.
Figure 4.4: The number of galaxies (no colour constraints) within $R_{200}$ in C4 mock halos as a function of the C4 halo mass. Galaxy counts shown here are for absolute magnitudes of $M_r < -21.2$. The solid line is the best fit HOD model (see Section 4.1.1) for masses $\geq 10^{14} h^{-1} M_\odot$, while the dotted line is for $\geq 5 \times 10^{13} h^{-1} M_\odot$. 
Figure 4.5: The same as Figure 4.4 but for “red” galaxies (within the ±2σ $(r - i)$ limit on the host cluster red sequence) within $R_{200}$ in C4 clusters.
Chapter 5

Cluster correlation function

5.1 The two point correlation function

Studying the correlations of galaxies and clusters on large scales in the Universe can provide constraints on cosmological parameters, such as $\sigma_8$, $\Omega_m$, $\Omega_\Lambda$ and $w$. To quantify the structure and characterise the large scale distribution of galaxies, a number of statistical methods has been developed and the most widely used is to measure the two point correlation function $\xi(r)$, defined as,

$$\xi(r) = \langle \delta(x)\delta(x+r) \rangle_x,$$

(5.1)

which represents a spatial average, where $\delta(x)$ is the density contrast, $\delta(x) = \frac{\rho - \bar{\rho}}{\bar{\rho}}$. The two point correlation function can also be defined as the excess probability with respect to a Poisson distribution of finding a pair of galaxies in volumes $dV_1$ and $dV_2$, which are separated by distance $r$, i.e.,

$$dP_{1,2} = \bar{n}^2[1 + \xi(r)]dV_1dV_2,$$

(5.2)
CHAPTER 5. CLUSTER CORRELATION FUNCTION

where \( dP_{1,2} \) is the joint probability and \( \bar{n} \) is the mean density of the sample. As we see from this equation, if \( \xi(r) = 0 \), the probability is that expected from a random distribution of galaxies. If \( \xi(r) > 0 \), there is a higher probability of finding a pair of galaxies separated by distance \( r \), i.e., the case of excess clustering where there are more galaxies than in the Poisson case. If \( \xi(r) < 0 \), galaxies are less clustered than random (anti-clustering case).

The observed \( \xi(r) \) is generally described by a power law as a function of distance \( r \),

\[
\xi(r) = \left( \frac{r_0}{r} \right) ^ \gamma,
\]

(5.3)

where \( r_0 \) is a characteristic scale length which is estimated when \( \xi = 1 \). However, this simple power-law description does not account for the complex physics of structure formation which must include non-linear gravitational dynamics, gas physics and magnetic fields to name but a few.

The interesting advantage of \( \xi(r) \) is that it is related to the power spectrum \( P(k) \) in such a way that the Fourier transform of \( \xi(r) \) gives \( P(k) \),

\[
P(k) = \int \xi(r) e^{ikr} d^3r.
\]

(5.4)

Although we can obtain the same information from both, \( \xi(r) \) is an easy statistic to measure.

5.1.1 The estimators

As discussed in Kerscher et al. (2000), there are various ways of measuring the two point correlation function, \( \xi(r) \), in the literature which are practically based on counting pairs as a function of their separation distance \( r \). The correlation function gives us the ability to compare the distribution of our data to a random
sample. Therefore the $\xi$ compares the number of pairs of galaxies (DD) with pairs of randoms (RR) at a fixed separation $r$ (note that the random points must be distributed over the same volume as the real data) and is written as,

$$\xi(r) = \frac{DD}{RR} - 1. \quad (5.5)$$

Some other estimators are written as $\xi(r) = \frac{DD}{DR} - 1$ (DR is the number of data–random pairs) by Davis & Peebles (1983), $\xi(r) = \frac{DD-DR}{RR}$ by Hewett (1982), $\xi(r) = \frac{DDRR}{DR} - 1$ by Hamilton (1993), and $\xi(r) = \frac{DD-2DR+RR}{RR}$ by Landy & Szalay (1993). The main difference between the various estimators is usually the way that they treat the edge correction. Kerscher et al. show that on small scales, all the estimators are comparable. From a practical point of view, they recommend the Landy & Szalay as preferable over other methods. However, the crucial point in calculating the correlation function lies with the modelling of the random catalogue. The random catalogue must take into account the selection effects of the real data, if an accurate comparison is to be made, and this is a challenge (Percival, 2007). The main selection effects which are included in our random sample are: the survey angular mask, the redshift distribution (smoothed on an appropriate scale to avoid large scale structure), and fibre collisions (Percival et al. 2007).

### 5.2 Applying the cluster correlation function to the halo model

As discussed in Section 1.5, clusters are an excellent tool for studying the large scale structure and examining the theories of structure formation since they are the most massive gravitationally collapsed structures in the universe and they
play an important role in the hierarchy of structure formation. Clusters trace the mass distribution of the universe and they are strongly biased with respect to the matter distribution (Peacock & Dodds, 1994). The two point cluster correlation function is the best statistical technique for studying the cluster distribution directly from the observational data.

In Section 5.1, we discussed and defined the galaxy correlation function, but here we derive the cluster correlation function. This means that we calculate the cluster correlation function by counting the cluster pairs and comparing their distribution to a random distribution of clusters (which must have the same volume). We wish to apply our cluster correlation function to our halo model and to constrain the cosmological parameters such as $\sigma_8$. It is possible to derive $\sigma_8$ from the mass function, $n(M)$, or from the $1h$ term + $2h$ term power spectra (see Section 4.1).

We decided to use the newer power–spectrum approach as the focus on the “clustering” statistics ties in well with the observational aspects of our halo model. Our observables are,

- $\langle N|M \rangle$ (our HOD),
- $n(M)$ (mass function),
- cluster correlation function as a function of mass ± errors.

Therefore, deriving the observed $P_{\text{cluster}}(k)$ (related to $\xi_{\text{cluster}}(r)$), one can predict the $2h$ term and compare with the observations. As we see from Equation 4.3, we need to calculate the power spectrum, $P_{\text{dm}}^{(\text{lin})}(k)$. For this purpose, we use CAMB (Code for Anisotropies in the Microwave Background, Lewis et al. 2000) for which we specify an input cosmology. We compare the predictions of CAMB with our data and so determine an estimate of $\sigma_8$. Therefore, the next step is to calculate our cluster correlation function.

\footnote{Available from ‘http://camb.info/’ webpage.}
5.2.1 Cluster correlation function

To measure the cluster correlation function, we use the estimator of Landy & Szalay described in Section 5.1.1,

\[ \xi(r) = \frac{DD - 2DR + RR}{n_r(n_r-1)/2}, \]

(5.6)

where the denominators of DD, DR and RR are the normalisation factors, \( n_d \) is the number of data points and \( n_r \) is that of the random points. Note that \( n_r \) must be much larger than \( n_d \), because larger \( n_r \) will reduce the noise. The random catalogue created in our work is 3 times the number of objects in the main data catalogue.

The random catalogue used in our work was designed for the DR5 main sample by Percival et al. (2007), and includes angular completeness issues. We use only RA and DEC from his catalogue and not the redshifts. We assign them redshifts which are taken randomly from our own real cluster redshifts, since the cluster redshift distribution does not match exactly the distribution of the main DR5 galaxy sample.

We used the NPT code (see Moore et al. 2001) to make our measurements, which was developed as part of a Computational Astrostatistics collaboration in Carnegie Mellon University (see Nichol et al. 2001). The code is based on the structures known as ‘kd-trees’. A kd-tree is a way of organising a set of data in k-dimensional space in such a way that once built, any query requesting a list of points in a neighbourhood can be answered quickly without going through every single point.

We measured the cluster correlation function for 18 bins in the pair distance range of 0.5 to 122 Mpc. The errors have been derived using the “jackknife” method, i.e., we divided the whole sample into 23 separate regions where all the
regions have almost the same area. We measured the two point cluster correlation function 23 times and every time we excluded one of the 23 regions. Finally, the “jackknife” error estimate is calculated for the variance as (Lupton 1993),

\[ \sigma^2_\xi(r_i) = \frac{N - 1}{N} \sum_{j=1}^{N} [\xi_j(r_i) - \bar{\xi}(r_i)]^2, \]

(5.7)

where \( N \) is 23 for our case and

\[ \bar{\xi}(r_i) = \frac{1}{N} \sum_{j=1}^{N} \xi_j(r_i). \]

(5.8)

We measured the correlation function (and hence bias) as a function of mass by splitting our cluster sample into two ranges of masses: \( M \geq 10^{14}M_\odot \) and \( M < 10^{14}M_\odot \). Figures 5.2 and 5.3 show the cluster correlation function for the large (\( \sim 600 \) clusters) and low mass (\( \sim 1100 \) clusters) respectively. As we see from Figure 5.6 (where the high and low mass correlation functions are overlapped), the correlation function is larger for higher mass clusters which means that they are more strongly clustered, as expected. We measured the best-fit to \( \xi(r) \) (see Section 5.1 and Equation 5.3) over the range 2–40\( h^{-1}\)Mpc. The correlation function has the correlation length \( r_0 = 22.5 \pm 6.0h^{-1}\)Mpc and \( \gamma = 1.7 \pm 0.2 \) for the mass range \( > 10^{14}M_\odot \), and we obtained \( r_0 = 16.8 \pm 5.0h^{-1}\)Mpc and \( \gamma = 1.9 \pm 0.3 \) for the mass range \( < 10^{14}M_\odot \). The results are presented in Figures 5.2 and 5.3.

For the whole mass range of our cluster sample, the correlation function has the correlation length \( r_0 = 16.8 \pm 4.9h^{-1}\)Mpc and \( \gamma = 1.9 \pm 0.3 \) (see Figure 5.5). We checked the best-fit values if fitting is applied on larger scales (up to 80\( h^{-1}\)Mpc), and the slopes become steeper by 10%.

We wished to investigate and study the peak in the two point correlation function.
function which occurs at the scale $\sim 110h^{-1}\text{Mpc}$. This peak is the Fourier transformation of the oscillations in the observed power spectrum. At this large scale, the two point cluster correlation function should show the evidence of baryon acoustic oscillations (BAO) (Angulo et al. 2005). For this purpose, we extend our 18 bins to 23 bins until the scale 384 Mpc (see Figure 5.4) and for the whole range of mass. But we could not see this peak and our cluster correlation function has a flat tail at large scales. The errors are also very large for $r > 40h^{-1}\text{Mpc}$, and cosmic variance dominates for large scales, leading to a strong correlation between bins. We have performed a large number of tests in order to investigate this problem.

- We excluded data from near the edge of the survey, to eliminate edge effects.
- We used a different estimator which is defined in Equation 5.5 as $\xi(r) = \frac{DD}{RR} - 1$.
- By looking at the redshift distribution, $n(z)$, of our cluster data, we see that there is an over-density region around $z = 0.08$ (SDSS Great Wall, see Section 4.3). We cut this part out, which has a large clustering, and then smoothed the $n(z)$ (using a Gaussian of width 3000 km/s) and used it to assign redshifts to our random points. Figure 5.4 represents the histogram of our cluster data with redshifts.
- To check if the NPT code works properly, we measured the random–random correlation function to see if the answer is zero and we found that it was.

For all these cases, we measured the correlation function and we saw that the problem of the flat tail of the correlation function persisted. A possible reason could be that at large scales we do not have enough signal and it is very noisy and this makes it hard to measure the cluster correlation function. Possibly there may also be a problem with the random catalogue that we used, or there may be a systematic error which we have not been able to identify. Figure 5.5 shows the
Figure 5.1: Histogram of the 1713 C4 clusters with redshifts. The solid curve is the estimated redshift distribution of our sample using a Gaussian of width 3000 km/s (we have cut the SDSS Great Wall out which has a large clustering, and then smoothed the $n(z)$. See text for more details.)

correlation function which is cut at $100h^{-1}\text{Mpc}$.

5.3 Comparison with Numerical Simulation

In order to test the robustness of the C4–SDSS cluster correlations, we compared our two point cluster correlation function with that measured in an N–body simulation (Sabiu & Nichol, 2008), to determine the accuracy of our results. The cosmology was set using WMAP5 parameters (Komatsu et al. 2008) and the particles were evolved from $z = 50$ to $z = 0.3$, using the publicly available GADGET2 code (Springel 2005). A halo catalogue containing 1491 objects was then extracted from the dark matter particle distribution, where the halos are

---

2The GADGET2 code is a TreeSPH code, capable of following a collision–less fluid with the N–body method, and an ideal gas by means of smoothed particle hydrodynamics (SPH).
Figure 5.2: Two point cluster correlation function for the mass range $> 10^{14} M_\odot$. The errors are calculated using the “jackknife” method. The solid line indicates our best–fit to $\xi(r)$ (see Section 5.1 and Equation 5.3). The fit is over the range $2–40 h^{-1}$ Mpc. The correlation function has the correlation length $r_0 = 22.5 \pm 6.0 h^{-1}$ Mpc and $\gamma = 1.7 \pm 0.2$. 
Figure 5.3: Two point cluster correlation function for the mass range $< 10^{14} M_\odot$.
The errors are calculated using the “jackknife” method. The solid line indicates our best–fit to $\xi(r)$ (see Section 5.1 and Equation 5.3). The fit is over the range 2–40$h^{-1}$ Mpc. The correlation function has the correlation length $r_0 = 16.8 \pm 5.0 h^{-1}$ Mpc and $\gamma = 1.9 \pm 0.3$. 
Figure 5.4: Two point cluster correlation function for the whole mass range of our cluster sample. This is the same as Figure 5.5, but for 23 bins. The errors are calculated using the “jackknife” method (see Figure 5.5 for the best-fit to $\xi(r)$).
Figure 5.5: Two point cluster correlation function for the whole mass range of our cluster sample. The errors are calculated using the “jackknife” method. The solid line indicates our best–fit to $\xi(r)$ (see Section 5.1 and Equation 5.3). The fit is over the range 2–40$h^{-1}$Mpc. The correlation function has the correlation length $r_0 = 16.8 \pm 4.9h^{-1}$Mpc and $\gamma = 1.9 \pm 0.3$. 
Figure 5.6: Square symbols show the two point correlation function for the high mass clusters \( (M > 10^{14}M_\odot) \) and the diamond symbols show the two point correlation function for the low mass clusters \( (M < 10^{14}M_\odot) \). The errors are calculated using the “jackknife” method.
selected using a friends–of–friends algorithm with a linkage parameter, \( l = 0.2 \).

We built a random catalogue, which is 3 times the number of objects in the halo catalogue, and proceeded to measure the correlation function in the same way as explained above (note that we also got the mass thresholds correct between the data and simulations). The result of this estimation is shown in Figure 5.7. The correlation function for the simulated data has the correlation length \( r_0 = 9.3 \pm 3.7 h^{-1} \text{Mpc} \) and \( \gamma = 1.7 \pm 0.3 \). (The correlation function for our real data has the correlation length \( r_0 = 16.8 \pm 4.9 h^{-1} \text{Mpc} \) and \( \gamma = 1.9 \pm 0.3 \) as shown in Figure 5.3 as well.) As we see from Figure 5.7, there is not an exact match between our correlation function and that of the N–body simulations. We investigated the differences, which could be differences in cosmology or redshift evolution, i.e., are the simulations at \( z = 0 \) (corresponding to the range of the redshift of our cluster sample)? We checked this and realised that the simulations are at \( z = 0.3 \) which may be causing an issue. For the cosmological model, throughout our work, we assumed a flat, ΛCDM cosmology with \( H_0 = 100 \text{h} \text{km} \text{s}^{-1} \text{Mpc}^{-1} (h = 1) \), \( \Omega_m = 0.3 \) and \( \Omega_\Lambda = 0.7 \), which are not the same values for \( h \) and \( \Omega_m \) according to the values predicted by WMAP5 results (Komatsu et al. 2008) and this may be causing the discrepancy between simulations and observations.

Moreover, we also wished to compare our cluster correlation function with that measured in the literature. This was not fully possible since the difference of our work with the literature is that we have calculated the cluster correlation function as a function of mass and this has not been derived in the literature. However, by looking at some of the cluster correlation functions in the literature, such as Bahcall & Soneira (1983), Nichol et al. (1992), Postman et al. (1992), Collins et al. (2000), Basilakos & Plionis (2004), we see a power–law behaviour (see Section 5.1 and Equation 5.3) in our cluster correlation function (of the whole range of mass of our cluster sample) for the scales up to \( r = 40 h^{-1} \text{Mpc} \), which
agrees with the results of the cluster correlation function derived in the literature. We have also compared results to those in the literature which have used different mass proxies such as X–ray luminosity. Collins et al. (2000) have derived the spatial two–point correlation function from the X–ray luminosity survey REFLEX (containing 499 X–ray clusters with redshifts), and have found the correlation length of \( r_0 = 18.8 \pm 0.9 \, h^{-1}\text{Mpc} \) and \( \gamma = 1.83 \) (the power–law fit is made to \( \xi(r) \) over the range \( 4 - 40 \, h^{-1}\text{Mpc} \)). Bahcall et al. (2003) have derived the cluster correlation function for 1108 clusters (from SDSS early data) for different richness thresholds. They have obtained the correlation length of \( r_0 = 21.18 \pm 2.77 \, h^{-1}\text{Mpc} \) and \( \gamma = 2.0 \) for \( N_{\text{gal}} \geq 20 \), and \( r_0 = 17.32 \pm 1.31 \, h^{-1}\text{Mpc} \) and \( \gamma = 2.0 \) for \( N_{\text{gal}} \geq 15 \). In our analysis as we see in Figures 5.2, 5.3 and 5.5, we found \( r_0 = 22.5 \pm 6.0 \, h^{-1}\text{Mpc} \) and \( \gamma = 1.7 \pm 0.2 \) for the mass range \( > 10^{14}M_\odot \).

We also found that the correlation function for the lower mass range of our cluster sample (mass \( < 10^{14}M_\odot \)) has the correlation length \( r_0 = 16.8 \pm 5 \, h^{-1}\text{Mpc} \) and \( \gamma = 1.9 \pm 0.3 \). Therefore, we see an agreement on the value of \( r_0 \) and \( \gamma \) between our larger mass range and that of the X–ray (obtained by Collins et al. 2000), and that of the larger richness (\( N_{\text{gal}} \geq 20 \)) obtained by Bahcall et al. (2003). And for our lower mass range of our cluster sample, we found an agreement with that of the smaller richness (\( N_{\text{gal}} \geq 15 \)) derived by Bahcall et al. (2003). Nichol et al. (1992) have used the Edinburgh–Durham Southern Galaxy Catalogue (containing 79 clusters), and have derived the amplitude \( r_0 = 16.4 \pm 4.0 \, h^{-1}\text{Mpc} \) which is in good agreement with our results of the correlation length for the whole mass range of our cluster sample.

But our correlation function starts to flatten for \( r > 40 \, h^{-1}\text{Mpc} \), with large error bars; cosmic variance dominates for \( r > 40 \, h^{-1}\text{Mpc} \), leading to a strong correlation between bins. As explained before in Section 5.2.1 this could be because of the lack of enough signal at large scales, which makes it difficult
Figure 5.7: Square symbols show the two point correlation function for the halos from the simulations (Sabiu & Nichol, 2008). Diamond symbols show the two point correlation function for the clusters of our sample and the errors are calculated using the “jackknife” method. The solid lines indicate our best-fit to $\xi(r)$ (see Section 5.1 and Equation 5.3). The correlation function for the simulated data has the correlation length $r_0 = 9.3 \pm 3.7 h^{-1} \text{Mpc}$ and $\gamma = 1.7 \pm 0.3$. The correlation function for our real data has the correlation length $r_0 = 16.8 \pm 4.9 h^{-1} \text{Mpc}$ and $\gamma = 1.9 \pm 0.3$ as shown in Figure 5.5 as well.

to measure the cluster correlation function, or some problem with calculating the average density, i.e., a problem with the random catalogue that we used to measure the cluster correlation function.

5.3.1 The cluster correlation function for mock catalogue

In this section, we compare our results of the cluster correlation function for the real data (C4) to mock SDSS catalogue to ensure the catalogue is a fair representation of the SDSS. Figure 5.8 presents the results of the two point correlation function repeated for a C4 mock catalogue of clusters containing 388
objects. We measured the correlation function in the mock galaxies using the same techniques that we applied for the real data. In Figure 5.9, the correlation functions for the real and mock data are overlapped. The correlation function for the mock data has the correlation length $r_0 = 16.27 \pm 5.5h^{-1}\text{Mpc}$ and $\gamma = 2.1 \pm 0.2$. As we see in Figure 5.5 shown in Section 5.2.1, the correlation function for the whole mass range of our cluster sample (real data), has the amplitude $r_0 = 16.8 \pm 4.9h^{-1}\text{Mpc}$ and $\gamma = 1.9 \pm 0.3$. These values of $r_0$ and $\gamma$ are in good agreement. We carried out an analysis of the correlation function for both data and mock, and investigated how the correlation length ($r_0$) varies with cluster mass. For this purpose, we used four mass bins which have been chosen to have approximately equal numbers of clusters in each and therefore similar errors on the fitted values, resulting a more effective comparison. Then, we measured the $r_0$ values for each of the mass bins. The results are presented in Tables 5.1 and 5.2. As we see from the results, the $r_0$ values for both the mock and the real data are in agreement (within the error bars), and they are smaller for the smaller bins of mass and they increase for the larger bins of mass. This is reasonable since the $r_0$ value describes the clustering strength, and higher $r_0$ for higher masses (also higher in bias), means stronger clustering. This agreement of the results between the correlation functions for the real and mock data, indicates that the C4 catalogue is a fair representation of the SDSS.

Following the discussion in Nichol et al. (1992), the measured cluster correlation length, $r_0$, has given controversial results in cosmology. Bahcall & Soneira (1983) have measured $r_0 = 25h^{-1}\text{Mpc}$, indicating that the clustering strength of clusters is $\simeq 15$ times more than that of galaxies. Postman, Huchra & Geller (1992) have found a lower value of $r_0 = 20 \pm 4.0h^{-1}\text{Mpc}$, which still indicates an inconsistency with a flat universe dominated by cold dark matter (White et

---

3This catalogue was kindly provided by Chris Miller (2000), private communication.
Table 5.1: The values of the correlation length \((r_0)\) for C4 mock clusters as a function of mass.

<table>
<thead>
<tr>
<th>Mass range ((10^{13} M_\odot))</th>
<th>No. clus</th>
<th>(r_0) ((\text{Mpc}))</th>
<th>(\gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6 - 6.0</td>
<td>82</td>
<td>18.26 ± 5.50</td>
<td>1.40 ± 0.21</td>
</tr>
<tr>
<td>6.0 - 9.9</td>
<td>82</td>
<td>21.47 ± 5.46</td>
<td>1.72 ± 0.22</td>
</tr>
<tr>
<td>9.9 - 20</td>
<td>83</td>
<td>22.34 ± 5.39</td>
<td>1.39 ± 0.27</td>
</tr>
<tr>
<td>20 - 200</td>
<td>90</td>
<td>25.74 ± 5.45</td>
<td>2.28 ± 0.23</td>
</tr>
</tbody>
</table>

Table 5.2: The values of the correlation length \((r_0)\) for C4 real clusters as a function of mass.

<table>
<thead>
<tr>
<th>Mass range ((10^{13} M_\odot))</th>
<th>No. clus</th>
<th>(r_0) ((\text{Mpc}))</th>
<th>(\gamma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04 - 3.0</td>
<td>416</td>
<td>16.32 ± 5.54</td>
<td>1.76 ± 0.18</td>
</tr>
<tr>
<td>3.0 - 6.6</td>
<td>421</td>
<td>15.99 ± 5.45</td>
<td>1.89 ± 0.23</td>
</tr>
<tr>
<td>6.6 - 130</td>
<td>385</td>
<td>22.23 ± 5.51</td>
<td>2.18 ± 0.20</td>
</tr>
<tr>
<td>130 - 200</td>
<td>402</td>
<td>24.45 ± 5.52</td>
<td>1.93 ± 0.19</td>
</tr>
</tbody>
</table>

al. 1987). These results may be caused by the projection effects based on the cluster catalogues used (Sutherland 1988), and after correction, the value of \(r_0\) decreases to \(r_0 = 14h^{-1}\text{Mpc}\). Nichol et al. (1992) who have minimised the projection effects, have derived \(r_0 = 16 \pm 4.0h^{-1}\text{Mpc}\), in agreement with our results \((r_0 = 16.8 \pm 4.9h^{-1}\text{Mpc})\). The C4 catalogue used in our analysis, was created by an algorithm which minimises the projection effects (see Section 2.1).
Figure 5.8: Two point correlation function for the C4 mock cluster catalogue containing 388 objects. The errors are calculated using the “jackknife” method. The solid lines indicate our best–fit to $\xi(r)$ (see Section 5.1 and Equation 5.3). The fit is over the range $2-40h^{-1}\text{Mpc}$. The correlation function for the mock data has the correlation length $r_0 = 16.27 \pm 5.5h^{-1}\text{Mpc}$ and $\gamma = 2.1 \pm 0.2$. 
Figure 5.9: Square symbols show the two point correlation function for the C4 mock cluster catalogue containing 388 objects. Diamond symbols show the two point correlation function for the clusters of our real sample of data and the errors are calculated using the “jackknife” method. The best-fit lines and values to $\xi(r)$, are shown in Figures 5.8 and 5.9.
Conclusion

In this work we were motivated by the “halo model” which describes the relationship between the dark matter and the luminous galaxies. This model was applied to 1713 SDSS-C4 clusters in the redshift range $0.03 < z < 0.13$. The Halo Occupation Distribution (HOD) was used to measure the mean number of galaxies in a dark matter halo as a function of halo mass ($\langle N \rangle \propto M^\beta$). This was achieved by counting galaxies in the clusters as a function of cluster mass. For this purpose, we measure $M_{200}$ and $R_{200}$ (virial mass and radii of clusters) as the main part of our work. The mass was calculated from the optical luminosity (the summed cluster r-band luminosity) and the radius was calculated from the stacked radial profiles of the clusters as a function of mass.

A comparison of our results with those obtained from stacked weak lensing studies of clusters found excellent agreement between these relationships. The derived $R_{200}$ values are found to be robust (to less than a 20% uncertainty) against systematic errors associated with the cluster centroid positions, fiber collision effects and the color and luminosity cuts imposed on the sample. Then to compute our HOD, we measured the number of galaxies within the virial radius ($R_{200}$) of each cluster and found a strong correlation with $M_{200}$ for clusters of
mass $> 10^{13} M_\odot$. Subsequently we measured the slope of the HOD, $\beta$, which is dominated by systematic uncertainties and we conservatively estimate our data is fully consistent with $\beta = 1 \pm 0.1$, in agreement with previous published measurements. An investigation of the scatter about the relation between $M_{200}$ and $\langle N \rangle$ found that it is consistent with a Poisson variation, as expected from simulation. We also investigated the environmental dependence of our HOD and found no evidence to support this hypothesis. After examining the dependence on the color of cluster galaxies, we found $\beta \simeq 0.6$ for “red” galaxies and we found the same result for all galaxy luminosities including Luminous Red Galaxies (LRG). For “blue” galaxies, we found the HOD slope decreases with increasing galaxy luminosity from $\beta \simeq 0.5$ to $\beta \simeq 0.9$ for the brightest galaxies to the faintest blue galaxies.

We repeated the HOD analysis for a C4 mock catalogue of clusters containing 388 objects. We measured the HOD in the mock galaxies using the same techniques that we applied for the real data, and derived the best fit HOD $\beta$ values for “all” and “red” galaxies. One interesting consequence of our result is the fact that for the “red” mock galaxies, we see that the slope of the HOD appears to be the same ($\beta \simeq 0.6$) as the slope of the HOD for the real data, and this is robust. We measure a value of $\beta \simeq 0.6$ for red galaxies, which is different from the HOD slope for all galaxies ($0.71 \pm 0.06$ & $0.79 \pm 0.05$ for the mass range $> 5 \times 10^{13} h^{-1} M_\odot$ and $M > 10^{14} h^{-1} M_\odot$ respectively). This difference matches the result obtained from the real data (see Section 4.1).

Moreover, we computed the cluster correlation function as a function of mass and found a power–law behaviour (in agreement with the literature) for scales up to $r = 40 \text{Mpc} h^{-1}$. We measured the best–fit to $\xi(r)$ (see Section 5.1 and Equation 5.3) over the range $2–40 h^{-1} \text{Mpc}$, and found that the correlation function has the correlation length $r_0 = 22.5 \pm 6.0 h^{-1} \text{Mpc}$ and $\gamma = 1.7 \pm 0.2$ for the mass range
> 10^{14}M_{\odot}$, and $r_0 = 16.8 \pm 5.0h^{-1}\text{Mpc}$ and $\gamma = 1.9 \pm 0.3$ for the mass range $< 10^{14}M_{\odot}$. For the whole mass range of our cluster sample, the correlation function has the correlation length $r_0 = 16.8 \pm 4.9h^{-1}\text{Mpc}$ and $\gamma = 1.9 \pm 0.3$. We also compared our cluster correlation function with that of the N-body simulations and found a close but not exact match between them. The correlation function for the simulated data has the correlation length $r_0 = 9.3 \pm 3.7h^{-1}\text{Mpc}$ and $\gamma = 1.7 \pm 0.3$. Moreover, we repeated the analysis of the two point correlation function for a C4 mock catalogue of clusters containing 388 objects. We measured the correlation function in the mock galaxies using the same techniques that we applied for the real data. The correlation function for the mock data has the correlation length $r_0 = 16.27 \pm 5.5h^{-1}\text{Mpc}$ and $\gamma = 2.1 \pm 0.2$. And the correlation function for the whole mass range of our cluster sample (real data), has the amplitude $r_0 = 16.8 \pm 4.9h^{-1}\text{Mpc}$ and $\gamma = 1.9 \pm 0.3$. These values of $r_0$ and $\gamma$ are in good agreement. We carried out an analysis of the correlation function for both data and mock, and investigated how the correlation length ($r_0$) varies with cluster mass. For this purpose, we used four mass bins and measured the $r_0$ values for each of the mass bins. The results that we obtained is that the $r_0$ values for both the mock and the real data are in agreement (within the error bars), and they are smaller for the smaller bins of mass and they increase for the larger bins of mass. This is reasonable since the $r_0$ value describes the clustering strength, and higher $r_0$ for higher masses (also higher in bias), means stronger clustering. These agreements of the results indicates that the C4 catalogue is a fair representation of the SDSS.

As possible future work, the aim would be to apply our cluster correlation function to our halo model and then to constrain the cosmological parameters such as $\sigma_8$ and $\Omega_m$. 
A detailed understanding of the Halo Model will help in many areas of Cosmology, as it is a natural language to interpret the data. From secondary contributions to the CMB, Kinetic and Thermal SZ effects to Weak Lensing studies, the Halo Model will prove an invaluable tool in future research. The halo model can be useful in making several key predictions, including understanding why galaxy clustering produces a power–law correlation function, estimating statistical biases in large scale structure weak lensing surveys, and calculating the full covariance matrix associated with certain large scale structure observations, such as the angular correlation function of galaxies in the SDSS. This will only be possible if a concentrated effort is applied to up and coming observational cluster projects using large galaxy surveys like SDSS–III, DES, Pan–STARRS (all are optical surveys), eROSITA (X–ray survey).


[Baum(1959)] Baum, W. A. 1959, The Hertzsprung-Russell Diagram, 10, 23


BIBLIOGRAPHY


[de Vaucouleurs(1948)] de Vaucouleurs, G. 1948, Annales d’Astrophysique, 11, 247


Miller, C. J. 2000, Ph.D. Thesis,


[Sabiu & Nichol (2008)] in preparation


[Schindler(2002)] Schindler, S. 2002, Memoriedella Societa Astronomica Italiana, 73, 965


[Torki et al. (2008)] Torki, M., Nichol, R. C., Miller, C. J., The halo occupation distribution of galaxies in SDSS C4 clusters, unpublished


