Simulating Large Cosmology Surveys with Calibrated Halo Models

Stuart Lynn

Institute for Astronomy

School of Physics & Astronomy

University of Edinburgh

Doctor of Philosophy

September 2009

Abstract

In this thesis I present a novel method for constructing large scale mock galaxy and halo catalogues and apply this model to a number of important topics in modern cosmology. Traditionally such mocks are created through first evolving a high resolution particle simulation from a set of initial conditions to the present epoch, identifying bound structures and their evolution, and finally applying a semi-analytic prescription for galaxy formation. In contrast to this computationally expensive procedure, I use low resolution simulations to obtain a density field that traces large scale modes. From this background I sample the population statistics of halos: the number of halos which are typically found within a region of a given overdensity, to produce a halo catalogue.

From the halo catalogue I then produce galaxies by appealing to the halo model. In this model the expected number of galaxies within a halo and the distribution of their properties is dependent on halo mass alone. By sampling conditional luminosity functions for a number of populations of galaxies, I produce a galaxy catalogue with luminosity and colour properties.

The aim of developing algorithm is not to probe the mechanics of galaxy formation in great detail. It is instead intended as a method of rapidly producing mock galaxy and halo catalogues rapidly on modern desktop computers. The approach we will take is to try to distill the minimal algorithm required to achieve this and still provide useful catalogues for observational cosmologists.

Both the conditional mass function and conditional luminosity functions required for the algorithm are calibrated from the Millennium Simulation, one of the highest resolution cosmology simulations to date, and its associated semi-analytic catalogues. In Chapter 2 I examine these statistics and provide fits to the quantities of interest. As a test of the method, in Chapter 3 I produce a halo and galaxy catalogue from the same large scale modes as the Millennium Simulation. The clustering statistics of galaxies and halos within this re-simulation are calculated and compared with those of the original.

Confident of the accuracy of the method, in Chapter 4 I populate a number of

simulations, each 8 times the volume of the Millennium Simulation, and study the evolution of the Baryon Acoustic Oscillation signal. For each population (dark matter, halos and galaxies) I fit the BAO in the power spectrum to obtain the shift in the BAO peak.

In Chapter 5 I extend the algorithm to produce lightcones: simulated skies in which the evolution of the Universe along the line of sight is accounted for. I simulate the geometry and limitations of a major pending survey and calculate the expected clustering signature I expect to see in both. The redshift space distortions induced by peculiar velocities of galaxies along the line of sight are determined and their ability to distinguish between gravity models is also explored.

In Chapter 6 I detail a further extension to the algorithm for simulating weak gravitational lensing surveys. I use the analytic 2D surface density profiles of NFW profiles to dress each dark matter halo on a lightcone. The sum of these profiles over the entire population can be used to construct high resolution maps of the convergence. From these maps I calculate the spectrum of the convergence and compare with theoretical predictions.

Finally in Chapter 7 I discuss further possible applications and extensions of the algorithm I have developed in this thesis.

Declaration

I declare that this thesis is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University. I further state that no part of my thesis has already been or is being concurrently submitted for any such degree, diploma or other qualification.

This thesis is the outcome of my own work except where specifically indicated in the text.

Stuart Lynn
Edinburgh,
September 2009.

Acknowledgements

I cant imagine what the 16 year old me would have said looking at this thesis when he started university all those years ago. I doubt he would have believed that he was capable of producing something like this or surviving the process needed to get here. He would have been right as well if it had not been for a number of wonderful people who have supported, educated and kept me sane over the years its taken to write this thesis. I would like to thank each of those people here.

First and foremost is my supervisor Professor John Peacock. His wisdom patience and support have been instrumental in the production of this thesis. I would also like to thank him for giving me so many glimpses of understanding of the beautiful Universe around us.

I would also like to thank many of the rest of the faculty at the Edinburgh Royal Observatory for providing such a wonderful environment to work in and to single out Alan Heavens, Andy Taylor and Peder Norberg for their extra support and collaboration.

The students and postdocs of the ROE also have a large role to play in this thesis. I would like to thank all of them for pints down the pub, late nights at work, wine fridays and all the laughs through the years. In particular I would like to thank Rita Tojeiro for always being a great warm friend and a person to share awe and the beauty of the Universe with, Eric Tittley for being the best scot not from Scotland, Simon Reynolds for being a wonderful flatmate, Berian James and Alina Kiessling for contribute work and everyone else who has touched my life over the years.

The road to a phd is a long and hard one and you need distractions to keep you going and sane. I have to acknowledge the wonderful staff of the ROE visitors centre, Tania Johnston, Dave Charlton and Dan Hillier for providing me with such a great distraction in the form of outreach. You are all fantastic people who I will miss dearly and I loved our weekends to random places in scotland, long nights showing people telescopes on the roof and chaotic open days. These will be my best memories of the ROE.

Finally for their constant support belief and just all round awesomeness, I would

like to thank my family: John, Anne, Craig, Kirsten and Aileen. You are all the most inspirational, loving, fantastic people I know and I couldn't have done this without you by my side every step of the way.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background cosmology

It is widely thought that the Universe that surrounds us today evolved from an small region through an inflation period in which some mechanism exponentially expanded a region by around $\gtrsim 10^{26}$ times. The same process drives quantum fluctuations which give rise to a spectrum of inhomogeneities.After the inflationary period ends, the Universe is left in a hot dense state of plasma in which hydrogen, helium and a few other trace elements are produced. Within this plasma, charged baryons are tightly coupled with photons through Compton scattering leading to pressure supported medium which supports sound like fluctuations.

As the Universe expands further, the plasma cools until it reaches a temperature of $T \sim 0.3$eV Kolb et al. (1990) at which point electrons and proton in the plasma combine to form neutral atoms. Now free to travel, the photons are observed today as the Cosmic Microwave Background (CMB). Without pressure support, gravity becomes the dominant force in the Universe and acts on the small inhomogeneities, amplifying overdense regions and diluting underdense ones.

The resulting arrangement of matter, referred to as the cosmic web, is an intricate structure made up of sheets, filaments and haloes of dark matter. Within these dense dark matter haloes, baryons can radiate and condense until they are

dense and hot enough to ignite nuclear fusion thus forming stars. This initial population of stars, made of mostly hydrogen and helium, are massive and short lived and produce a number of the higher mass elements in their cores. Once they have burned as much of their hydrogen fuel as they can, these stars explode in powerful supernovae polluting the surrounding gas with heavy elements which leads to a second generation of less massive, longer lived stars.

Collections of these stars, along with dust, gas and dark matter form galaxies, which are initially numerous and small. As the cosmic web evolves, the dark matter haloes that house the galaxies merge hierarchically; in the process, galaxies become satellites of one and other or merge together to produce larger more luminous galaxies. Today, the largest structures which have formed out of this hierarchical merging are giant super clusters containing many thousands of galaxies.

1.1.1 Isotropy and homogeneity

One of the founding ideas on which we base our understanding of the Universe is that of the Cosmological principle. An extension of the Copenician view that we are not at the centre of the solar system, the cosmological principle states that our position in the Universe is in no way special and that there is no preferred location or direction in the Universe. On sufficiently large scales we observe this to be true: the distribution of galaxies is broadly isotropic and homogeneous. Even more compellingly, the microwave light we observe from the early Universe as the cosmic microwave background is homogeneous to a level of one part in 100,000. While on small scales this is clearly not true, it serves as a good starting point from which the evolution of non-uniform structures can be thought of as perturbations.

1.1.2 Robertson-Walker metric

The theoretical framework used to understand the large scale Universe is that of General Relativity, in which gravity is described as a geometrical effect: the warping of 4D space time by the presence of matter. This distortion is described by a metric which relates the proper time interval between events, $\delta\tau$, to infinitesimal

displacements along the four directions in spacetime : δx, δy, δz, δt. We begin by considering a truly isotropic and homogeneous Universe with no fluctuations. In such a Universe the curvature of space-time must be the same everywhere and this constraint forces the form of the metric to be :

$$c^2\delta\tau = -g_{\phi\varphi}dx^{\phi}dx^{\varphi} = c^2dt - R(t)(dr^2 + S_k^2(r)(d\theta^2 + \sin^2(r)d\phi^2)), \tag{1.1}$$

where $R(t)$ is the scale factor which represents the evolution of the separation of any two particles and we have taken r to be the comoving separation between objects. The three possible geometric solutions which obey the constraints of homogeneity and isotropy are encapsulated in $S_k(r)$ which takes on the values:

$$S_k(r) = \begin{cases} \sin(r) & \text{if } k = 1 \\ r & \text{if k=0} \\ \sinh(r) & \text{if k=-1} \end{cases}, \tag{1.2}$$

for a Universe with positive constant curvature ($k = 1$), no curvature ($k = 0$), and negative constant curvature ($k = -1$). In the positive curvature case the Universe is wrapped around back on itself in such a way that it is finite in extent and two parallel geodesics, the paths of light through the metric will eventually converge. In both the $k = 0$ and $k = -1$ cases the Universe is infinite in extent and parallel geodesics will remain parallel in the first case and diverge from each other in the second. Space time is curved in response to the presence of energy and matter so to complete the description of the Universe we have to also specify the matter it contains and how it relates to its curvature. In General Relativity, the various contents of the Universe are described by the energy-momentum tensor $T^{\mu\nu}$ which is related to the metric via :

$$G_{\nu}^{\mu} + \Lambda g^{\mu\nu} = \frac{8\pi G}{c^4}T^{\mu\nu}, \tag{1.3}$$

where $G^{\mu\nu}$ is a tensor of the metric components and their derivatives. The constant Λ was originally introduced by Einstein to counteract the attractive force of gravity to allow a solution to the equation in which the Universe was static. We now know that the Universe is expanding and that the expansion is getting faster which might be attributed to the inclusion of a non-zero cosmological constant.

In general, the above equation must be solved for the Universe and its contents and the solution to this equation for a homogenous Universe is discussed in the

next section. However, our Universe is not entirely homogenous and the equations above are hard if not impossible to solve for this more general case. Thankfully we can treat the dynamics of our Universe as small perturbations to the homogeneous case and appeal to Newtonian gravity in many cases.

It is useful at this point to define a few dimensionless quantities related to the metric: the dimensionless scale factor $a(t)$ and the conformal time η:

$$a(t) = \frac{R(t)}{R(0)} \tag{1.4}$$

$$\frac{d\eta}{dt} = \frac{1}{a}, \tag{1.5}$$

where $R(0)$ is the value of the scale factor today.

1.1.3 Friedmann Equation

Solving the equations of general relativity for the simple case of a homogenous Universe amounts to determining the time evolution of the scale factor. The derivation will not be explicitly given here but can be found in Robertson (1935). The result is the Friedmann equation:

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R^2}, \tag{1.6}$$

where ρ is the combined energy density of the contents of the Universe and $(kc^2)/(a^2)$ is the effect of curvature. It is clear from equation 1.6 that the curvature of the Universe is linked with its density. For the critical case where $k = 0$ we find we require the density of the Universe to be :

$$\rho_c = \frac{3H^2}{8\pi G}, \tag{1.7}$$

where we have introduced :

$$H = \frac{\dot{R}}{R} = \frac{\dot{a}}{a} \tag{1.8}$$

If the density of the Universe is greater than ρ_c, the Universe is spatially closed and the evolution of the scale factor is also bound: rising to a maximum and then being halted and reversed by gravity, decreasing to zero in a big crunch. Conversely if the density is less than ρ_c, the Universe is spatially unbound and there is insufficient mass to stop the expansion.

The evolution of the scale factor is therefore dependent on the contents of the Universe ρ and its curvature k. We define the density parameter

$$\Omega = \frac{\rho}{\rho_c}, \tag{1.9}$$

which can be split in to the various components which make up the Universe.

1.1.4 Components of the Universe

We have seen in the previous section that the dynamics of the scale factor depend on the energy density of the Universe. This energy density is split up in to a number of components ρ_i and for each we can define a density parameter:

$$\Omega_i = \frac{\rho_i}{\rho_c} \tag{1.10}$$

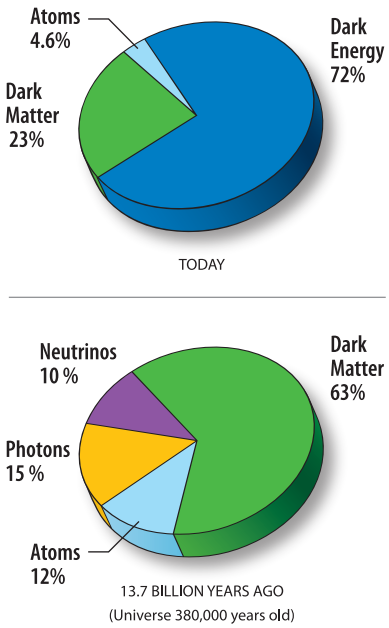

Figure 1.1: Current and past contributions to the energy budget

During the Universe's history the dominant contribution to the energy budget has evolved. Initially when the Universe was still filled with plasma, the dominant contribution came from radiation. As the Universe expanded the energy density of radiation was diluted away and the dominant factor became the cold dark matter which drove the gravitational collapse of matter forming structures. At

the current epoch the expansion of the Universe has diluted the contribution from dark matter enough that dark energy now dominates and is in the process of accelerating the expansion. Each of forms of matter/energy respond to the expansion, and contribute to the Friedman equations in different ways.

Radiation and relativistic matter

We observe the Universe through the radiation which baryons emit, absorb and reflect and as such it is an important component. However, in terms of structure formation: radiation is only dominant at early times. Their are two main sources of radiation in our Universe: primordial photons from the cosmic microwave background and those emitted by astrophysical processes. In terms of energy density the primordial contribution is dominant. However, at the current epoch this component is negligible, $\omega_r h^2 \simeq 4.2 \times 10^{-5}$. As the Universe expands the number of photons in a given volume is diluted by a^3 and the photons are redshifted by the expansion of the Universe by another factor of a leading to a relation of :

$$\rho_r(a) = \rho_{r0} a^{-4} \tag{1.11}$$

While negligible now at early times the energy density in radiation equalled and surpassed that matter with the redshift of equality being :

$$1 + z_{eq} \simeq 23.900(\Omega_m h^2)^{-1} \tag{1.12}$$

Baryons

The component of the Universe which is most familiar to us is the matter which makes up galaxies, stars and planets. Consisting of baryons: quarks, electrons and other particles of the standard model of particle physics, baryonic matter interacts with itself through the four known forces: electromagnetic, strong, weak and gravity. Despite being responsible for every visible object in the night sky, luminous matter is only a fraction of the baryons in the Universe, with most existing as the gas which is present between stars and galaxies. In total the baryon content is currently believed to make up only 4% of the total energy budget of

the Universe and only 10% of this is in the form of stars, with a further 20% as hot gas and the remanning 70% as cold gas. Like radiation, the contribution of baryons to the energy density of the Universe is diluted by the expansion in terms of number density :

$$\rho_b(a) \propto a^{-3} \tag{1.13}$$

Despite consisting of a small fraction of the global energy budget, in some regions of the Universe the baryonic content is dominant. The ability of baryons to interact through the electromagnetic force mean that they are collisional and can radiate energy through the emission of photons. This allows baryons in a sufficiently dense potential well to condense to form the compact dense structures of stars and galaxies.

Dark matter

Dark matter is thought to be the most dominant form of matter in the Universe. Although not as of yet observed directly, a growing body of indirect evidence points to a component of the Universe not described by the standard model of particle physics.

Originally proposed by Zwicky in the 1930s to explain the discrepancy in the velocities of galaxies in the Coma cluster and the observed mass in baryons dark matter has gained general acceptance with further evidence coming from the rotation curves of galaxies (de Blok and McGaugh, 1997), the perturbations of the CMB and the clustering of galaxies.

Perhaps the most striking evidence to date comes from gravitational lensing (see chapter 6) measurements of merging clusters. The bullet cluster shown in Figure 1.2 is a pair of colliding clusters which has been imaged in the optical and x-ray wavelengths, and has also had its mass distribution determined using weak gravitational lensing. This allows us to map the locations of the three components of the cluster, stars and cold gas in the form of galaxies visible in the optical, hot gas which is visible in the x-ray and the total mass distribution, including dark matter from the weak lensing analysis.

As the two clusters have collided the galaxies have passed by each other, while

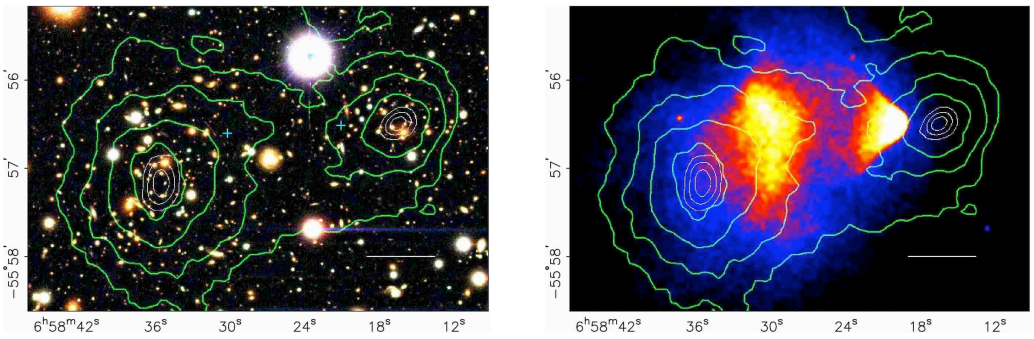

Figure 1.2: From Clowe et al. (2006), the bullet cluster showing the initial phases of two clusters merging. On the left we see the mass contours obtained from gravitational lensing overlaid on the optical view of the cluster galaxies while the right image shows the distribution of gas in the cluster. The displacement of lensing contours from the gas suggests a dominant component of invisible collision-less matter is present in both clusters.

the gas in the clusters has, due to its collisional nature, lagged behind. The drag on the gas of smaller of the two clusters has formed a bow shock in the gas of the larger. The gas in the clusters is more massive than the combined mass of the galaxies by many times, so if the cluster contained only baryonic matter we should expect to find most of the mass measured by lensing in the centre of the collision. What is actually seen, is the mass clustered around the galaxies. This strongly suggests an unseen collisionless component which constitutes most of the mass of the cluster.

It is thought that this missing matter is dark matter, a collisionless form of matter not present in the current standard model of particle physics. Dark matter is thought to interact only through the weak force and gravity, as a result it is unlikely to emit or absorb photons and has a small interaction cross-section with both ordinary matter and itself.

Like ordinary matter the density of dark matter is diluted by the expansion of the Universe by a factor :

$$\rho_{cdm}(a) \propto a^{-3} \tag{1.14}$$

Dark energy

Dark energy is perhaps the largest mystery which exists in modern physics. Evidence from supernovae in distant galaxies suggests that there exists a component of our Universe with negative pressure. Supernove observed by Riess et al. (1998) and Perlmutter et al. (1999) in high redshift galaxies act as standard candles allowing the distance redshift relation to be probed beyond the local Universe. The results show a Universe inconsistent with a slowing expansion, instead they point to a Universe in which the rate of expansion is increasing.

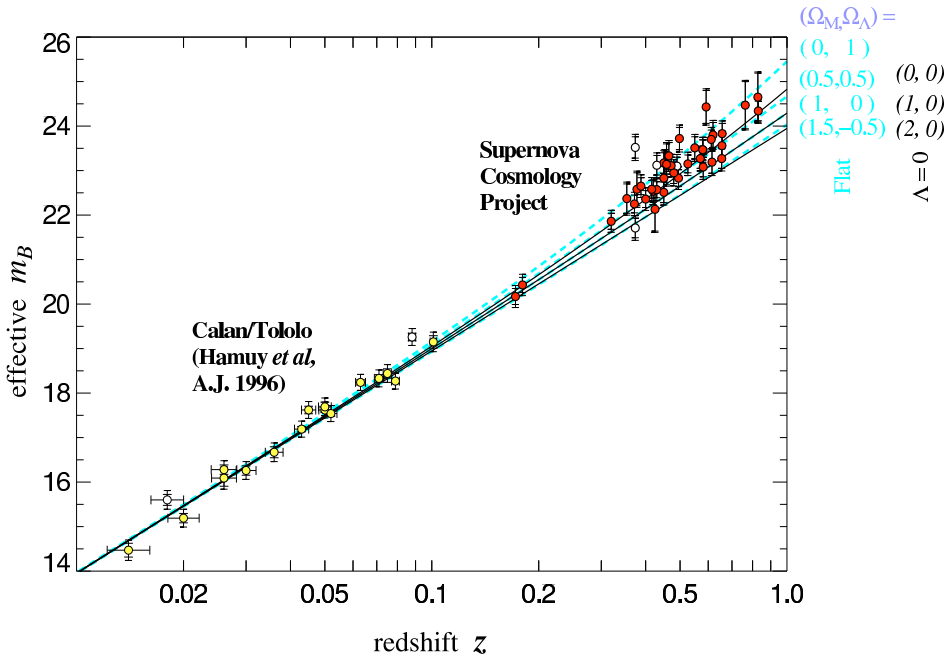

Figure 1.3: From Perlmutter et al. (1999). Inclusion of high redshift galaxies in the Hubble diagram is strongly inconsistent with a slowing expansion rate. The tendency of supernovae to be fainter and therefore more distant at high redshift than is predicted for a world with $\Lambda = 0$ provides evidence of a component of Dark Energy in the Universe increasing the rate of the expansion.

Not much is known about the nature of dark energy and so a general model in which it is described as a spatially uniform and possessing an equation of state w which relates its density to its pressure :

$$p(x) = w(x)\rho(x)c^2 \tag{1.15}$$

To exhibit an accelerating Universe, left hand side of the Friedman equation is required to be > 0 (ie $\rho + 3P/c^2 > 0$) leading to the condition $w < -1/3$:

$$-\frac{4\pi G}{3}(1+3w) > 0. \tag{1.16}$$

A large number of models have been proposed for dark energy attributing its effects to one or many quantum scalar fields, a break down of general relativity on large scales or the interaction of brane worlds. However, the simplest explanation was originally proposed by Einstein as a constant term to General Relativity. In this theory, known as the cosmological constant λ, the effects of dark energy are attributed to a non-zero vacuum energy. As it is associated with empty space this model has $w = -1$. Such a vacuum energy can be caused by the constant but brief creation and annihilation of particle/ antiparticle pairs in empty space. However, calculations from particle physics currently puts the value of the cosmological constant on the order of 10^{120} times greater than that required for the observed expansion rate.

In more complicated theories the equation of state of dark energy is not necessarily a constant but may evolve in time. A simple parameterisation of this evolution is one in which the equation of state varies linearly with the expansion factor a

$$w(a) = w_0 + (1-a)w_a \tag{1.17}$$

Putting all of the above contributions together, for the case where w is constant, we can write the Friedmann Equation in terms of present day observable quantitates

$$H^2 = H_0^2(\Omega_m a^{-3} + \Omega_r a^{-4} + (1-\Omega)a^{-2} + \Omega_v a^{-3(1+w)}) \tag{1.18}$$

where the $(1-\Omega)a^{-2}$ term comes from curvature.

1.2 Structure formation

Until now we have modelled the Universe as completely homogeneous and isotropic. However in such a Universe no structures can ever form. To proceed we consider perturbations around this background.

As we have already quantified the evolution of the scale factor with the Friedmann equation we want to remove the effects of the expansion from the description of perturbations. To this end we use the comoving coordinate system in which we factor out the expansion of the Universe :

$$\mathbf{x}(t) = a(t)\mathbf{r}(t) \tag{1.19}$$

$$\partial\mathbf{v}(t) = a(t)\mathbf{u}(t) \tag{1.20}$$

where the comoving peculiar velocity $\mathbf{u}$ is the time derivative of $\mathbf{r}$ and so $\dot{\mathbf{x}} = \dot{a}\mathbf{r}+a\dot{\mathbf{r}} = H\mathbf{x}+a\dot{\mathbf{r}}$ which is just the Hubble flow plus the peculiar velocity $\delta\mathbf{v} = a\mathbf{u}$.

Linear Theory

We measure the departure from homogeneity of the Universe as the overdensity field: δ:

$$\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}(\mathbf{x})} - 1. \tag{1.21}$$

The equation of motion for this field can be found by first differentiating $\mathbf{x} = a\mathbf{r}$ twice to obtain

$$\ddot{\mathbf{x}} = a\dot{\mathbf{u}} + 2\dot{a}\mathbf{u} + \frac{\ddot{a}}{a}\mathbf{x} = \mathbf{g}_o + \mathbf{g}, \tag{1.22}$$

where $\mathbf{g}_0$ is the acceleration a particle feels in a homogeneous universe and $\mathbf{g}$ is the peculiar acceleration field. Subtracting off the equation of motion in a homogeneous universe $(\ddot{a}/a)\mathbf{x} = \mathbf{g}_0$ and using the linear approximation for the continuity equation as experienced by fundamental observers moving with the Hubble flow:

$$\dot{\delta} = -\bigtriangledown \cdot \mathbf{u}, \tag{1.23}$$

we obtain the equation for the evolution of δ as :

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\rho\delta. \tag{1.24}$$

A solution to this can be seen to be:

$$\frac{d\ln\delta}{d\ln a} \approx \Omega_m(a)^\gamma \tag{1.25}$$

In general relativity $\gamma \simeq 0.55$ (Linder, 2005; Guzzo et al., 2008). However, a large number of alternative gravity theories : Brane worlds, modified gravity etc the growth of perturbations can be shown to agree with the above expression but with varying values of γ. This implies the rate of growth of structures, is a potentially important probe of alternative gravity.

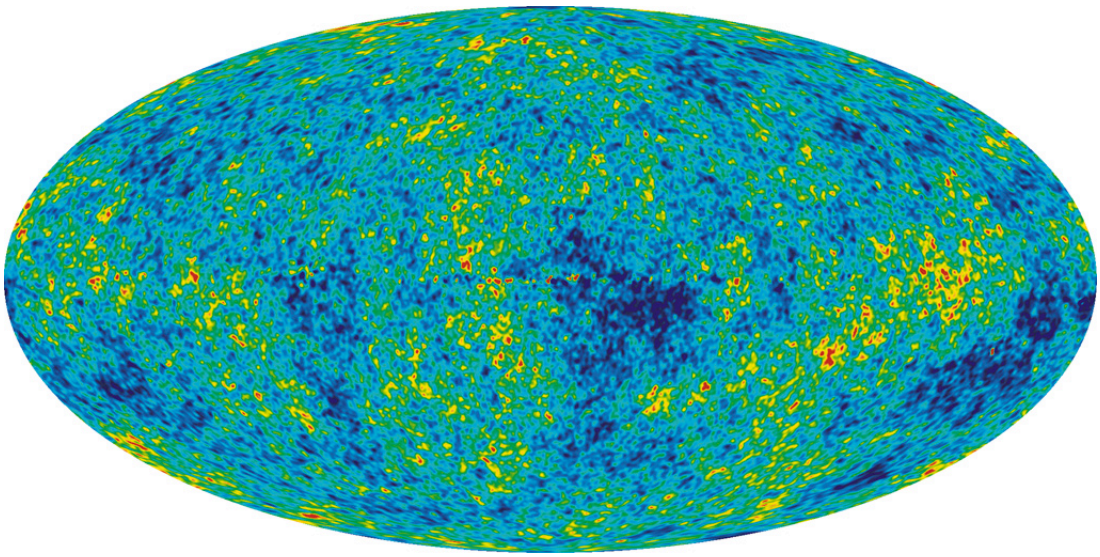

Figure 1.4: Fluctuation in the early Universe as seen in the cosmic microwave background radiation

1.3 Statistics of the Density field

1.3.1 Power Spectrum

The primordial density fluctuations are thought to be very close to a Gaussian random field, and as such its properties are totally characterised by a single function: the power spectrum with all higher order moments equal to zero. As the Universe evolves and gravity produces non linear structures this will not always be true and the bispectrum, trispectrum and higher become significant. However the power spectrum remains the most natural statistic with which to describe the density field.

The power spectrum is defined as the ensemble average of the Fourier modes of wave vector k :

$$P(\mathbf{k}) = \langle \delta^2(\mathbf{k}) \rangle \tag{1.26}$$

However as we expect the density field to be isotropic we usually simplify the power spectrum to $P(k) = \langle |\delta_k^2|(k) \rangle$.

Expressed in this way the density field can be seen to be a superposition of waves each with amplitude that has the expectation value of $\sqrt{P(k)}$ and a random phase. While any two realisations or any two sufficiently large regions of the Universe will have the same power spectra, if they have different phases the structures present in each will be different.

It is more common to find the the power spectrum expressed in a dimensionless form as the variance per $ln(k)$ giving:

$$\Delta^2(k) = \frac{V}{(2\pi)^3} 4\pi k^3 P(k) \tag{1.27}$$

which unlike $P(k)$ is now independent of the normalisation volume V. The functional form of the power spectrum can be thought of as being a combination of two factors: the initial spectrum of fluctuations produced at inflation and the imprint of physical processes that have occurred in the Universe since the end of inflation. At time of writing, the mechanism which drove the initial rapid expansion of the Universe and seeded the perturbations from homogeneity in our Universe is unknown. However most predict a power law distribution with a power index n_s of very close to 1 and the current best measurements of the CMB fluctuations place $1 - n_s = 0.037^{+0.015}_{-0.014}$ (Dunkley et al., 2009).

While the initial fluctuations lack any features the subsequent evolution of the Universe through the radiation dominated era subjects the density field to physical processes which imprints a number of signatures on the power spectrum. These departures from a power law are encapsulated in the the transfer function $T(k, z)$. Taking both these contributions together let us express the power spectrum as :

$$P(k) \propto k^{n_s} |T(k, z)|^2 \tag{1.28}$$

A full fitting formula for the power spectrum which includes all of the effects described below is presented in Eisenstein and Hu (1998) and shown in figure 1.6.

Baryon Acoustic Oscillations

While still in a hot plasma state, the early Universe supported sound waves within the tightly coupled baryon photon fluid. These compression waves are free to travel through the Universe until recombination at $z \simeq 1000$ at which point free electrons and protons combine to form neutral atoms, and pressure support within the Universe is lost. At this point the furthest a pressure wave could have travelled is simply the speed of sound in the medium multiplied by the age of the Universe. This gives a sound horizon at recombination of :

$$s = \int_0^{t(z_d)} c_s(1+z)dt \tag{1.29}$$

$$= \frac{2}{3k_{eq}}\sqrt{\frac{6}{R_{eq}}}\ln\frac{\sqrt{1+R_d}+\sqrt{R_d+R_{eq}}}{1+\sqrt{R_{eq}}} \tag{1.30}$$

where z_d is the redshift of decoupling, c_s is the sound speed, k_{eq} the equality wavenumber and R_{eq}, R_d are the weighted ratio of the density of baryons and radiation evaluated at the redshift of equality and decoupling respectively, which can be evaluated with (Eisenstein and Hu, 1998):

$$R(z) = \frac{3\rho_b(z)}{4\rho_\gamma(z)}. \tag{1.31}$$

One of the most interesting aspects of the Baryon Acoustic Oscillations is their role as a standard ruler. After recombination the frozen in sound waves form a excess of matter of scales of around the fundamental mode. These excesses are where we would expect to see galaxies forming and so in the galaxy correlation we expect a similar BAO bump. Indeed this has been detected by both the SDSS and 2dF galaxy surveys Figure 1.5

As the Universe expands the physical scale of the BAO signature will increase, which implies that if we have a measurement of the BAO scale at a number of redshifts we can directly probe the expansion rate of the Universe. This promises to give precise measurements of the equation of state and evolution of dark energy. We discuss this aspect more in chapter 4.

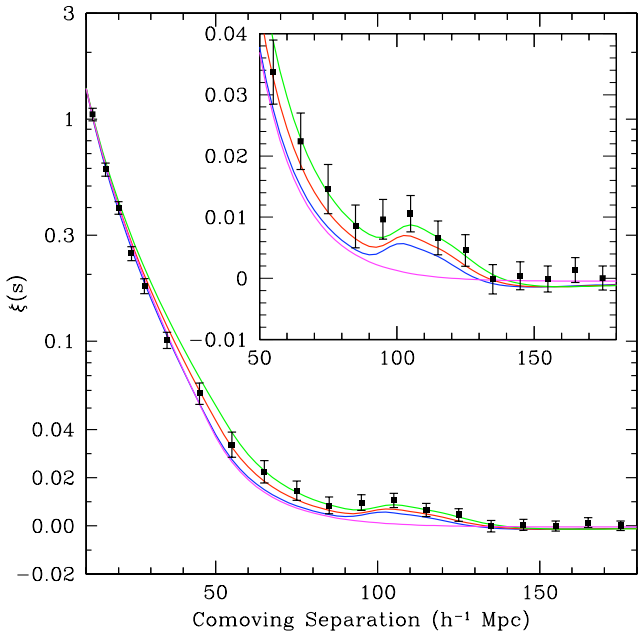

Figure 1.5: The detection of the BAO peak in the Luminous Red Galaxy catalogue of the Sloan Digital Sky Survey (SDSS) (Eisenstein et al., 2005).

The location of the BAO peak at de-coupling is determined by both the matter density and the baryon density and is given in k-space by (Eisenstein and Hu, 1998)

$$k_{peak} = \frac{5\pi}{2s}(1 + 0.217\Omega_o h^2), \tag{1.32}$$

where

$$s = \frac{44.5 \ln (9.83/\Omega_0 h^2)}{\sqrt{1 + 10(\Omega_b h^2)^{3/4}}} Mpc \tag{1.33}$$

1.3.2 Correlation function

While the power spectrum is usually the theorist choice of description, for most observational data its real space counterpart, the correlation function is preferred :

$$\xi(r) = \langle \delta(x)\delta(x + r) \rangle \tag{1.34}$$

The correlation function can be thought of as the fractional over-density of the number of pairs with a shell of radius r. The correlation function and power

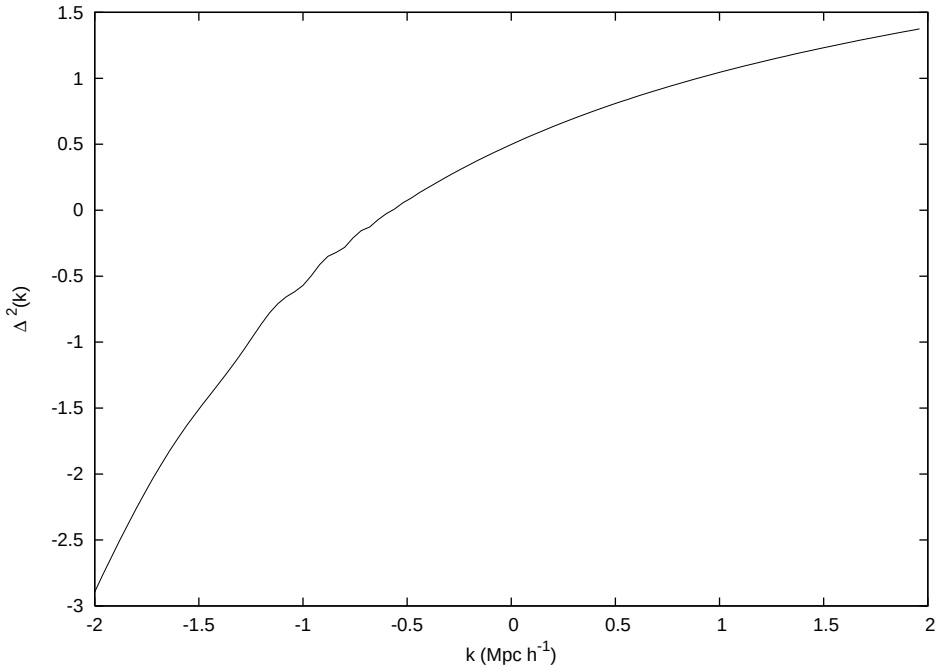

Figure 1.6: Power Spectrum of fluctuations as calculated by the fitting formula in Eisenstein and Hu (1998)

spectrum are related by:

$$\xi(r) = \int_0^\infty \Delta^2 \frac{dk}{k} \frac{\sin r}{kr} \tag{1.35}$$

1.4 Dark Matter haloes

1.4.1 Non-linear models of collapse

As gravitational forces act on matter in the Universe, regions of high density will evolve from the linear to non-linear regime. The end point of this process for the highest peaks in the density field is the formation of gravitationally bound structures. To follow this process in detail requires numerical simulations, but there are two well known models which are worth discussing here, which try and probe beyond the linear regime. The spherical collapse model follows the growth of an isolated perturbation and is useful for defining a number of characteristics of the collapsed structures we expect to observe, while the Zeldovich approximation attempts to follow the formation of structures using a kinematical approach, essentially evolving the particle's along their initial velocity vector.

Zeldovich approximation

In the Zeldovich approximation (Zel'Dovich, 1970) we attempt to follow the formation of structure by identifying the initial displacement of a particle and assuming it will continue to move in that direction. We write the proper coordinate of a particle as

$$\mathbf{x}(t) = a(t)\mathbf{q} + b(t)\mathbf{f}(\mathbf{q}), \tag{1.36}$$

where $\mathbf{q}$ is the Lagrangian coordinate of the particle which at time $t = 0$ is simply the particles comoving coordinate, $\mathbf{f}(\mathbf{q})$ is the time independent displacement field and $b(t)$ simply scales the displacement field. If we neglect the second term this is simply a Hubble expansion.

In the Lagrangian coordinate system, the density is by definition constant in time. To convert to the usual Eulerian density we use the Jacobian of the transformation between $\mathbf{x}$ and $\mathbf{q}$ to give

$$\frac{\rho}{\rho_0} = \left[\left(1 - \frac{b}{a}\alpha\right)\left(1 - \frac{b}{a}\beta\right)\left(1 - \frac{b}{a}\gamma\right)\right]^{-1}, \tag{1.37}$$

where α, β and γ are the eigenvalues of the tensor $\partial f_i/\partial q_i$. We see then that the collapse of an overdense region occurs along the axis with the largest positive eigenvalue. If only one of the eigenvalues is negative the region will collapse to form a sheet of matter known a pancake. If a further eigenvalue is negative the collapse can continue along two axes to form a filament and if all three eigenvalues are negative then the collapse occurs along each axis, forming a spherically collapsed structure known as a halo.

The Zeldovich approximation works well, out performing Eulerian linear theory in accuracy, but eventually breaks down when the particles streams' cross each others' paths. At this point, where gravity would first halt and eventually reverse the direction of particles, in the Zeldovich approximation they continue along the same path. The result is that bound structures fly apart and the approximation becomes unreliable. Aside from illustrating the process of collapse, the Zeldovich approximation has an important practical application. The Zeldovich approximation can be re-written as

$$\mathbf{x}(t) = a(t)[\mathbf{q} + D(t)\mathbf{f}(\mathbf{q})], \tag{1.38}$$

where $D(t)$ is the linear density growth factor. This is an incredibly useful expression and it allows us to set up the initial particle conditions for n-body simulations in the following way. Given particles distributed randomly within cells, we can displace them from their positions $\mathbf{q}$ by a factor $D(t)\mathbf{f}(\mathbf{q})$ to obtain a particle distribution which conforms to a density field with a given power spectrum. This is accomplished by noting that the Fourier modes of the displacement field, $\mathbf{f}_k$ are related to the Fourier modes of the density field, δ_k as

$$\mathbf{f}_k = -i\frac{\delta_k^2}{k^2}\mathbf{k}. \tag{1.39}$$

.

The Spherical collapse model

The Zeldovich approximation shows us that the inevitable fate of overdense regions of the Universe is to form collapsed structures, with the most dense of these forming dark matter halos. It does not tell us what happens after collapse, as

structures will fly apart after first crossing. To gain further insight into the process of structure formation we appeal to a simple model of an isolated spherical overdensity. This problem is entirely analogous to a closed Universe which has the parametric solution

$$r = A(1 - \cos(\theta))t = B(\theta - \sin(\theta)) \tag{1.40}$$

where $A = GMB^2$. Expanding these relations to fifth order in θ gives

$$r \simeq \frac{A}{2}\left(\frac{6t}{B}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{6t}{B}\right)^{2/4}\right]. \tag{1.41}$$

From this we can identify the density within the sphere as

$$\delta \simeq \frac{3}{20}\left(\frac{6t}{B}\right)^{2/3}. \tag{1.42}$$

At early times the overdensity expands with the Hubble flow until it reaches turn around, the point at which the radius is briefly stationary. This occurs when $\theta = \pi$ and $t = \pi B$ giving a density enhancement relative to the background of around 5.55. Gravity then forces the sphere to collapse in on itsself potentially to a singularity at the halos' centre. This does not happen in practice as small offsets of particles from a true sphere will cause the kinetic infall motion of the particles to be converted into random thermal motions. Once the condition for equilibrium, the virial theorem (ie the potential energy of the system is twice the kinetic) , is met the halo becomes a stable body. This occurs at $\theta = 3\pi/2$ and corresponds to a density enhancement of 147.

A key use of this model is to define the density in linear theory which corresponds to the collapse threshold of a halo. By taking the time of collapse in the model and computing the equivalent linear density perturbation, we can define a collapse threshold $\delta_{sc} \simeq 1.63$. As we will see this is useful when we want to calculate the abundance of haloes. We can use linear theory and simply identify haloes are regions which have a density greater than the collapse threshold.

1.4.2 Halo density profiles

In numerical simulations haloes are found with a wide range of shapes for a given mass and can be irregular. However when spherically averaged they are

well described by simpler density profiles. One of the most successful and widely adopted of these is the Navarro Frenk and White (NFW) profile Navarro et al. (1997):

$$\frac{\rho(r)}{\overline{\rho}} = \frac{\delta_c}{(r/r_s)(1+r/r_s)^2}, \tag{1.43}$$

where $r_s = r_{200}/c$ is the scale radius, r_{200} is the radius at which the average density of the halo is $200\overline{\rho}$ and c is the concentration. $\overline{\rho}(z)$ can either be taken as the mean density of the Universe or the critical density for collapse obtained from the spherical collapse model. In this work we will take $\overline{\rho}$ to be the background (mean) density of the Universe. The main motivation for this is that we want to compare with halos identified from simulations using a Friends of Friends algorithm which links particles together with a linking length designed to find structures with a contrast of 200 times the mean background particle distribution. Each halo is therefore described as a slowly varying power law with a shallow inner slope and a steeper outer with the transition radius controlled by the concentration c.

Massive haloes are typically more centrally concentrated than smaller ones, the general trend is described by Bullock et al. (2001) as

$$c(m) \approx \frac{9}{(1+z)}\left(\frac{m}{m_*}\right)^{-0.13} \tag{1.44}$$

Obviously there is a large variation in the actual shape of haloes of a given mass but this can be broadly taken in to account by having a lognormal distribution with variance $\Delta(\log c) \approx 0.2$.

$$p(c|m,z)dc = \frac{d\ln c}{\sqrt{2\pi\sigma_c^2}}\exp\left(\frac{\ln^2[c/\bar{c}]}{2\sigma^2_{\ln c}}\right) \tag{1.45}$$

In this work we will treat, unless otherwise stated, all haloes as having an NFW profile with a concentration selected with the mean and variance described by the distribution given above.

Both the concentration and overall normalisation of the density profile are correlated with the formation redshift of the halo. This usually refers to the epoch at which the halo had assembled half its present day mass. The overall normalisation of the profile, δ_c, is related to the formation redshift of the halo as

$$\delta_c = 3000(1+z_f)^3. \tag{1.46}$$

1.4.3 Abundance

The most fundamental question we can ask about dark matter haloes is how many of them should we expect to find in our Universe of a given mass. This amounts to counting the peaks in the density field which are greater than some threshold for collapse δ_c when the density field is smoothed over a scale R_f . As the density field is Gaussian, the probability that a region has attained this collapse threshold is

$$p(\delta > \delta_c | R_f) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\delta_c}{\sqrt{2}\sigma(R_f)}\right)\right], \tag{1.47}$$

where $\sigma(R_f)$ is the rms fluctuations on the filtering scale. These regions will have therefore have a mass of $M \sim \rho_0 R_f^3$. Press and Schechter (1974) argued that this could be taken to be proportional to the probability that a given point has even been part of a collapsed object of scale R_f. This argument tends to only count those regions which are overdense and therefore misses half the mass in the Universe. To account for this we multiply by a factor of 2. Therefore the fraction of the Universe which is contained in objects with a mass $> M$ is

$$F(> M) = 1 - \mathrm{erf}(\nu/\sqrt{2}), \tag{1.48}$$

where we have defined

$$\nu = \frac{\delta_c}{\sigma(M)} \tag{1.49}$$

Another way of arriving at the PS result is to consider a point in the density field and calculate the density when smoothed over a region R_f around it. If we take the filtering function to have a sharp truncation in k space then as we decrease the filtering scale from $R_f = \infty$ at which point $\delta = 0$ we are simply adding new spherical shells in k space. As the field is Gaussian the trajectory of δ will therefore execute a random walk. We want to find is the first value of R_f at which it crosses the collapse threshold. After initially crossing the threshold the field can increase or decrease on its random walk. However, this doesn't matter as the point will still be associated with an object of the first crossing scale. After crossing the threshold the field is just as likely to increase as decrease and all points which lie over the threshold at some point R_1 must have previously crossed at a higher value of R. The survival probability is then

$$\frac{dP_s}{d\delta} = \frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{\delta^2}{2\sigma^2}\right) - \exp\left(-\frac{(\delta - 2\delta_c)^2}{2\sigma^2}\right)\right] \tag{1.50}$$

giving

$$1 - P_s = 1 - \mathrm{erf}\left(\frac{\delta_c}{\sqrt{2}\sigma}\right) \tag{1.51}$$

The abundance of haloes is usually quoted in the form of the multiplicity function which is the fraction of mass contained by objects in a unit range of $\ln M$. We can write the multiplicity function in a general form:

$$\frac{M^2 f(M)}{\rho_0} = \left|\frac{d\ln\sigma}{dlnM}\right| F_{ps}(\nu) \tag{1.52}$$

where

$$F_{ps} = \frac{2}{\pi}\nu \exp\left(-\frac{\nu}{2}\right) \tag{1.53}$$

In the PS approach the threshold is taken to be a constant corresponding to linear collapse threshold we obtained previously from the spherical collapse argument :

$$\delta_{sc} = 1.69(1+z). \tag{1.54}$$

However in comparisons with numerical simulations this prescription tends to over predict the abundance of low mass haloes while under predicting the number of high mass ones. A number of authors have attempted to improve on the PS result by considering different forms for the collapse threshold. Most notably Sheth et al. (2001) considered the effect of using a threshold corresponding to elliptical collapse which results in replacing F_{ps} in eqn 1.52 with the form:

$$f_{st}(\sigma) = A\sqrt{2a}\pi\left[1 + \left(\frac{1}{2\nu^2}\right)^p\right]\nu \exp\left(-\frac{a\nu^2}{2}\right) \tag{1.55}$$

On fitting this form to numerical simulations we obtain the best fit parameters $A = 0.3222$, $a = 0.707$ and $p = 0.3$. This achieves a much better match to what is seen in numerical simulations. Figure 1.7 shows a comparison of the these two multiplicity functions.

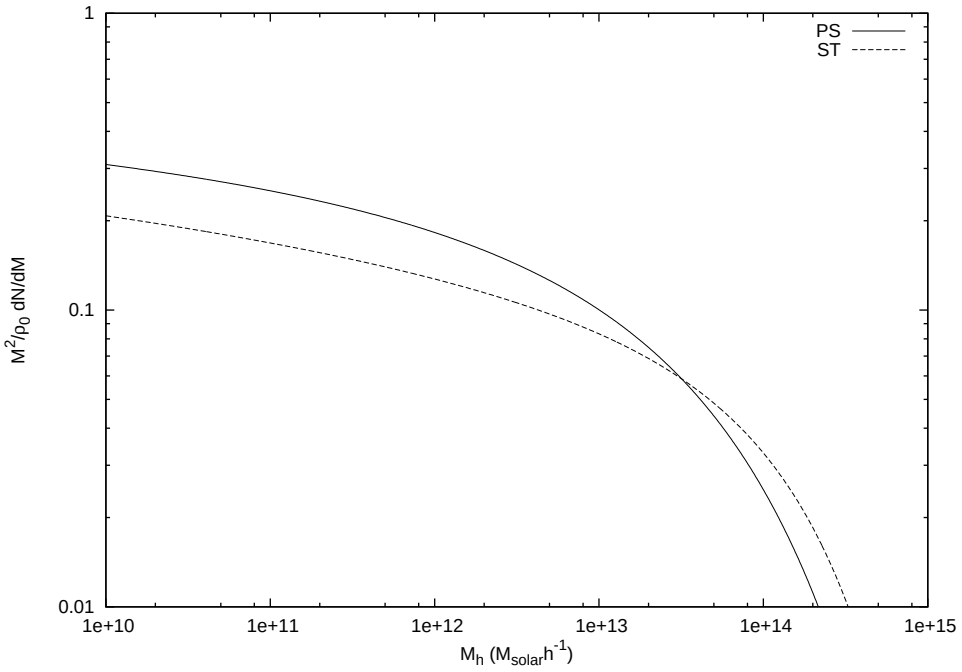

Figure 1.7: The Press-Schecter and Sheth - Torman $z = 0$ multiplicity functions

1.5 Peak background split

As gravity acts on the early smooth Universe it causes over-dense regions to collapse though successive spatial dimensions to form sheets, filaments and finally gravitationally bound dark matter haloes. On small scales then the non-linear dynamics of these structures become hard to follow and we require high resolution particle simulations to accurately calculate their structure and evolution. Even at present redshift however the large scale dynamics are largely linear. We can treat these two components separately as a background of near linear large scale fluctuations and small scale density peaks.

As I discussed previously, the dark matter field can be described as a superposition of Fourier modes with random phase and amplitudes given by the power spectrum P(k):

$$\delta(x) = \sum_k \delta_k e^{-ix\cdot k} \tag{1.56}$$

It is helpful to think of this field as consisting of two components: the high frequency fluctuations which are boosted by a low frequency background as shown schematically in Figure 1.8

In standard cosmological simulations the strategy is to use a large number of simulation particles to map these small high density fluctuations. An alternative approach which both PTHALOS (described below) and my own algorithm(described in Chapter 3) take is to describe the statistics of these fluctuations relative to a smooth background, developing a Monte Carlo approach to sampling these statistics.

1.5.1 Conditional Mass Function

We have examined expressions for the mass functions of haloes which describe the expected number of haloes of a given mass in the Universe. The mass function only applies to the halo population when averaged over scales where the density field could be considered uniform, in smaller local volumes with an over-density δ, the halo mass function will not in general be a simple scaling of the universal

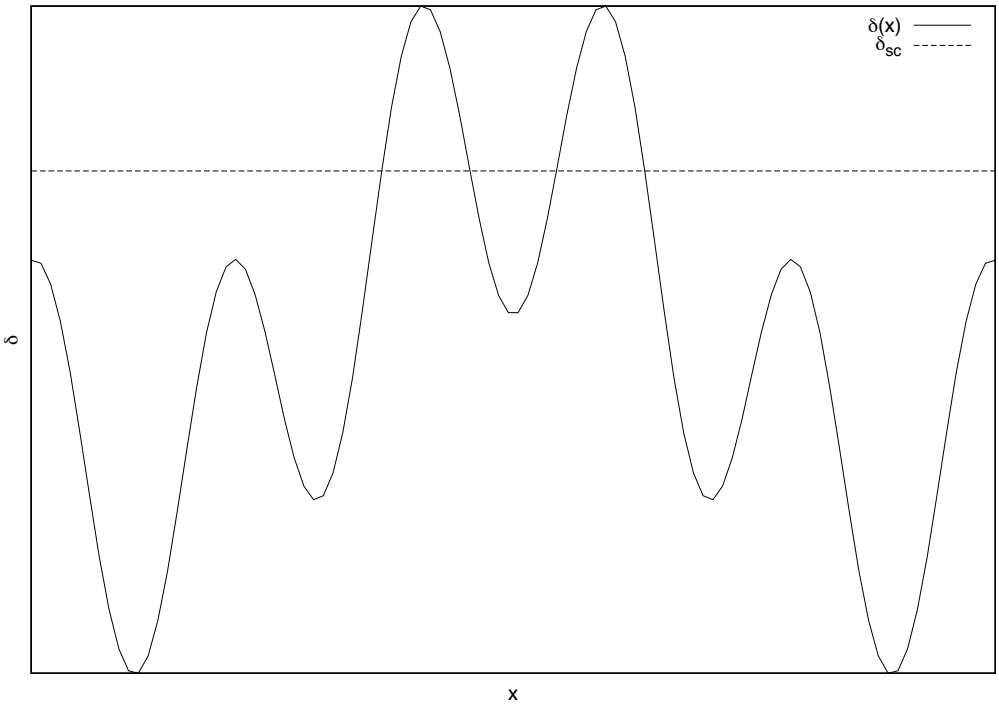

Figure 1.8: An illustration of the peak background split. The small scale peaks in the density field which represent the locations of dark matter haloes are boosted by a slowly varying background. In regions where this boost is present the peaks will reach the collapse threshold (the dotted line) earlier than similar small scale peaks in underdense regions.

one. If we wish to accurately populate the density field then we need to describe these local variations in the distribution of haloes.

Formally the statistic we seek is the conditional mass function $\mathrm{CMF}(\mathrm{M_h}|\delta)$, the number of haloes of mass M_h which reside in a region of space with a background density contrast of δ.

If a region of space with a density $\delta > 0$ is considered, then the local small scale fluctuations which will collapse to form haloes are embedded within a larger scale density fluctuation effectively boosting the small scales' fluctuations' density compared with that of a low density region. This boost implies that small fluctuations within a dense region will cross the critical threshold for collapse δ_{sc} before similar fluctuations in under-dense regions. These regions therefore form their haloes earlier in the history of the Universe, and once formed, these haloes have more time to merge to produce higher mass haloes.

This way of thinking has led a number of authors (Bond et al., 1991; Lacey and Cole, 1993) to adapt the universal mass function of haloes to a local Conditional Mass function through a rescaling of the variance and collapse criteria:

$$\delta_{sc}(x) \rightarrow \delta(x) + \delta_0(x) \tag{1.57}$$

$$\sigma^2(x) \rightarrow \sigma^2(x) - \sigma_0^2(x) \tag{1.58}$$

where $\delta_0(x)$ is the density of the cell and σ_0 is the variance of fluctuations within the cell and d_{sc} is the collapse threshold in the spherical collapse model. For example the conditional Press-Schecter mass function becomes:

$$CMF_{ps}(m|\delta_0, \sigma_0) = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{m} \left| \frac{d\sigma}{dm} \right| \frac{\sigma(\delta_{sc} - \delta_0)}{(\delta^2 - \delta_0^2)^{3/2}} \exp\left\{ -\frac{(\delta_{sc} - \delta_0)^2}{(\sigma^2 - \sigma_0^2)} \right\}. \tag{1.59}$$

This scaling is valid for mass functions which are related to the crossing of a constant barrier. However, a number of Universal Mass Functions (UMF's) are more accurate than that of Press Schecter. In recent work by Rubiño-Martín et al. (2008), the authors develop a more general approach to rescaling.

Later we will want to compare these theoretical predictions with those measured from simulations, and so we need a way of relating the non-linear densities measured in the simulations to the linear quantities used in excursion set theory and thus the expressions above. If the density of a region of a simulation is given

as : $1 + \delta = M/(V\bar{\rho})$ and δ_0 is the density that we would have expected from linear theory, the two are related by the fitting formula proposed by Mo and White (1996) and modified by Scoccimarro and Sheth (2002) :

$$\delta_0 = \left[1.68647 - \frac{1.35}{(1+\delta)^{2/3}} - \frac{1.12431}{(1+\delta)^{1/2}} + \frac{0.78785}{(1+\delta)^{0.58661}}\right]\frac{\delta_{sc}(z)}{1.68647} \tag{1.60}$$

where δ_0 is the density we would expect in linear theory and δ_{sc} is the spherical collapse density threshold (which introduces a dependence on cosmology).

Halo Bias

As a result of the dependence of halo properties on local density, the haloes of different masses will be biased tracers of the underlying mass field. The most massive haloes will typically live in the rarest peaks of the density field while smaller massed halos are more evenly distributed. This affects the clustering of haloes and the galaxies they contain. The number of haloes within a region of density δ, $n(M, \delta|z)$ can be modelled as (Cole and Kaiser, 1989; Mo and White, 1996; Sheth and Tormen, 1999)

$$n(M, \delta|z) = [1 + b(M|z)\delta]n(M|z), \tag{1.61}$$

where $b(m|z)$ is bias of a halo of mass m and $n(M|z)$ is the universal mass function. In the ellipsoidal collapse model$b(m|z)$ can further be shown to be

$$b(m|z) = 1 + \frac{\alpha\nu^2 - 1}{\delta_{sc}(z)} + \frac{2p/d_{sc}(z)}{1 + (a\nu^2)^p}, \tag{1.62}$$

where $\nu = \delta_{sc}/\sigma(M)$ δ_{sc} is the collapse threshold from the spherical collapse model, $\sigma(M)$ is the variation of the density field on scales corresponding to M", and p and a are free parameters to be fit to simulations (Cole and Kaiser, 1989)Mo and White (1996)Sheth and Tormen (1999).

This has a big implication for how we think about tracers of the dark matter field. Often it is easiest to calculate predictions for the dark matter in our Universe, but in reality we observe galaxies which inhabit haloes and so require a clear understanding of the bias between these two populations to be able to link observations and theory.

1.6 Galaxies and haloes

The current model of galaxy formation was put forward by White and Rees (1978). They argue, that hot baryonic gas in the Universe requires a large potential well in which it can radiate and cool to form stars and galaxies. They suggest that dark matter haloes provide such a place and so galaxies should form within and be associated with dark matter haloes.

The population of galaxies within haloes is described by the halo model. The basic assumption of which is that the luminous contents of a halo is a function of haloes' mass alone. The number of galaxies per halo and their luminosity distribution are modeled by the halo occupancy $N(M)$ and the conditional luminosity function $CLF(M, mag)$, the total number of haloes within each halo and the distribution of their absolute magnitudes. From such statistics it is possible to predict galaxy clustering and by comparing these predictions to real world measurements it is possible to determine the parameters of the halo model Cooray (2006)

If we consider a limiting luminosity: L_{cut} then we expect the number of galaxies brighter than this that a halo contains to increase with its mass. There will also be a mass at which haloes will simply lack the baryon content to produce a L_{cut} galaxy. Larger haloes will have enough gas to have collapsed a single blue galaxy at their cores with larger haloes still having multiple satellite galaxies.

We follow Cooray and Milosavljević (2005) and Cooray (2006) who proposed the following model for the conditional luminosity function. In this model central galaxies (the dominant galaxy which resides at the centre of the dark matter halo) and satellite galaxies (smaller galaxies which orbit within the potential well of the halo) are treated separately as $\Phi^{cent}(L|M,z)$ and $\Phi^{sat}(L|M,z)$:

$$\Phi(L|M,z) = \Phi^{cent}(L|M,z) + \phi^{sat}(L|M,z) \quad (1.63)$$

$$\Phi^{cent}(L|M,z) = \frac{f_{cen}(M,z)}{\sqrt{2\pi}\log_{10}(\sigma_{cen}L)}) \quad (1.64)$$

$$\times \exp\left(\frac{\log_{10}[L/L_c(M,z)]^2}{2\sigma_{cent}}\right) \quad (1.65)$$

$$\Phi^{sat}(L|M,z) = A(M,z)L^{\gamma(M)}g_s(L) \quad (1.66)$$

where $f_{cen}(M, z)$ is the selection function which accounts for the efficiency of galaxy formation taking in to account the inability of some low mass haloes to produce a galaxy at all

$$f_{cent} = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\log(M) - \log(M_{cent-cut(z)})}{\sigma}\right)\right], \tag{1.67}$$

and g_s is

$$g_s(L) = 0.5\left[1 + \text{erfc}\left(\frac{\log(L_c/2.0) - log(L)}{\sigma_s}\right)\right] \tag{1.68}$$

$M_{cent-cut}$ takes a numerical value of $10^{10}M_{sun}$ and $\sigma = 0.5$

$$L_c = L_0(1+z)^{\alpha}\frac{(M/M_1)^a}{[b + (M/M_1)^{cd(1+z)^{\nu}}]^{1/d}} \tag{1.69}$$

For the normalisation of the satellite portion of the formula see Cooray (2006). Figure 1.9 shows the conditional luminosity function for a number of different halo masses:

1.7 Galaxy Surveys

To test the predictions of the theories described in this chapter and of the ΛCDM cosmology as a whole we need to compare these results with the Universe around us. The primary way this has been achieved is through galaxy redshift surveys. Despite their dominant role in the formation of structure and evolution of the Universe, dark matter and dark energy are not visible to us. The more physically complex luminous baryons in our Universe are the most readily available probes of structure we have out our disposal. To that end, a number of large galaxy surveys have been undertaken to map the Universe around us and provide data to test against theories.

The challenge of galaxy surveys is to find as many galaxies as possible in the sky and determine their position on the sky and their distance from us. Obtaining the location of a galaxy on the sky is relatively straight forward but to obtain a measure of the distance of a galaxy from us is harder. We cannot measure the distance directly and so have to infer the distance of the galaxy from other factors. The most widely used is that of redshift. Due to the expansion of the

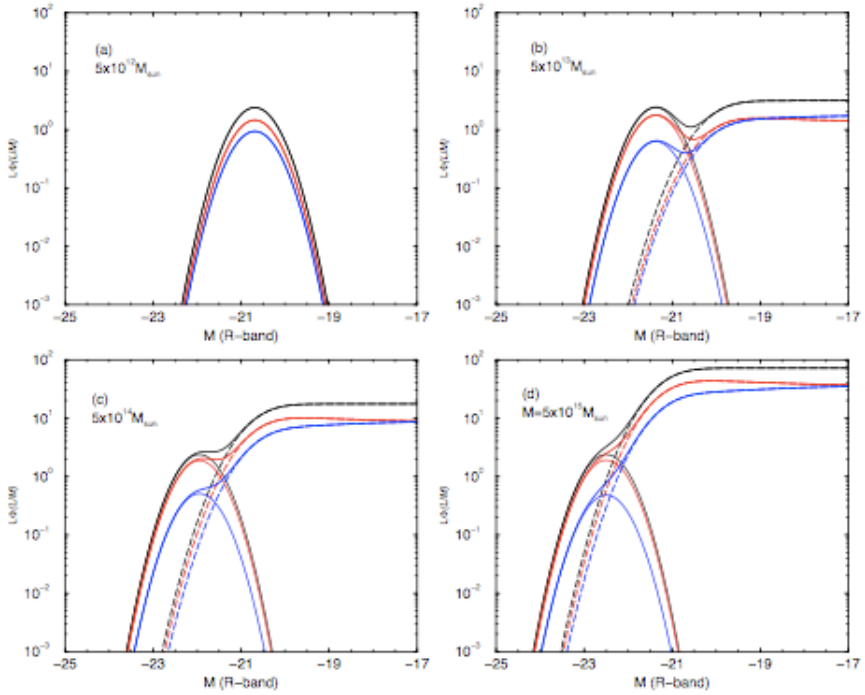

Figure 1.9: Conditional luminosity functions used by Cooray (2006). The black line shows the entire galaxy population, while solid blue and red lines represent the contribution of blue and red central galaxies and dotted lines their satellite equivalents.

Universe, there is a direct relation between the recessional velocity (measured as its redshift) of a galaxy and its comoving distance from us

$$R_0 dr = \frac{c}{H(z)} dz \tag{1.70}$$

$$= \frac{c}{H_0} \left[(1-\Omega)(1+z)^2 + \Omega_v + \Omega_m (1+z)^4\right]^{-1/2} dz \tag{1.71}$$

where we have used the expression for the Hubble parameter in equation 1.18 and taken $\Omega_r << 1$.

It is also useful to be able to calculate various physical properties of an object at a given redshift. The two most important of these quantities are relations which allow us to calculate the proper transverse size of an object, and the relation between the flux density and luminosity of an object at redshift z. The proper transverse size of an object with comoving size $S_k(r)d\phi$ is given by

$$d\ell = \frac{d\phi R_0 S_k(r)}{(1+z)} \tag{1.72}$$

while the relation between the luminosity of a galaxy and its flux density is given by

$$S_\nu(\nu_0) = \frac{L_\nu([1+z]\nu_0)}{4\pi R_0^2 S_k^2 (1+z)}, \tag{1.73}$$

where ν_0 is the frequency of the emitted photon, and L_ν is measured in units of WHz^{-1}. The second relation is valid only for isotropic emission and takes into account not only the effect of the increased proper surface area over which the radiation has spread, $4\pi[R_0 S_k(r)]^2$, but also the effect of redshift on the photon.

Form these two equations we can define two effective distances, the angular diameter distance and the luminosity distance :

$$D_A = (1+z)^{-1} R_0 S_k(r) \quad \text{angular-diameter distance} \tag{1.74}$$

$$D_L = (1+z) R_0 S_k(r) \quad \text{luminosity distance.} \tag{1.75}$$

The main complication with redshift surveys is that we do not measure the position of galaxies exactly but rather we measure their recessional velocity and imply a distance from the Hubble flow. Galaxies also have peculiar velocities which also contribute to the redshift giving

$$(1+z_{\mathrm{obs}}) = (1+z_{\mathrm{cos}})(1+z_{\mathrm{vel}}), \tag{1.76}$$

where $z_{\rm cos}$ is the redshift induced by the Hubble flow and $z_{\rm vel}$ that of its peculiar velocity. This leads to an incorrect estimate of galaxy distance which is more pronounced in regions of high peculiar velocity, usually around large clusters. Figure 1.10, a slice of the 2dFGRS survey (Colless, 1999), illustrates this effect well. Along the radial direction the position of galaxies in clusters have a tendency to be smeared out, giving distinctive "fingers of god" features.

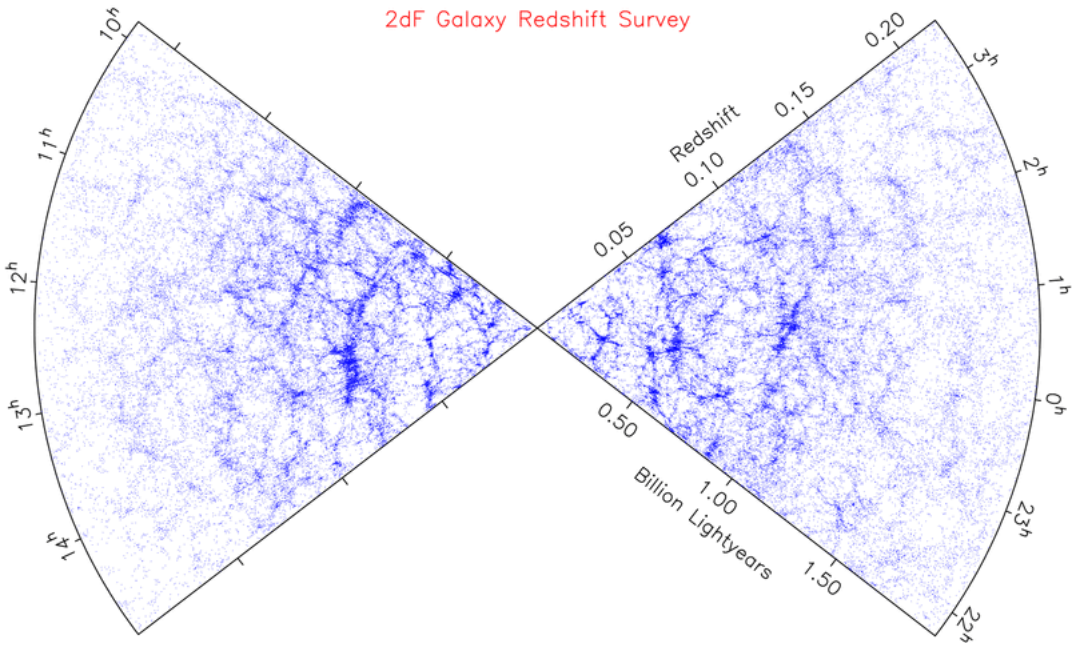

Figure 1.10: A slice through the 2dFGRS (Colless, 1999) survey. The cosmic web can clearly be seen along with the radial smearing of galaxy positions due to the peculiar velocities of galaxies

.

1.8 Numerical Simulations

Theoretical models can help us predict the statistics of the halo and galaxy population, but as we have seen in the previous section the practicalities of measuring these quantities in the real Universe can often complicate things. It is therefore useful, both to check the predictions of theoretical results and to better approximate what is actually being measured in galaxy surveys, to use computer modelling to produce realisations of the Universe. As the physics of baryonic

matter is complicated and heavily dependent on the cosmic web placed down by dark matter, the natural starting point is to simulate a Universe in which dark matter is the only dynamic component. The dominance of dark matter over baryonic matter, means that it is the primary driver of structure formation after pressure support for baryons is lost at recombination. This is fortunate as the weakly interacting properties of dark matter simplify its dynamics considerably: with no way in which to interact with the electromagnetic field, dark matter can be treated as a collisionless fluid with gravity as the only relevant force. On a discrete system such as modern computers we sample this fluid with a number of simulation particles and evolve these particles through a number of small discrete time steps.

As with all numerical work, we require that the quantities involved in the simulation, the particle comoving positions and peculiar velocities, are dimensionless. The natural way of obtaining this is to scale the particle position by the width of the simulation box L, the scaling choice for the peculiar velocities are less clear and in this work we take them to be in units of the expansion velocity across the box HL (where H is the Hubble factor) to give:

$$\mathbf{X} = \mathbf{x}/\mathbf{L} \qquad \mathbf{U} = \mathbf{u}/\mathbf{HL} \tag{1.77}$$

In these co-ordinates the equation of motion becomes:

$$\frac{d}{d\ln a}(f(a)\mathbf{U}) = \frac{3}{8\pi}\Omega_m(a)f(a)\frac{1}{N}\sum_i \frac{\mathbf{X}_i - \mathbf{X}}{|\mathbf{X}_i - \mathbf{X}|^3} \tag{1.78}$$

where $f(a)$ is proportional to $a^2H(a)$ and has arbitrary normalisation (which we take to be unity at the initial epoch). The largest computational task involved in evolving the density field is the calculation of forces. We need to compute the contribution of the force on a given particle over every other particle in the simulation. This is a processes which has computational complexity of $O(N^2)$: if we double the number of particles in a simulation we increase the run time by a factor of 4. This is clearly prohibitive to running simulations with high enough resolution to resolve detailed structures. A number of methods exist to overcome this N^2 problem and we discuss the workings of an advanced code, GADGET2, in chapter 2. Most of the simulations which we will run in this thesis will use a simpler scheme than that of GADGET2 which it is worth detailing here.

Instead of calculating the force directly we can implement what is known as a particle mesh (PM) code. A PM code (Efstathiou et al., 1985) utilises the existence of fast algorithms for computing Fourier Transforms to turn the problem from $O(N^2)$ to a more manageable $O(N \log N)$. The particle field is first binned on to a mesh of a given size N producing a density field. This is then Fourier transformed, the three Fourier components of force computed and then finally transformed back to real space. The resulting force field can then be interpolated to the position of each particle and their velocities and position updated appropriately. This is repeated until we reach the desired epoch.

The other ambiguous step is in how we assign the mass of each particle to the grid. There are many different schemes to accomplish this, the simplest of which is simply to assign the entire particle to the nearest grid point. As ultimately the resolution of the simulation is governed by the resolution of the mass/force mesh, this simple approximation is sufficient. In other cases within this thesis, we will want to map mass to a regular grid with more care. In these cases we will typically use the cloud in cell (CIC) scheme which distributes the mass of the particle between the surrounding cells. If i and j are cell indices, and x,y and z are the particles' position such that $x_i < x < x_{i+1}$, $y_j < y < y_{j+1}$ and $z_k < z < z_{k+1}$, then the (CIC) algorithm maps mass as

$$M_{i,j,k} = \frac{M_p}{dxdydz}|(x - x_i)||(y - y_j)||(z - z_k)|, \tag{1.79}$$

where $dx = x_{i+1} - x_i$ etc, $M_{i,j,k}$ is the mass assigned to the cell (i, j,k) and M_p is the mass of the particle.

This smooths the particle out between cells, which is desirable as each particle actually represents not a point but a region of space.

1.8.1 Initial Conditions

Having described a method to evolve the particles we need to set up their initial positions and velocities. As the linear density field is Gaussian this consists of ensuring the initial field has the correct power spectrum at the initial redshift. As we discussed before the Zeldovich approximation of structure formation gives a good way of accomplishing this. If the initial redshift is high enough then the

initial displacements of the particles are simply proportional to the velocities :

$$\delta\mathbf{X} \propto \mathbf{U} \tag{1.80}$$

and the Zeldovich approximation relates the modes of the displacement field $\mathbf{f}$ to the power spectrum through:

$$\mathbf{f}_k = -i\frac{\delta_k}{k^2}\mathbf{k} \tag{1.81}$$

Given a power spectrum it is possible to populate a grid in Fourier space at which each mode has a power given by the power spectrum and random phase. This can then be Fourier transformed to obtain a displacement field. Particles placed at random within cells can then be displaced by the correct amount to obtain a particle population with the correct power spectrum.

1.8.2 Hierarchical clustering and Semi-analytics

Over the years the standard cosmological model has become that of ΛCDM or cold dark matter with dark energy. If the dynamic mass in the Universe is in the form of cold dark matter, then in the early Universe it undergoes very little free streaming and small scale fluctuations in the primordial power spectrum are preserved (Eisenstein and Hu, 1998). In this scenario it is the smallest fluctuations which are the first to undergo gravitational collapse and, these collapsed haloes merge to form larger and larger structures. This bottom up formation of structure if known as the hierarchical clustering model and it has seen a number of great successes. Most notable of these is the incredibly accurate predictions that ΛCDM makes for power spectrum of the CMB. Figure 1.11 show the tight agreement between the theoretical predictions and the CMB power spectrum as measured by the WMAP satellite. Another area where ΛCDM has produced good results is that of galaxy clustering. Again ΛCDM agrees to a high level of accuracy with large surveys such as the SDSS and 2dFGRS, which measure the clustering spectrum of galaxies in the Universe. The agreement of the model in both the early and late Universe and over so many orders of magnitude, as depicted in Figure 1.11 is compelling evidence.

While the ΛCDM model is successful at recovering the characteristics of the large scale structure of the Universe, there is still some debate about how well it

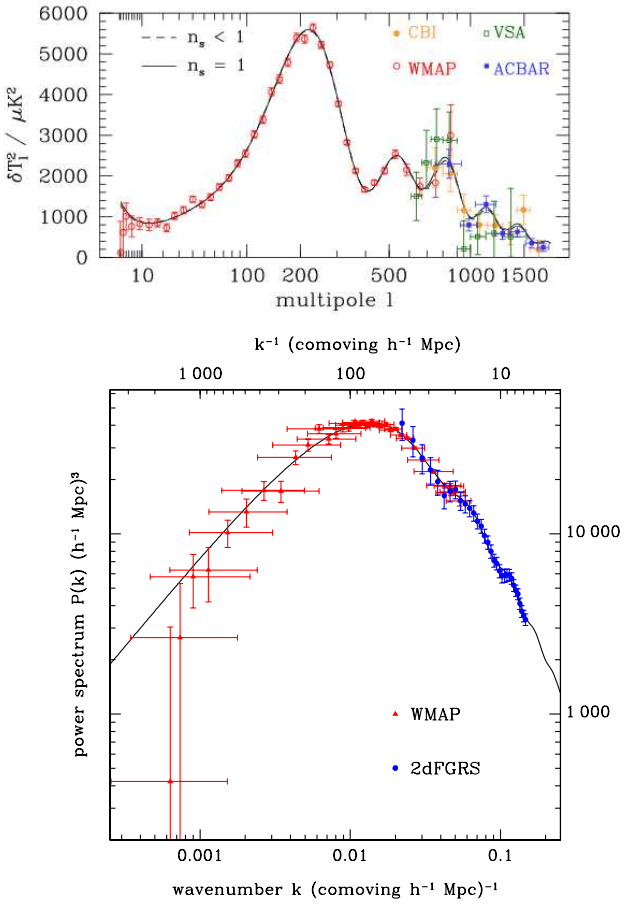

Figure 1.11: Top: The power spectrum of fluctuations in the CMB (Hinshaw et al., 2003; Readhead et al., 2004; Dickinson et al., 2004; Kuo et al., 2004) and the best fitting CDM model. Bottom: The power spectrum of density fluctuations measured from the 2dFGRS (Cole et al., 2005) and WMAP combined on the same plot. Plot taken from (Sánchez et al., 2006)

.

performs on the smaller scales. The two main challenges are the abundance of sub-haloes within galactic sized halos (also known as the satellite problem) and the shape of the inner profile of dark matter haloes (the "cuspy core" problem).

In high resolution simulations of dark matter, galactic sized halos have a tendency to form a large number of sub-haloes which should correspond to a large number of satellite galaxies. The observed number of satellite galaxies around the Milky Way and Andromeda is over an order of magnitude less than those found in simulations (Moore et al., 1999a) .

The "cuspy core" problem can be seen when looking at the rotations curves of low surface brightness galaxies which are dominated by dark matter. The rotation curves from these galaxies fail to match predictions of CDM haloes, implying that the CDM haloes are too cuspy (Moore et al., 1999b).

While both of these issues pose challenges to the CDM model, on these small galactic scales baryonic matter has an important role to play. It may be, as some authors (Benson et al., 2002b,a; Somerville, 2002) have suggested, that these issues might be solved with a better treatment of the astrophysics of these environments.

It is easy to study the hierarchical build in dark matter using simulations however following the luminous matter in the Universe poses a tougher challenge. While dark matter interacts only through the gravitational force, baryons in the Universe take part in many processes including shock heating, star formation, supernovae , ionisation and turbulence. Producing a simulation that includes all these effects accurately while still maintaining a high enough resolution to be able to resolve galaxies in a cosmological context is a hard task.

Instead an alternative approach known as Semi-Analytics has become popular. Instead of following the gas particles directly, Semi-Analytics uses numerical approximations and a description of the merger history of a dark matter halo to try to calculate the properties of the luminous material it contains. The sets of equations used to accomplish this are obtained either from theory or from scaling relations seen in real galaxies, and can contain a number of free parameters which can be tuned to match the present dat Universe.

One of Semi-Analytics notable successes has been to explain the shape of the galaxy luminosity function in the local Universe. A simple prediction of the

galaxy luminosity function can be achieved by taking the halo mass function and applying a constant light to mass ratio. Doing this severely over predicts the abundance of small and large galaxies. Semi-Analytic models, such as Croton et al. (2006), which incorporate methods of feedback, notably supernova and AGN, can suppress the formation of galaxies on both large and small scales and provide an excellent reproduction of the local galaxy luminosity function.

While Semi-Analytics have been good at predicting properties of galaxies in the low redshift Universe, they are increasingly under attack from observations at high redshift. A number of authors (Drory et al., 2005; Glazebrook et al., 2004) have begun to see massive red galaxies at high redshift. This poses a problem for Semi-Analytics and the hierarchical clustering model in general. Such large galaxies should take time to merge and so are unlikely in the high redshift Universe. A number of models have been proposed to deal with these observations (Nagamine et al., 2004, 2005), however they have yet to match the high redshift and low redshift results at the same time.

1.8.3 PTHALOS

The problem of the prohibitive expense of computing time is not a new one. With each new advance in computing power comes higher resolution simulations which in turn pose questions which require even more computing power. One obvious approach to this problem is to simply hope that the availability of faster computers will increase with time. Other authors have tried to devise algorithms which can achieve equivalent results with less computing power. This is the approach that Scoccimarro and Sheth (2002) have taken in their PTHALOS algorithm. This approach argues that with sufficient knowledge of the large scale fluctuations of the density field, the small scale peaks, the halos and galaxies, can be added using a statistical approach.

To avoid the computationally intensive N-body codes normally required to explore galaxy statistics, the PTHALOS approach instead evolves a set of N particles, where N is the number of dark matter haloes we would expect to find in the simulation volume, from their initial conditions using second order Lagrangian perturbation theory.

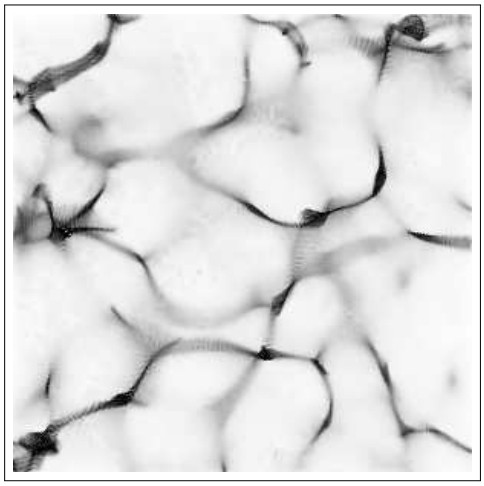

Figure 1. A slice of 2LPT, 150 Mpc/h a side and 6 Mpc/h thick.

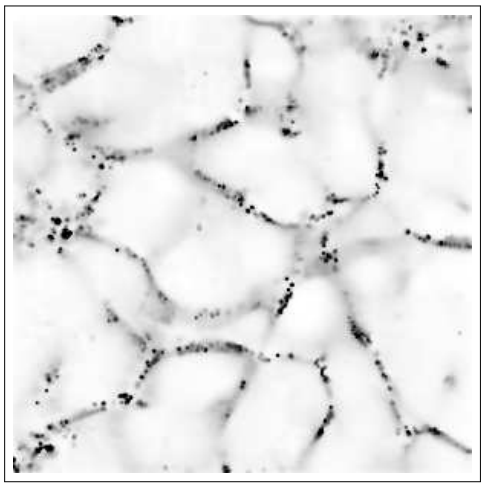

Figure 2. The same slice as previous figure for `PTHalos`.

Figure 1.12: A plot of the density field created using Lagrangian perturbation theory (top) and the mass field reconstructed by the halos created with the PTHALOS method. Taken from Scoccimarro and Sheth (2002)

.

These particles can then be binned to produce a density field as seen in Figure 1.12. This density field contains only the large scale modes. The small scale density fluctuations of haloes can be populated by randomly drawing halo masses from a conditional mass function and associating each mass with a perturbation theory particle. In this way a full description of the density field can be obtained with dramatically less computational effort than an equivalent N-body approach.

The main aim of this thesis is to update the PTHALOS approach to use statistics measured from large modern N-body simulations and semi-analytic methods which are run in tandem with them. We will use coarse N-body simulations instead of perturbation theory and use models fit to simulations, rather than theory, as our statistical description. The philosophy of achieving efficiency by only doing what we have to obtain an accurate and consistent, galaxy and halo catalogue will remain intact.

1.9 Thesis outline

One of the main aims of this thesis is to devise a new method of producing large scale mock galaxy catalogues with modest computational resources to aid in exploring the parameter space of current cosmological models and to produce the large scale mock catalogues needed by next generation surveys. As mentioned before the main inspiration for this work is taken from the work of Scoccimarro and Sheth (2002) and their algorithm PTHALOS. The key idea is to admit that for a lot of work what is required is not a comprehensive attempt to model the detailed physics of galaxy formation but simply a statistically accurate mock galaxy catalogue. With this in mind it makes sense to try to produce an algorithm which uses our knowledge of galaxy and halo statistics to populate a simulation rather than trying to follow the complicated and computationally intensive process of galaxy formation. The loss of knowledge about the individual history of a galaxy and its evolution is balanced by the speed of this approach, allowing larger areas of parameter space to be probed and larger computational volumes to be examined.

Our strategy will be to utilise the excellent work that has been carried out by those working on semi analytics galaxy models. These models (discussed in more detail in Chapter 2) can be used to populate high resolution simulations allowing

us to follow galaxy formation in tandem with structure formation. In Chapter 2 we will use the Millennium simulation (MS) and the semi-analytic galaxy catalogues of DeLucia, publicly available through the MS database (http://www.mpa-garching.mpg.de/millennium/), to construct the relevant statistics of how haloes populate the dark matter field and how galaxies, in turn, populate dark matter halos. We will quantify and provide fits for the conditional luminosity function of galaxies in two populations, red and blue. We then detail an algorithm to scale up these results in Chapter 3, allowing the population of low resolution simulations with haloes and galaxies. As a test of our algorithm we recreate the MS from scratch, starting with a low resolution version of the MS with a particle load 600 times less than that of the MS. We compare the clustering of galaxies and halos in the resimuation and the original MS to determine the accuracy of our method. In Chapters 4-6 we put the algorithm through its paces by populating larger volumes and using these to explore a number of topics. In chapter 4 we analyse the shift in the location of the baryon acoustic oscillation peak for galaxies, halos and dark matter. In chapter 5 we detail how the method can be applied to multiple simulation outputs in order to construct light cones. We construct a $5^\circ \times 5^\circ$ light cone and populate it with haloes and galaxies. We also explore the use of redshift space distortions to probe gravity models and the topology of galaxies, halos and dark matter in the Universe. Finally in Chapter 6 we detail a novel approach to producing lensing maps, involving the dressing of haloes on a lightcone with surface mass density profiles. We compare the results obtained by this method to a dark matter particle only run and show that not only can we recover the convergence spectrum, but that we can extend the dynamic range of this measurement by a number of orders of magnitude without having to resort to higher resolution simulations.

Chapter 2

Millennium Simulation

In this chapter we will discuss a number of statistics calculated from the Millennium Simulation and its associated semi-analytic galaxies. These statistics will ultimately be used to calibrate our halo model for use in populating larger volumes. The required statistics are the typical halo occupancy and variance of a cell of overdensity δ, the conditional mass function, the conditional luminosity function for various types of galaxy and the small scale bias of galaxies relative to the dark matter in their host haloes.

2.1 Simulation Overview

The Millennium Simulation (MS) is a landmark numerical calculation of cosmological structure formation. It uses the publicly available GADGET2 (Springel, 2005) numerical integration code to follow 2160^3 particles in a cubic region of $500\,h^{-1}\,\mathrm{Mpc}$ from a redshift of $z = 127$ until the present day. This unprecedented resolution allows the MS to resolve and follow structures over almost 6 orders of magnitude in mass from the dark matter haloes of dwarf galaxies ($M_h = 10^{10}h^{-1}M_\odot$) through to the largest clusters that are just forming in our Universe ($M_h = 10^{16}h^{-1}M_\odot$).

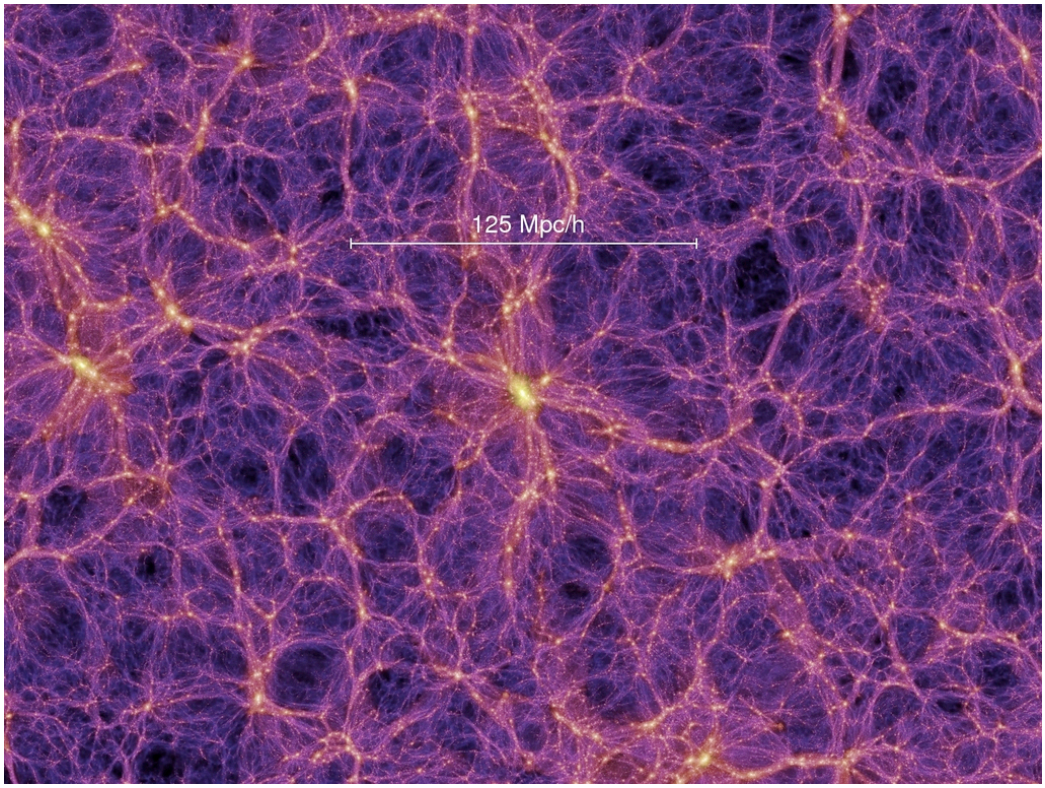

Figure 2.1: A $15\,h^{-1}$ Mpc slice of the Millennium Simulation showing a large dark matter cluster

To calculate the forces on simulation particles the GADGET2 code employs a dual approach. On small scales it adopts a tree code in which the forces are calculated by recursively dividing the volume into sub-regions. Starting with a root node which encompasses the entire simulation volume, each node is examined to determine if the force contribution of the particles it contains can be reasonably approximated by a multipole expansion. This will be true of groups of particles which are either small enough or far away enough from the current particle. If this criterion (usually referred to as the opening angle of the cell), is met then the multipole expansion is used; if it is not, then the current node is subdivided along the halfway point of each of its axis (Barnes and Hut, 1986). The same criteria is then applied to the 8 new daughter nodes. If the algorithm reaches a leaf node, one that contains a single particle, the force is calculated explicitly.

The order of the multipole expansion used within GADGET is a monopole requiring that at each node the total mass and centre of mass vector be stored. The accuracy of the algorithm is determined by the tuning of the opening criteria. The deeper the tree is traversed the more accurate the force calculation becomes. In GADGET2 the opening criteria for a cell containing mass M at a distance r and with extent l is taken to be:

$$\frac{GM}{r^2}\left(\frac{\ell}{r}\right)^2 \leq \alpha|\mathbf{a}|, \tag{2.1}$$

where $|\mathbf{a}|$ is the acceleration calculated in the last iteration and α is used to control the required force accuracy.

On larger scales, GADGET2 uses a more traditional particle mesh scheme in which the forces are calculated by constructing the density field of the volume by binning the particles. Using a cloud in cell (CIC) scheme, each particle's mass is assigned to a point on the mesh which is then Fourier transformed and multiplied by the Green's function for the potential: $-4\pi G/k^2$. To suppress the small scale forces which are being calculated by the tree part of the code, the force is damped at small scales. To account for the cloud-in-cell binning and the subsequent force interpolation the mesh is then divided by two factors of $\mathrm{sinc}^2(k_xL/2N_g)\mathrm{sinc}^2(k_yL/2N_g)\mathrm{sinc}^2(k_zL/2N_g)$. Following a second inverse Fourier transform GADGET2 obtains the potential which is then 4 way differenced to approximate the force in each Cartesian coordinate. The Millen-

Table 2.1: Millennium simulation parameters

σ_8	0.9	Power spectrum normalisation
h_0	0.73	Hubble parameter
Ω_m	0.25	Matter fraction
Ω_b	0.045	Baryon fraction
Ω_λ	0.75	Dark energy
w	-1	Equation of state of dark energy
n_s	1	Spectral index of primordial power spectrum

nium Simulation used a force mesh of 2560^3 cells.

Once the force has been calculated and interpolated to each particle position, the particles are updated using a leapfrog integration scheme. The size of each timestep was limited to constrain:

$$\Delta t \leq \sqrt{2\mu/|\mathbf{a}|}, \tag{2.2}$$

where again $\mathbf{a}$ is the particles acceleration and μ controls the integration accuracy. Using this criterion more than 11000 individual times steps where required to evolve the simulation to the current epoch.

To this dark matter framework prescriptions of semi-analytic galaxy formation have been applied to populate the density field with galaxies.

2.1.1 Properties of the MS Density Field

Often it will be useful for us to consider density fluctuations in cells of a given size and the distribution of cell densities. We measure this from the MS by binning each particle from the simulation into cells using the Cloud in cell algorithm discussed previously to assign each cell a density.

The distribution of cell densities for a number of scales can be seen in Figure 2.2 and follows the predicted lognormal distribution (Coles and Jones, 1991).

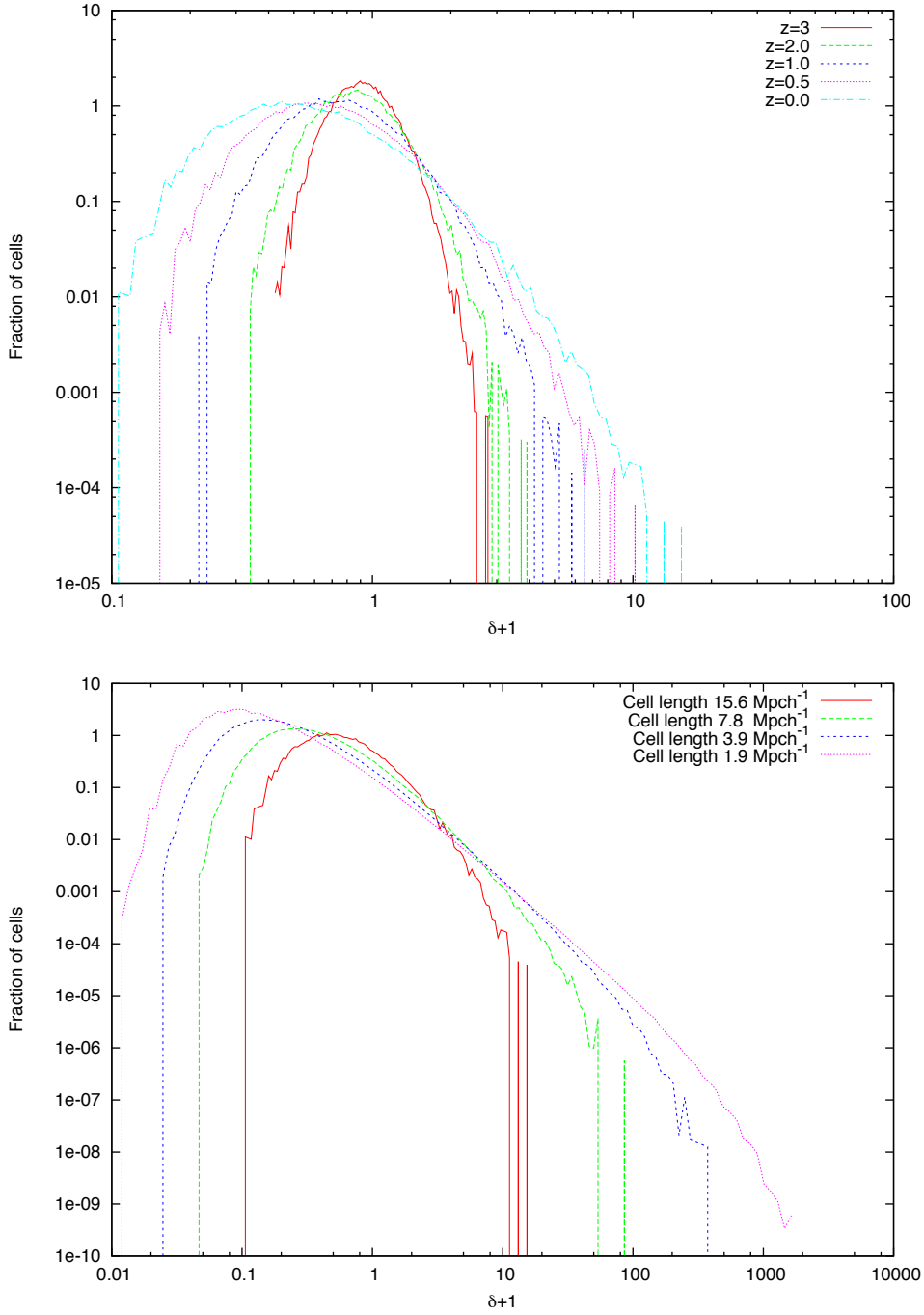

Figure 2.2: The distribution of cell densities calculated in cubic cells of size 15 h^{-1} Mpc and their evolution with time (top) and scale (bottom).

2.2 Haloes in the MS

2.2.1 Identifying haloes from the density field

All semi-analytic galaxy codes require two inputs to work with: 1) A halo merger tree and 2) physical properties of the haloes: mass, radial profile, angular momentum etc. The merger tree is a representation of the history of each halo at redshift $z = 0$, recording when progenitor haloe's merged. To accomplish it, haloes must first be identified from the dark matter field; this is a relatively intensive task and there is no definitive way to accomplish this, with many group finding algorithms in existence. Most authors (Kim and Park, 2006) attack the problem with Friends of Friends (FOF) halo finders which attempt to identify gravitationally bound structures by linking together chains of particles which are closer than a given distance from each other. The resulting structures will have a mean density which is determined by the linking length used in the algorithm. This is usually taken to be $b = 0.2\bar{n}^{-1/3}$ where n is the particle number density. This particular choice corresponds to a mean density of the order that we expect virialised groups to have. Practically this can be achieved by starting each particle in its own chain; if a candidate particle is found within the linking length, both chains are merged into one and the process continues.

Using a FOF algorithm it is possible, to identify dark matter haloes at each timestep, allowing a detailed picture of the merger history of haloes to be created. As a given halo can merge with another but in general will not split apart to form two separate new halos, the construction of a merger tree is the identification of the unique decedent of a given halo at each time step. This is accomplished by identifying the structures in which the particles which make up the current halo reside at a future time step. For each of these structures the number of identified particles is weighted by how gravitationally bound it is, and the structure with the highest value is labeled the descendent. Figure 2.3 shows an example of one such merger tree, showing clearly how larger haloes are hierarchically build up of smaller ones. Using these merger trees, semi-analytic codes are able to populate these haloes with galaxies

While on the whole successful, this approach can occasionally mistakenly identify structure. In some cases haloes passing close to each other can be combined

into one large halo when in fact the two structures are not dynamically bound. This has implications for both the measured properties of the haloes and can introduce phantom merger events into the tree.

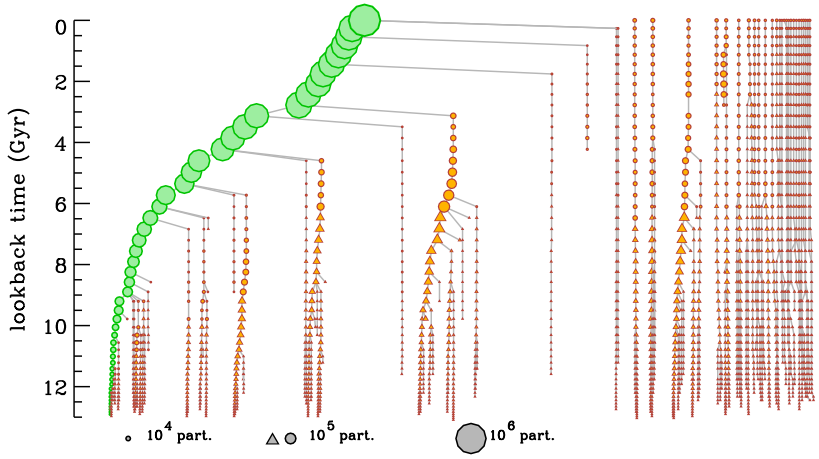

Figure 2.3: A merger tree from De Lucia and Blaizot (2007), showing the build up of a $10^{12}h^{-1}M_{\odot}$ halo at current redshift from a large number of smaller progenitors

The FOF algorithm is, however, not the only structure-finding algorithm employed by the MS. Substructure is also found using the SUBFIND algorithm (Springel et al., 2001). SUBFIND works in two stages: the identification of topologically distinct regions with a local overdensity and a subsequent gravitational unbinding processes. The first stage takes a catalogue of particles previously identified by a FOF algorithm and computes the local density at each particle. By kernel interpolation over particle neighbours, locally overdense regions are identified by defining an isodensity contour with a given threshold. As this threshold is lowered most regions of overdensity will simply grow in volume until they encounter another local region of overdensity and merge with it. At this point there exists a saddle point at the intersection. Each local overdense region bounded by an isodensity surface which traverses a saddle point is then considered a candidate for substructure. Each particle within a candidate region is then tested to determine if the particle is gravitationally bound to the region. This involves calculating the particles' total energy in physical coordinates. If this energy is negative then the particle is gravitationally bound, if positive it is removed from

the group. If a threshold number of particles survive this process they are identified as a sub-halo. At the end of this process we are left with particles within the FOF halo which belong to substructures and the remaining particles which are said to form the smooth background of the halo.

In this work, as we are concerned with halos themselves rather than their substructure, we will generally work with catalogues of halos determined from the FOF method.

2.2.2 Mass Functions

We use the MS halo catalogues produced by the Durham and MPA groups, which contain haloes from as small as $10^{10}h^{-1}M_{\odot}$ to $10^{16}h^{-1}M_{\odot}$. We measure the mass function of these haloes at a number of redshifts by counting the comoving number density of halos in a mass range $M < M_{halo} < M + dM$. Figure 2.4 shows the results in terms of the multiplicity function: the fractional mass contained in a unit range of $\ln M$. As is consistent with the hierarchical view of halo formation, as the redshift increases we see larger and larger structures being formed.

As we discussed in chapter 1 the mass function within regions where the mean density is not unity and will differ from the global mass function. While we have presented a number of theoretical models for this quantity none are guaranteed to give the correct answers and so we wish to compare with the distribution of haloes in the Millennium simulation. This poses a problem as the densities we measure from the Millennium simulation are non-linear while all previous expressions we presented for the conditional mass function are calculated from linear theory. We need to relate the non-linear density δ to the density that would have been obtained in a linear growth model. Mo and White (1996) and Scoccimarro and Sheth (2002) give a fitting formula for this relation to be :

$$\delta_0 = \left[1.68647 - \frac{1.35}{(1+\delta)^{2/3}} - \frac{1.12431}{(1+\delta)^{1/2}} + \frac{0.78785}{(1+\delta)^{0.58661}}\right]\frac{\delta_{sc}(z)}{1.68647}, \qquad (2.3)$$

where δ_{sc} is the collapse threshold obtained from the spherical collapse model and this final term was added by Scoccimarro and Sheth (2002) to account for the dependence of this quantity on cosmology. We check that this holds for the cosmology used in the MS by taking an early output of the simulation when

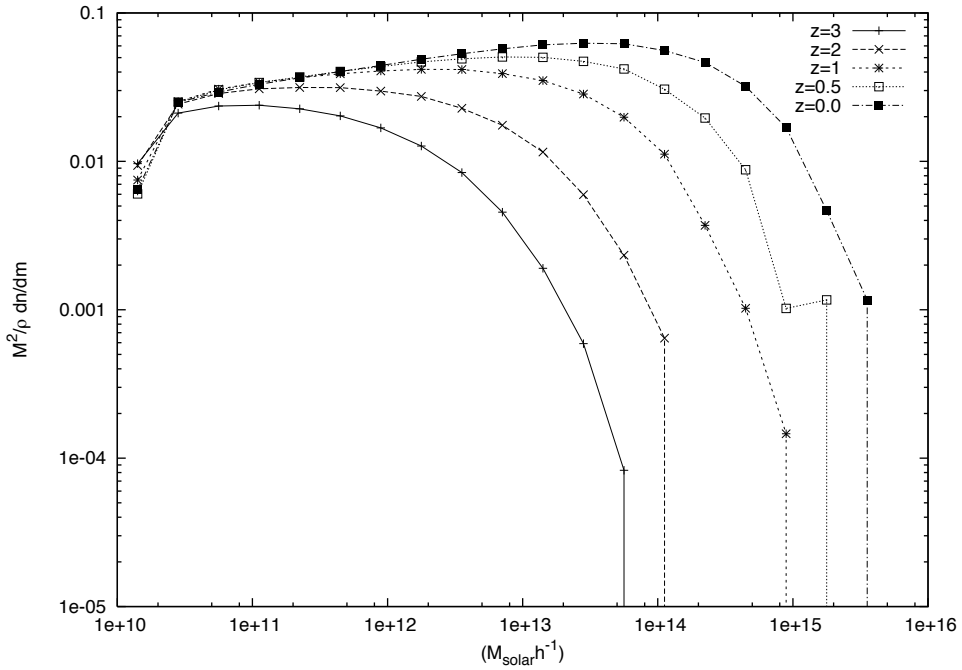

Figure 2.4: The evolution of the halo multiplicity function in the MS for a number of redshifts. The sharp fall-off at large halo mass at each redshift represents the largest objects to have formed by that epoch. As is consistent with the hierarchical model of halo formation larger halos are constructed form smaller ones at later epochs.

the dynamics are still linear and evolve the measured density to the present day using linear theory. The conditional mass function (CMF) is measured by binning haloes into cells of side $15\,h^{-1}$ Mpc and constructing the mass function in each cell before averaging the mass function over cells which have densities in the range δ to $\delta + d\delta$. At the same time we also construct the occupation number of each cell and the fraction of mass contained within haloes.

In PS theory the conditional mass function can be expressed in a universal form as:

$$\frac{m^2 N(M|\delta)}{\bar{\rho}} = \left| \frac{d \ln \nu_{12}}{d \ln m} \right| f(v_{12}). \tag{2.4}$$

What we measure from the simulation is $N(M|\delta)$, the average comoving number density of haloes in cells of a given size. However the functional form of the mass function is more easily seen if we express it in the universal form using:

$$\nu_{12} = \frac{\delta_{sc} - \delta_0}{\sqrt{\sigma_M^2 - \sigma_m^2}}, \tag{2.5}$$

as the variable of interest.

We calculate the conditional mass function of haloes by counting the number of haloes in a given mass range $M < M_{halo} < M + dm$ contained in a cell with overdensity in the range $\delta < \delta_{cell} < \delta_{cell} + d\delta$. Normalising by the frequency of such cells gives us a measure of the expected number of haloes in a given cell of a given mass. This quantity is displayed for cells of side $15\,h^{-1}$ Mpc for the $z = 0$ output of the MS in Figure 2.5 and the associated number of haloes contained in a cell of over-density δ and the scatter in this quantity is displayed in Figure 2.6.

As Figure 2.5 shows, the mass function split by cell density is not a straight forward scaling of the universal mass function. Less dense cells lack sufficient dark matter to produce the haloes with extreme masses, and so we see a sharp cut off of the mass function at some threshold value. As an example, the lowest density cell in the sample has overdensity $\delta + 1 = 0.13$. This corresponds to a mass within that cell of roughly $5 \times 10^{12} h^{-1} M_{\odot}$. As expected we find no halos greater than this mass contained within cells of this overdensity. This may at first seem trivial. However, as we are computing the conditional mass function by assigning a halo to a given cell by binning its centre of mass, there could arise

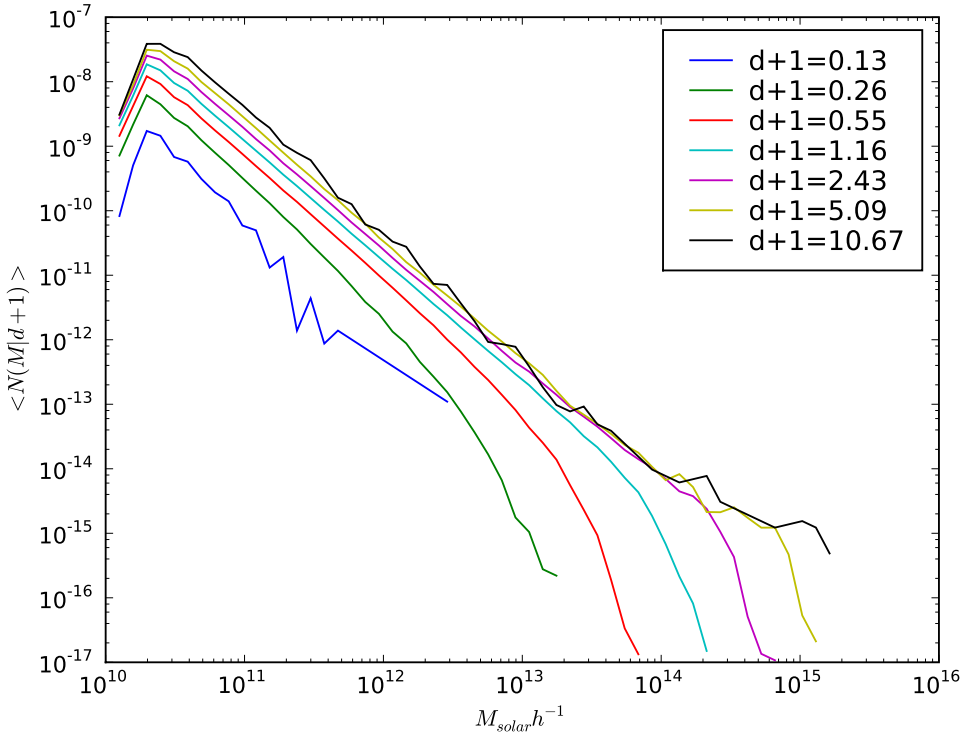

Figure 2.5: The conditional mass function as measured in the MS halo catalogue at z=0. The cells used to calculate the CMF are $15\,h^{-1}$ Mpc on each side. We see the distribution of halo masses within a cell is not a direct scaling and contains a cut off at a given mass. Lower density cells do not contain enough dark matter to produce larger mass halos.

a case where a halo of a given mass M_h could reside within a cell which contains a mass $M < M_h$. We expect this to become a problem when the size of our cells is comparable to the sizes of the largest haloes. If the cells used are too small they could essentially be entirely contained within a large halo. In this case the halo centre would be assigned to a conditional mass function bin with a misleading overdensity. We have found that cells of side $15\,h^{-1}$ Mpc are around the smallest that can be used to ensure that no cell contains a halo with more mass than the cell.

The dip at halo masses $M_h < 3 \times 10^{10} h^{-1} M_\odot$ illustrates the incompleteness in the halo catalogues used by the MS. In this work will typically err on the side of caution and take the confidence limit in terms of halo mass to be $10^{11} h^{-1} M_\odot$.

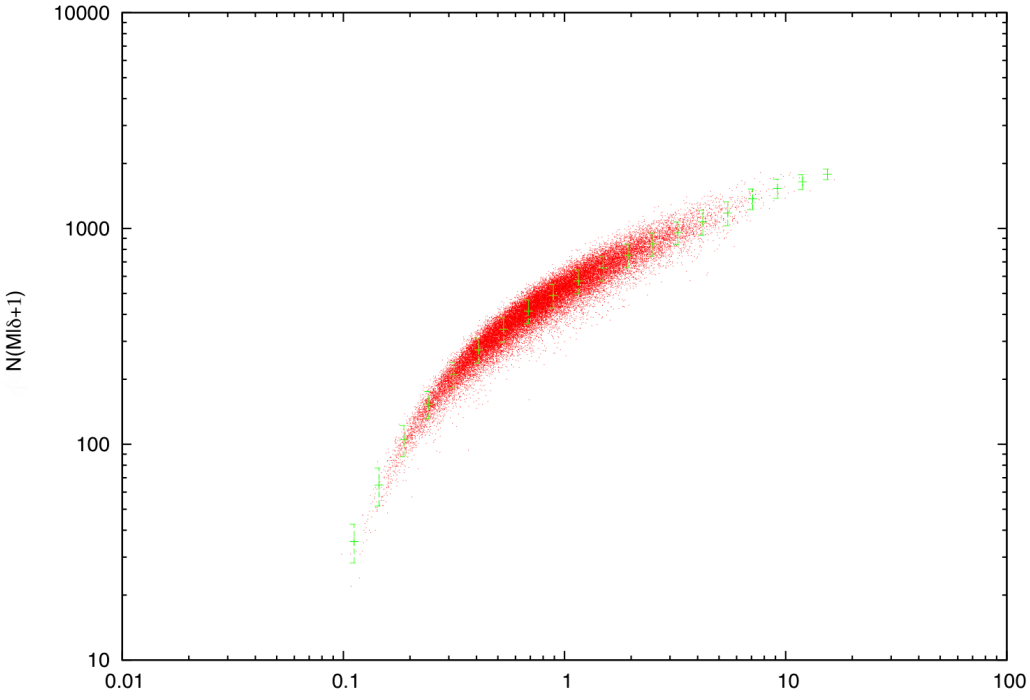

Figure 2.6: The scatter in the halo cell occupation number, the number of haloes contained within a cell of density δ. As expected the larger density cells contain many more halos than cells contained in voids. The scatter is well modelled by a Poisson distribution.

2.2.3 Velocity Statistics of haloes

From the continuity equation the peculiar velocity field (defined in Equation 1.23) in Fourier space is related to the density field by

$$\mathbf{v}_k = -iHf(\Omega)a\frac{\delta_k\hat{\mathbf{k}}}{|k|}. \tag{2.6}$$

The power of $1/|k|$ tells us that the velocity field is driven by large scale modes of the density field: the motion of matter in the Universe is dominated by bulk flows. This implies that on small local scales it does not in general matter if we consider an individual particle or the centre of mass of a collection of particles. We therefore expect the velocity distribution of dark matter haloes to be independent of their mass. Figure 2.7 shows the measured distribution of halo velocities in the Millennium simulation for a number of mass bins, as expected we see very little evidence of a dependence on halo mass.

2.3 Hierarchal clustering and Semi-analytic Models

In the hierarchal clustering model, most of the dynamic mass in the Universe is in the form of cold dark matter (CDM). One of the most important features of CDM is that it preserves small scale fluctuations in the primordial power spectrum, while if the Universe contained warm dark matter particles, much of the small scale structure is erased by free streaming in the early Universe. This means that in the CDM model the smaller fluctuations are the first to collapse and the formation of structure proceeds in a bottom up manner. The smaller fluctuations gradually over time merge to form larger and larger structures. This bottom up

While it is relatively easy to use dark matter simulations to follow the hierarchical merging of dark matter haloes, simulations which include baryons, to follow the merging of luminous material such as galaxies becomes much harder. Both in terms of the required computational resolution and having a good enough model with which to incorporate the varied gas physics (shocks, star formation etc).

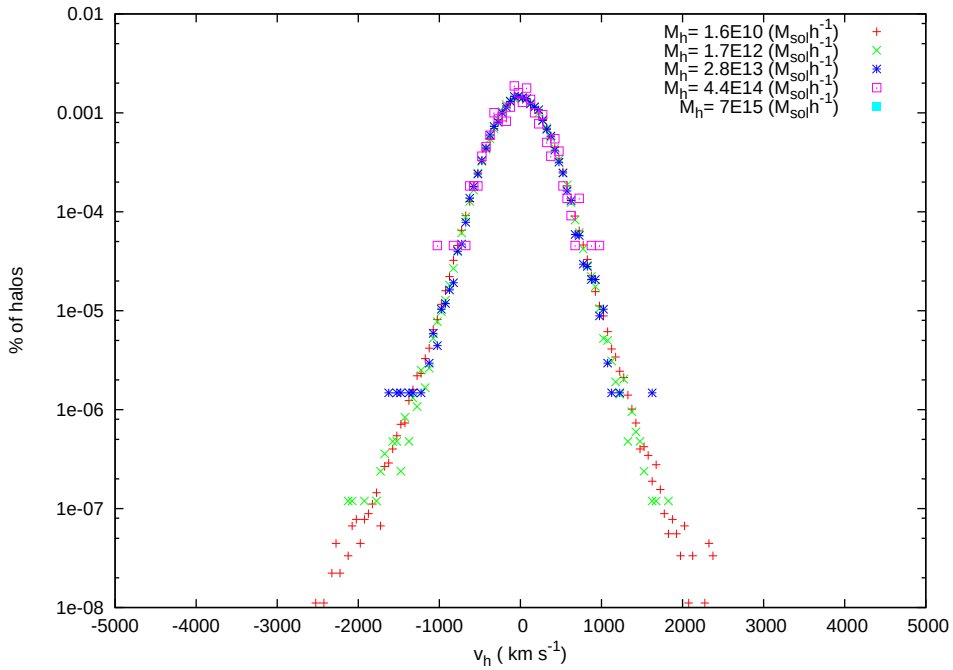

Figure 2.7: The distribution of centre of mass velocities of haloes of various masses found in the Millennium Simulation. The distribution is virtually identical for haloes of all masses.

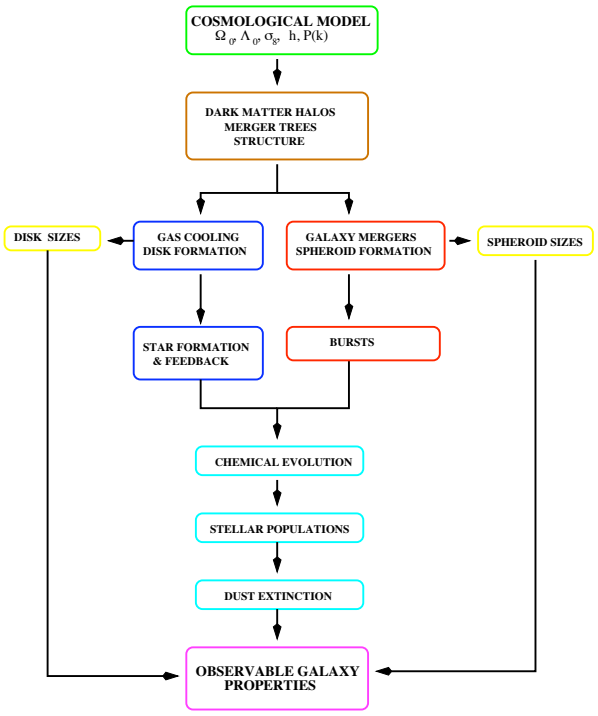

Figure 7. A schematic overview of the ingredients of a hierarchical galaxy formation model. Adapted from Cole et al. (2000).

Figure 2.8: A schematic of the semi-analytic taken from Cole et al. (2000). The role of merger activity in the formation of spherical galaxies compared with the passive evolution of disk galaxies is highlighted by the branching paths each takes.

Understanding the formation and evolution of galaxies from the primordial fluctuation of the CMB is perhaps one of the key goals of modern cosmology. In contrast to the clean simple physics of dark matter, understanding the process of galaxy formation is a complicated task. Not only do we need to contend with a fluid with pressure and viscosity, but we also have to take into account the fact that baryons can emit and absorb photons, chemically interact with the surrounding gases, ionize, form stars and undergo shocks from supernova winds. This makes full simulations of galaxy formation a hard task.

However, the hope has been that through analytic and semi-analytic treatments progress can be made. The key observation in both these approaches is that of White and Rees (1978) to note that baryons need to cool and condense to form stars and the only regions which have potential wells large enough to facilitate this process are dark matter haloes.

This ties the formation and evolution of the galaxy population to that of dark matter haloes. The approach taken by De Lucia and Blaizot (2007) , Bower et al. (2006) and others is to follow this process by following the collapse of dark matter haloes in high resolution simulations and then applying these semi-analytic algorithms.

In this section I will examine the main ingredients of various semi-analytic models and how they differ before moving on to compare the statistics of the resulting galaxy populations which will form the basis of the input parameters for my algorithm.

2.3.1 Galaxy Formation

Galaxy formation within the semi-analytic model takes place at the centre of dark matter haloes. The construction of stars requires the hot virialised gas in the dark matter halo to cool and condense onto the central galaxy. This is the main method by which new material becomes available to form stars. The gas can be thought of as having two states, the hot gas, which is shock-heated during mergers between haloes, to the virial temperature:

$$T_{\mathbf{vir}} = \frac{1}{2}\frac{\mu m_H}{k} V_H^2, \tag{2.7}$$

where $\mu = 1/1.71$ is the mean molecular mass of the gas, m_h is the mass of a hydrogen atom

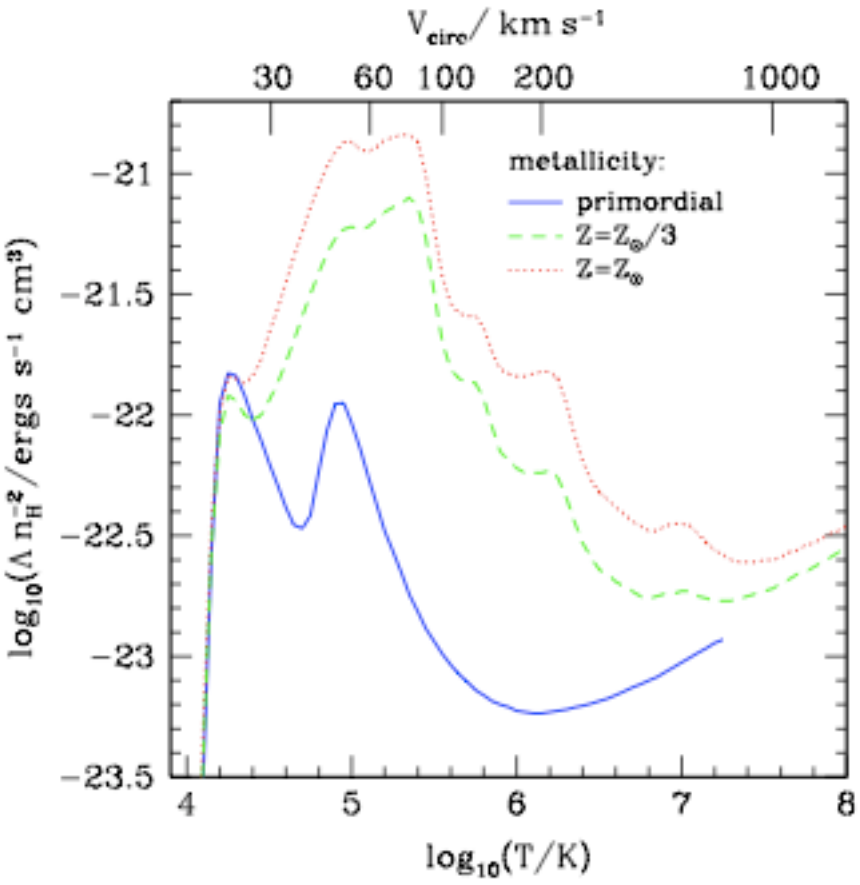

Figure 2.9: The cooling rate of hot halo gas as a function of metallically and temperature taken from Sutherland and Dopita (1993). As the gas temperature decreases the main cooling mechanism changes from thermal Bremsstrahlung radiation to the more complicated model of cooling through the recombination of metal ions.

Once there is a reservoir of cold gas available, the rate at which it can become stars needs to be estimated. In theory to calculate this rate would require a detailed understanding of the star formation process. A good approximation to the rate can however be obtained from a dimensional argument as:

$$\dot{M}_\star \propto \frac{M_{\rm cold}}{\tau}, \tag{2.8}$$

which balances the characteristic time-scale of star formation with the mass of cold gas available to the star formation process. The time-scale is taken to be a

typical dynamic time-scale of the galaxy $\tau_{\rm dyn} = r_{\rm gal}/v_{\rm gal}$ but may also include a dependence on the circular velocity of the galaxy.

2.3.2 Mergers

In the ΛCDM model, hierarchical mergers are the principal driving forces which control the growth of of galaxies. In addition major mergers (those in which the smaller of the two galaxies is greater than half the mass of the larger) are thought to be the cause of changes in morphology and colour of blue spiral galaxies into red elliptical galaxies. In semi-analytic algorithms the role of halo mergers is clear, but to apply these to galaxies requires extra physics. Galaxies are associated with the centres of bound structures and substructures. When these structures merge together the galaxy in the largest substructure becomes located at the centre of the new structure; this leaves any of the remaining galaxies to be dealt with. The remaining galaxies, both the central galaxy in the less massive halo and the satellite galaxies in both halos, become satellite galaxies within the new halo.

If there are substructures within the main halo, the satellite galaxies are associated with these sub-halos. These clumps of sub-structure are eroded over time by dynamical friction until they can no longer be identified. The galaxies themselves are thought to survive this process, as they are more compact than the dark matter structures they are embedded in. After a galaxy can no longer be associated with a given dark matter sub-structure, the assumption is that the galaxy decays in its orbit and will merge with the central galaxy in a dynamical-friction time scale.

2.3.3 Role of feedback

If the cooling of hot gas onto the central galaxy was the only mechanism of growth of galaxies then the galaxy luminosity function would have to be a scaled version of the halo mass function. In reality this is not the case, with the number density of both small and large galaxies being suppressed. To suppress the growth of these galaxies requires a mechanism by which the flow of gas to the central galaxy in a halo can be slowed. This is thought to come from feedback. There are a number of different sources of feedback which can affect the cooling of gas in small and

large haloes.

In small haloes the gas is heated and ionised by a background of UV photons generated by quasars and massive stars. These photons both heat the gas to $\approx 10^4 K$, increasing the pressure of baryons in halos with virial temperatures of less than this. The radiation also ionises the gas removing the efficient cooling by the excitation of atoms/ions in the gas, by collisions (Benson et al., 2002b,a).

In massive haloes the cooling rate is reduced by the injection of energy into the hot gas via two main mechanisms: supernovae winds and the accretion of material onto a central black hole. As the stars in a galaxy age, a number of them will eventually become supernovae, the largest creating high energy winds with the ability to eject and heat gas from the cold reservoir surrounding the galaxy (Larson, 1974; Dekel and Silk, 1986). Depending on the strength of these winds, the material may be ejected into the hot gas halo, to cool at a later time or it may be blown out of the halo all together.

2.4 Empirical Galaxy Properties in the MS

Having characterised how the dark matter haloes populate the density field, we now move on to examining the properties of the semi-analytic catalogues produced by applying the DeLucia and Bower models to the Millennium Simulation. In each case we will only consider galaxies with a rest frame magnitude brighter than $M_r < -18$ which is roughly the completeness limit within the MS. We split the galaxy properties into global: the colour distribution and over number density of galaxies of a given luminosity and local properties : the number density of galaxies of a given luminosity in haloes of a given mass and the spatial distribution of galaxies within haloes.

It is natural to split galaxies into two populations : star-forming and non star-forming galaxies. While we could do this using star-formation rates directly from the semi-analytic catalogues, most surveys we will be considering in this chapter do not have spectroscopy and so we cannot determine this quantity directly. Instead we take the two populations to be split by colour into red and blue galaxies using a colour cut of $M_b - M_v = 0.7$. While the precise level of this cut should depend on galaxy luminosity (Bell et al., 2004), the trend is gradual enough that

in this work we will take it to be static. To test if this is a sensible place at which to cut at, we calculate the colour distributions from the catalogues, which clearly show a bi-modal distribution. Figure 2.10 shows the colour-magnitude diagram and the distribution of galaxy colours obtained from the DeLucia model.

2.4.1 Luminosity Functions

The primary statistic of galaxies obtained from surveys is the luminosity function. The comoving number density of galaxies of a given magnitude is one of the most fundamental quantities for any model to compute. Semi-analytic models are usually tuned to reproduce the luminosity function at a specific epoch and for a specific class of galaxy. At redshift $z = 0$ both models agree well with each other for some bands but not others.

Most notably the De Lucia b band shows a slight excess at bright luminosities compared with Bower and a deficit at low luminosities. At higher redshift where the luminosity functions have not been tuned to give the right result the models predict differing numbers of galaxies in multiple bands. This can be up to a factor of 10 in density as demonstrated in Figures 2.11 and 2.12.

2.4.2 Conditional Luminosity Functions in the Millennium Simulation

As we have discussed previously the conditional luminosity function characterises the contents of haloes as a function of their mass. There is a clear distinction in both theory and the prescription of semi-analytics as to the role of central galaxies: those which reside at the bottom of the potential well of their parent halo, and satellite galaxies. In our analysis then we will consider these along with the colour of galaxies as four separate populations. We proceed by associating each galaxy with its parent halo and for each halo calculate its individual luminosity function for each of the four populations. We then proceed to average the individual luminosity functions over all haloes that reside in a log bin of width $\Delta \log M = 0.1h^{-1}M_{\odot}$. We also compute the integrated quantity of occupation numbers: the expected number of galaxies that reside in a halo of a given mass.

As can be seen in Figures 2.13 and 2.14 the total number of galaxies a halo

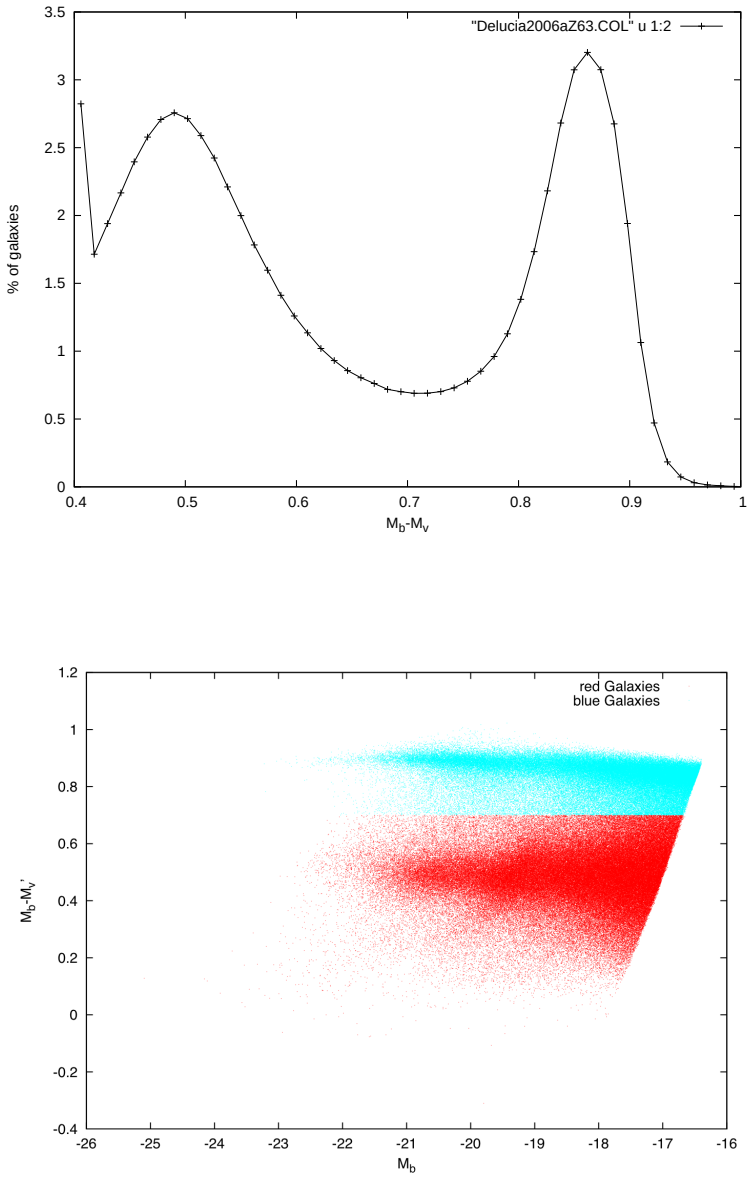

Figure 2.10: The distribution of galaxies with colour $M_b - M_v$ for the DeLucia semi-analytic mode (top). The red/blue galaxy modality can be clearly seen in the galaxy population. This split can also be clearly seen in the colour magnitude diagram of the same galaxy catalogue in the lower plot.

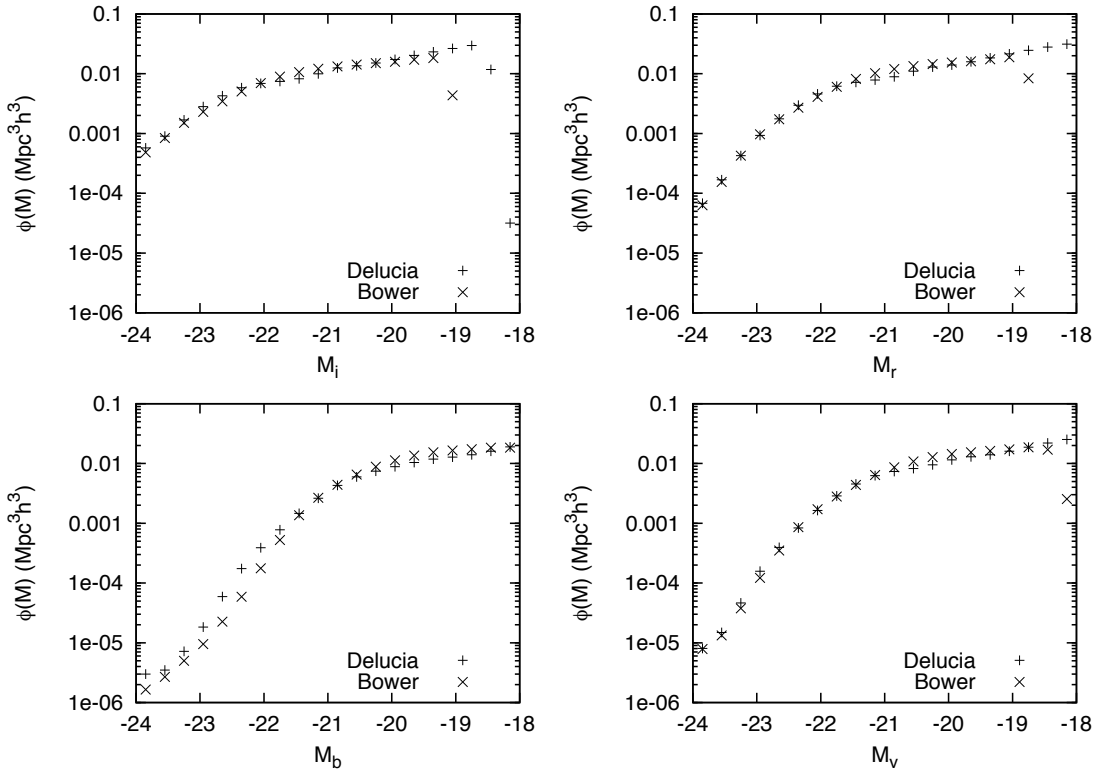

Figure 2.11: The z=0 luminosity function of the Bower and De Lucia models in the b,v,i and r bands.

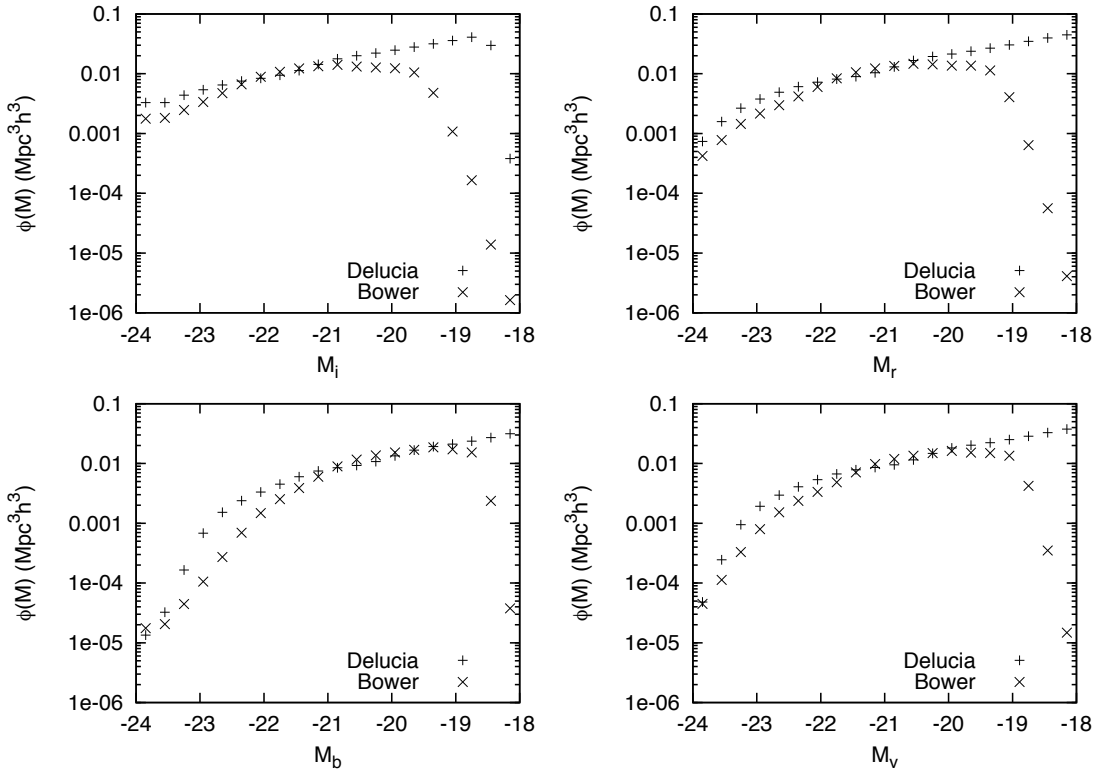

Figure 2.12: The $z = 1$ luminosity function of the Bower and De Lucia models in the b,v,i and r bands.

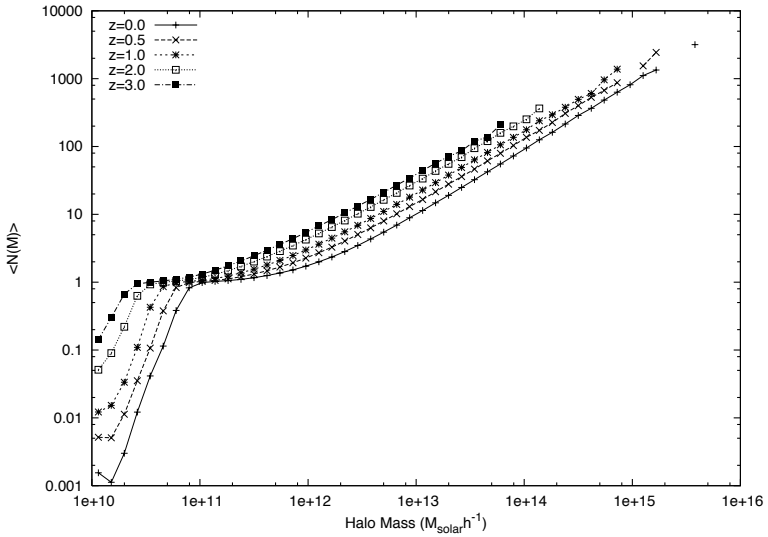

Figure 2.13: The evolution of the halo occupancy measured from the Millennium simulation and the DeLuica semi-analytic galaxy catalogues.

contains depends strongly on its mass and this dependence evolves with redshift. Below a mass of $M_h = 10^{11}h^{-1}M_\odot$ haloes are typically not large enough to contain enough baryons to produce a galaxy brighter than our magnitude limit, but due to the variation in baryon fraction from halo to halo, the drop off at small mass is not infinitely sharp. These galaxies are dominated as we would expect by blue central galaxies with an increasing number of satellites as the halo mass rises to that of the Milky Way. Around a halo mass of $M_h = 10^{12}h^{-1}M_\odot$ there is a change in the most likely colour of the central galaxy with the probability of its being blue gradually decreasing until a halo mass of $10^{13}h^{-1}M_\odot$ when virtually every central galaxy is red. Above $10^{13}h^{-1}M_\odot$ we begin to probe the region of galaxy clusters, which are dominated by red galaxies with a reasonably constant ratio of red to blue galaxies of 5.

At higher redshift the biggest change is the dominance of blue galaxies until much higher halo masses, where as at $z = 0$ the switch from blue to red central galaxies occurs at a higher mass $M_h = 3 \times 10^{12}$ while the ratio of red to blue satellite galaxies becomes almost equal.

Moving now to the luminosity functions of haloes we present in Figure 2.15

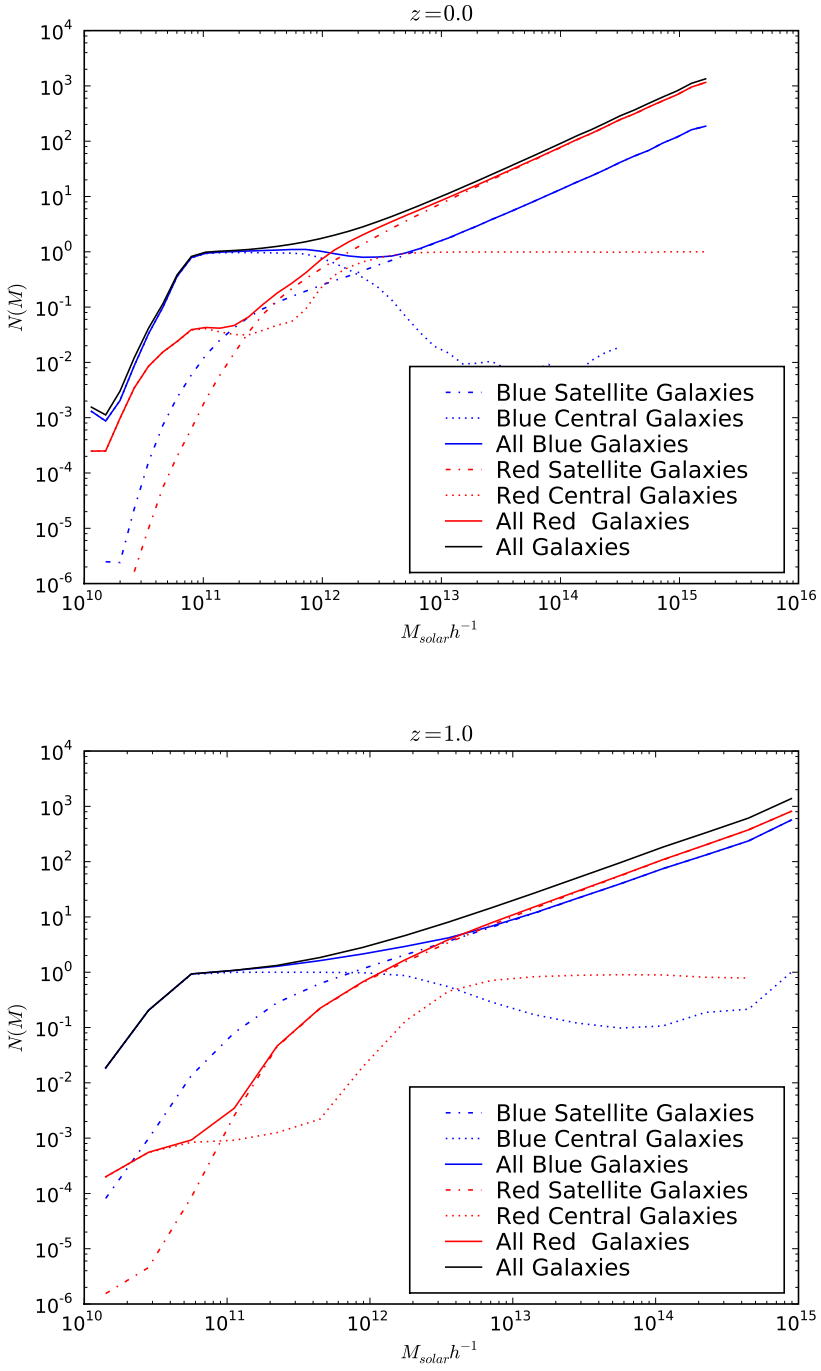

Figure 2.14: The halo occupation number as a function of halo mass split by type and colour as measured from the DeLucia semi-analytic model. Galaxies are split using the criteria $M_b - M_v < 0.7$ for red galaxies, $M_b - M_v > 0.7$ for blue galaxies. An overall magnitude cut of $M_r < -18$ is used.

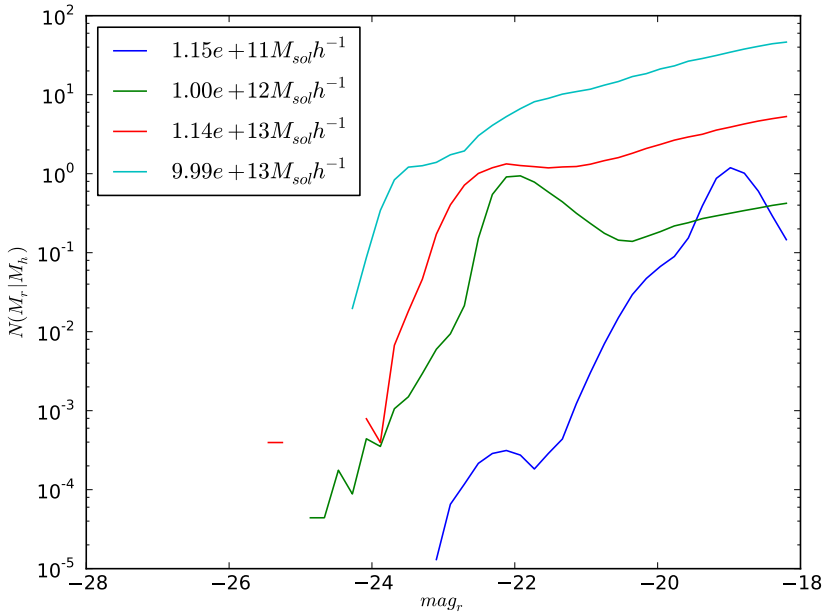

Figure 2.15: The conditional luminosity function for haloes of various masses

the breakdown of galaxy luminosities for haloes of $10^{11}h^{-1}M_{\odot}$, $10^{12}h^{-1}M_{\odot}$, $10^{13}h^{-1}M_{\odot}$ and $10^{14}h^{-1}M_{\odot}$. In practice we calculate the CLF in bins much finer than this but for clarity we will only present these cases as examples.

Figures 2.16 and 2.17 show the breakdown of halo contents as a function of galaxy magnitude, colour and type. The general shape of the CLF is a combination of a single peak around the mean central galaxy luminosity and a Schecter like function for the satellite galaxy. At around $M_h = 10^{12}h^{-1}M_{\odot}$ we see that this description breaks down. Below this mass haloes are dominated by blue galaxies and above it red galaxies dominate. At the transition mass the central galaxy distribution is a mix of red and blue galaxies and becomes a double peaked function.

To model the CLF then we use the sum of 4 separate populations

$$\begin{aligned} CLF(M_r, M_h) = CLF_{cr}(M_r, M_h) + CLF_{cb}(M_r, M_h) \\ + CLF_{sr}(M_r, M_h) + CLF_{sb}(M_r, M_h), \end{aligned} \tag{2.9}$$

where c or s denotes central or satellite respectively and r and b denote red and blue galaxies. As we have stated above we take the central galaxies to have a

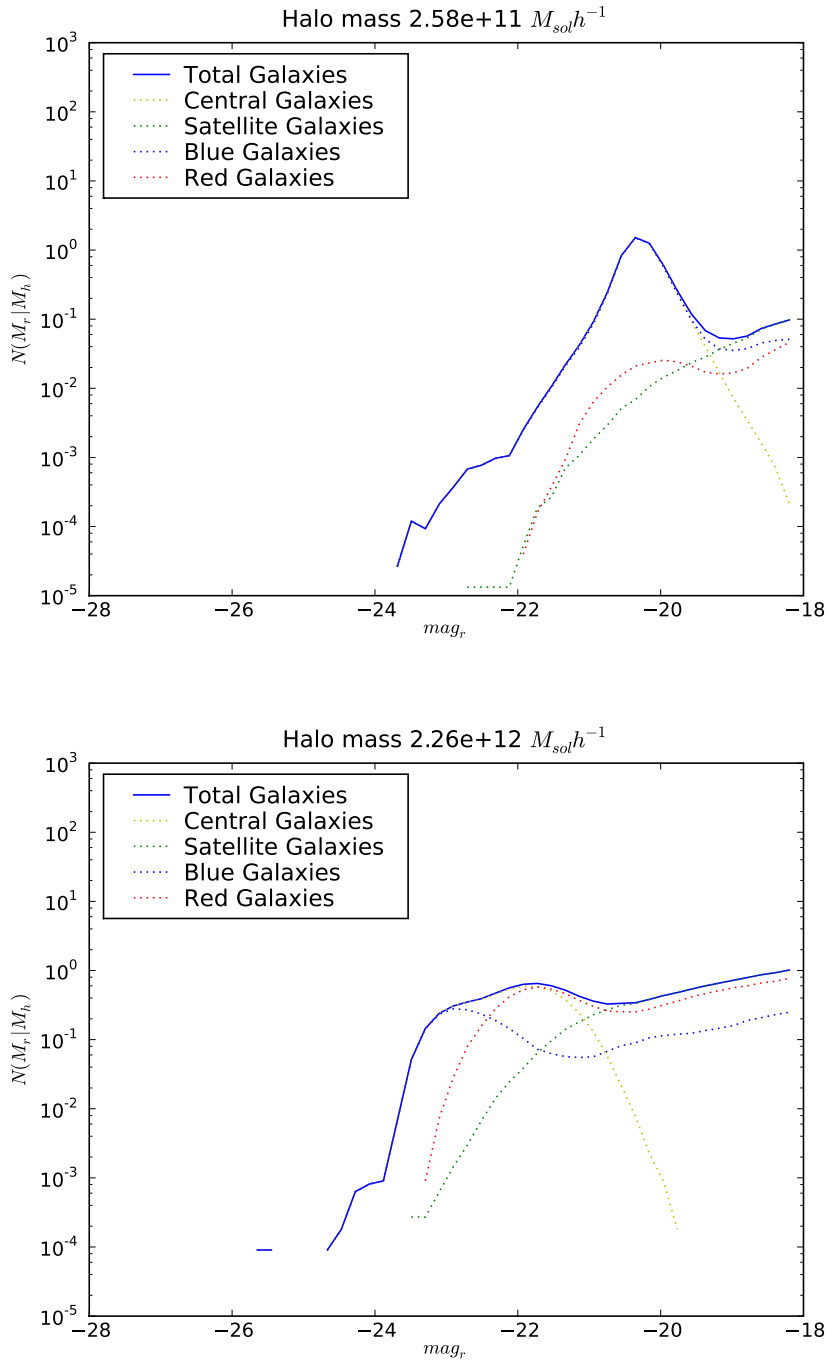

Figure 2.16: The r-band conditional luminosity function measured from the Delucia model. The contribution from central and satellite galaxies and both red and blue galaxies to the CLF are shown for haloes of various masses. The central galaxy contribution is roughly Gaussian for most masses while the satellite contribution is similar to a Schechter function. The distorted form of haloes around $10^{12}h^{-1}M_{\odot}$ is due to the contribution of both red and blue galaxies to the central galaxy.

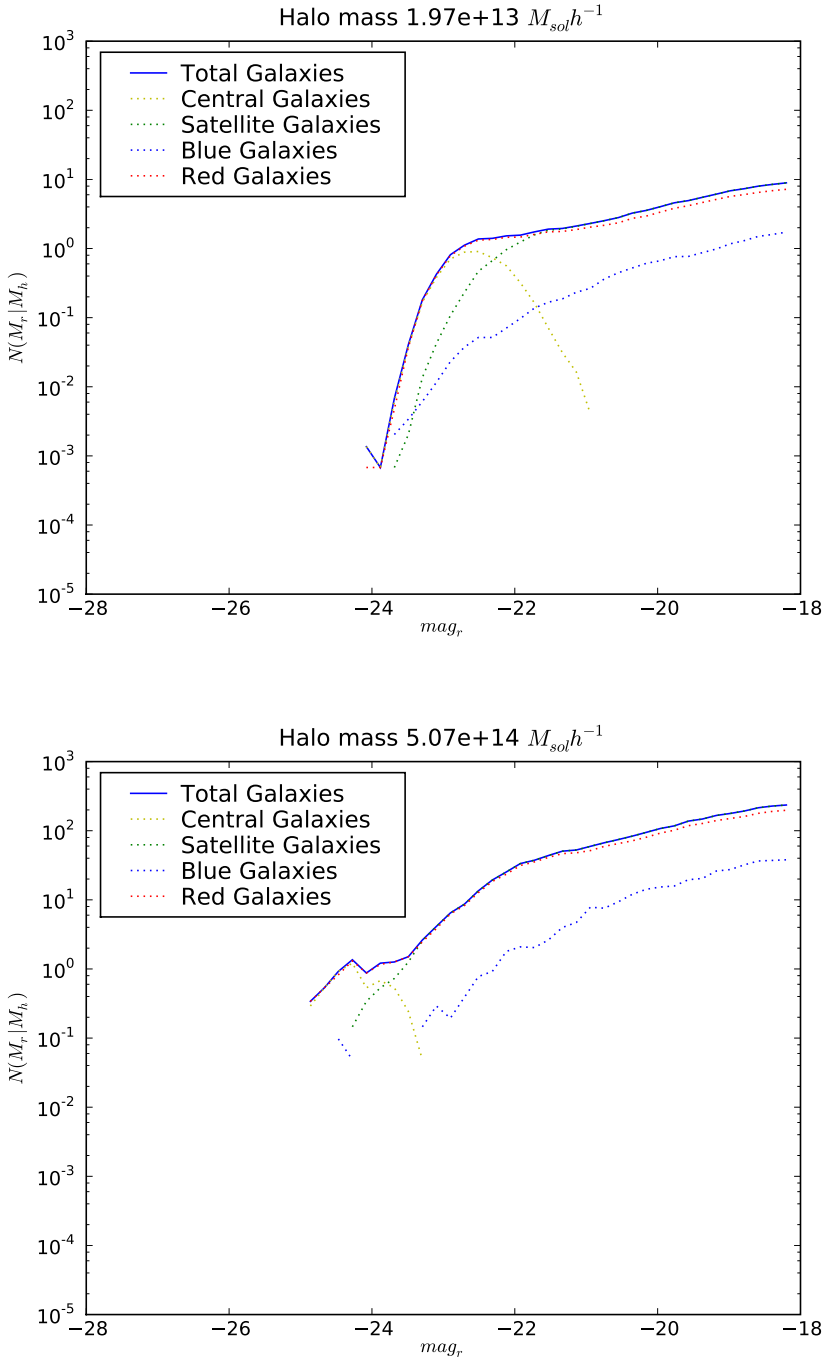

Figure 2.17: Continued from figure 2.16

Gaussian distribution, so we need to specify its mean and variance.

Following Cooray (2006) we take the form of the satellite galaxies CLF to be :

$$CLF_{sat}(M_r|M_h) = \frac{A_s}{2}10^{\gamma M_r}\left(1 - \mathrm{erf}\left(\frac{M^* - M_r}{\sigma_s}\right)\right), \tag{2.10}$$

and fit A_s , γ, M^* and σ_s to the values measured. Figures 2.20 and 2.21 show the fit we obtain to the satellite galaxy distribution.

2.4.3 Local light to mass ratios

The final quantity we measure from the galaxy population is the radial bias function we introduced in Eqn (3.8), which measures the departure of the radial distribution, of galaxies from an NFW profile. Here we want to measure how the galaxy population is spatially distributed relative to the dark matter field. Here we will make the assumption that each halo has an NFW profile correct for its mass with a concentration given by that in eqn (1.44); this will not always be strictly correct but as we are looking for an average trend over many haloes with similar masses it is a good approximation. We assume that all central galaxies sit exactly at the halo centre and do not include those galaxies in this analysis.

We first identify the halo in which each galaxy resides and bin them in shells of scaled radius $r^* = r_g/R_{200}$ from the halo centre. This gives us a number density profile of galaxies within that halo, $N_g(r^*)$, which we then divide by the halo's NFW profile to give a single halo bias function $g_h(r|M)$. We then take the average of this quantity over haloes $N_{haloes}(M)$ in the range M to $M + dM$ to obtain the expectation value of the radial bias:

$$\langle g(r^*, M)\rangle = \frac{\sum_{haloes} N_g(r^*)/\rho_{NFW}(r, M, c)}{N_{haloes}(M)} \tag{2.11}$$

We tend to find as in figure 2.22, that galaxies are less concentrated than the dark matter within the halo's NFW profile with a power law which evolves slowly with the mass of the halo.

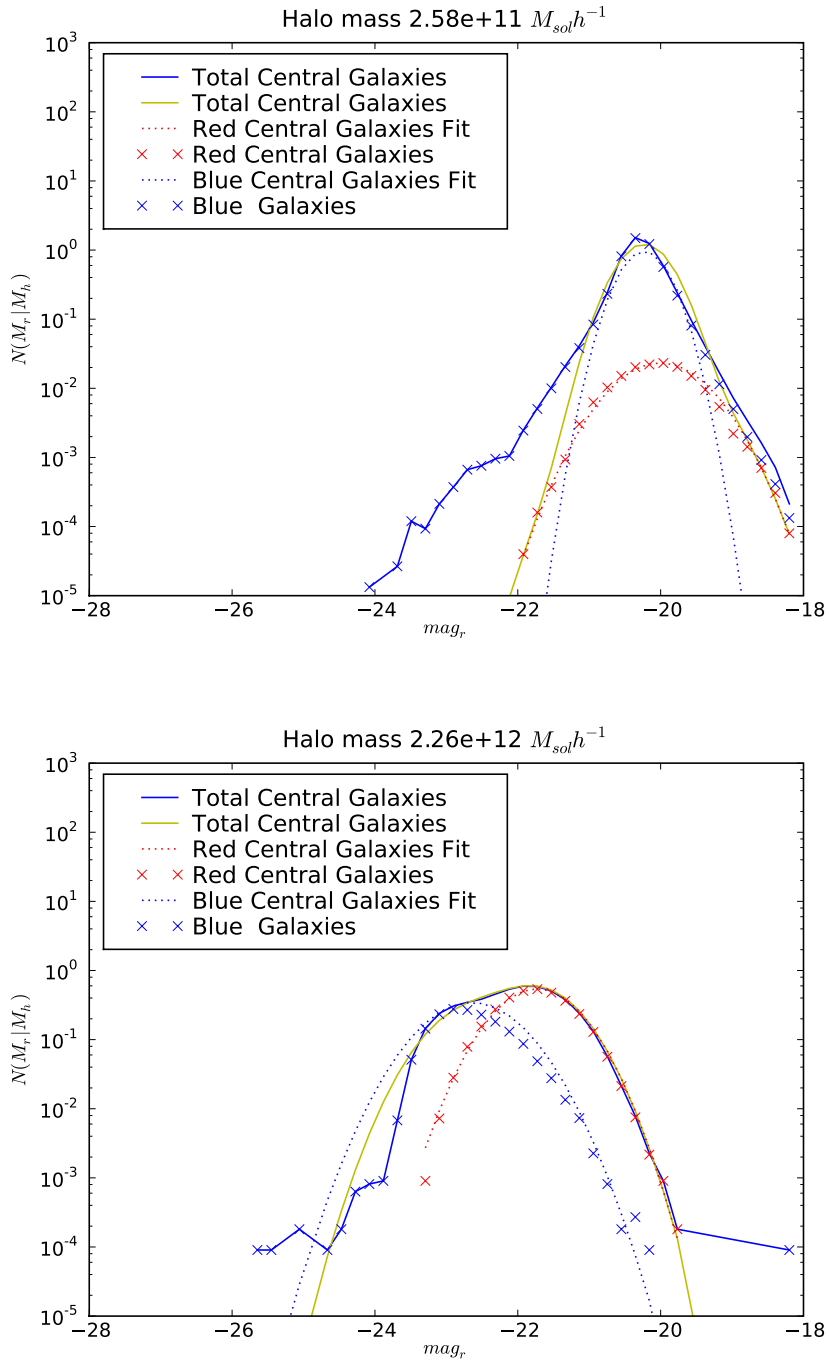

Figure 2.18: The r-band conditional luminosity function measured from the DeLucia model. The contribution from central and satellite galaxies and both red and blue galaxies to the CLF are shown for haloes of various masses. The central galaxy contribution is roughly Gaussian for most masses while the satellite contribution is similar to a Schechter function. The distorted form of haloes around $10^{12}h^{-1}M_{\odot}$ is due to the contribution of both red and blue galaxies to the central galaxy

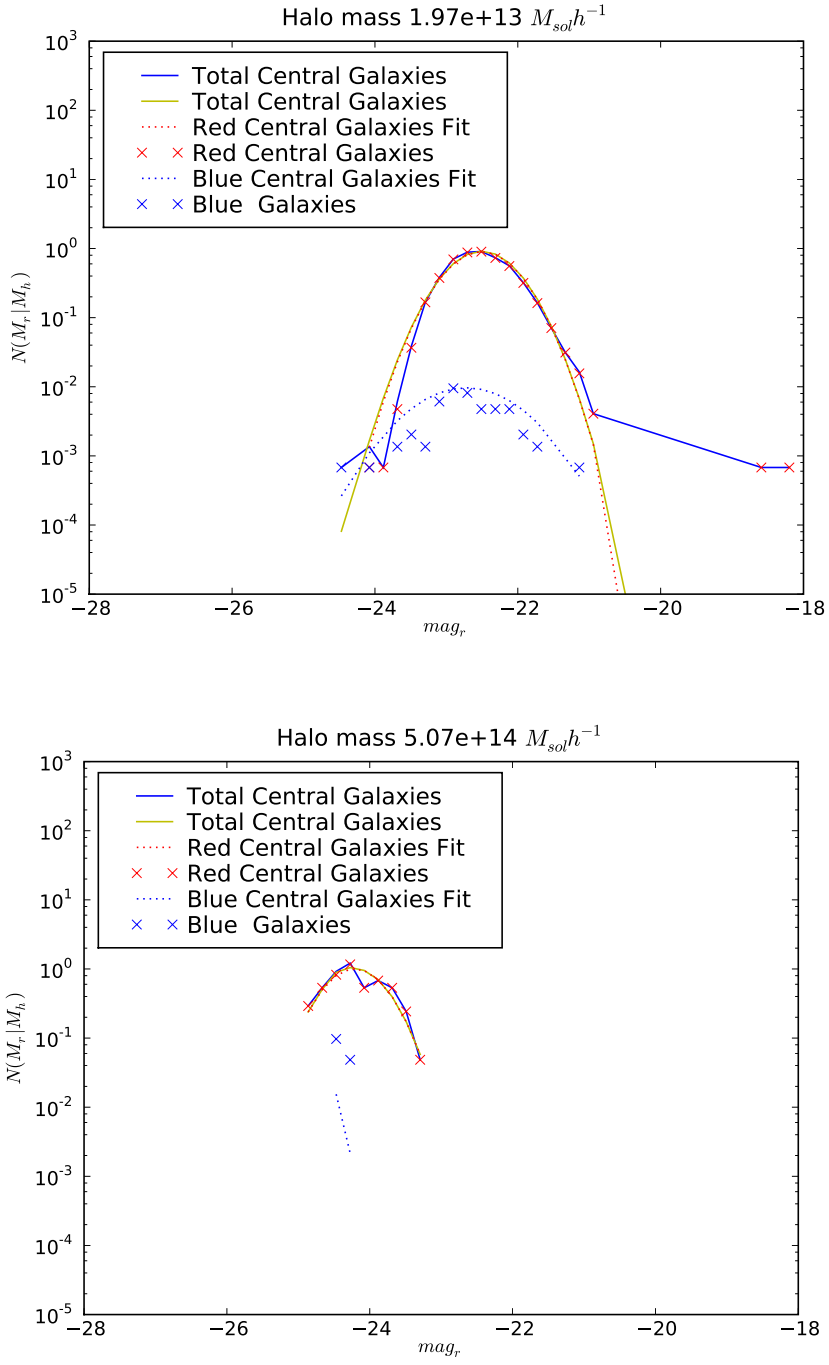

Figure 2.19: Continued from Figure 2.18

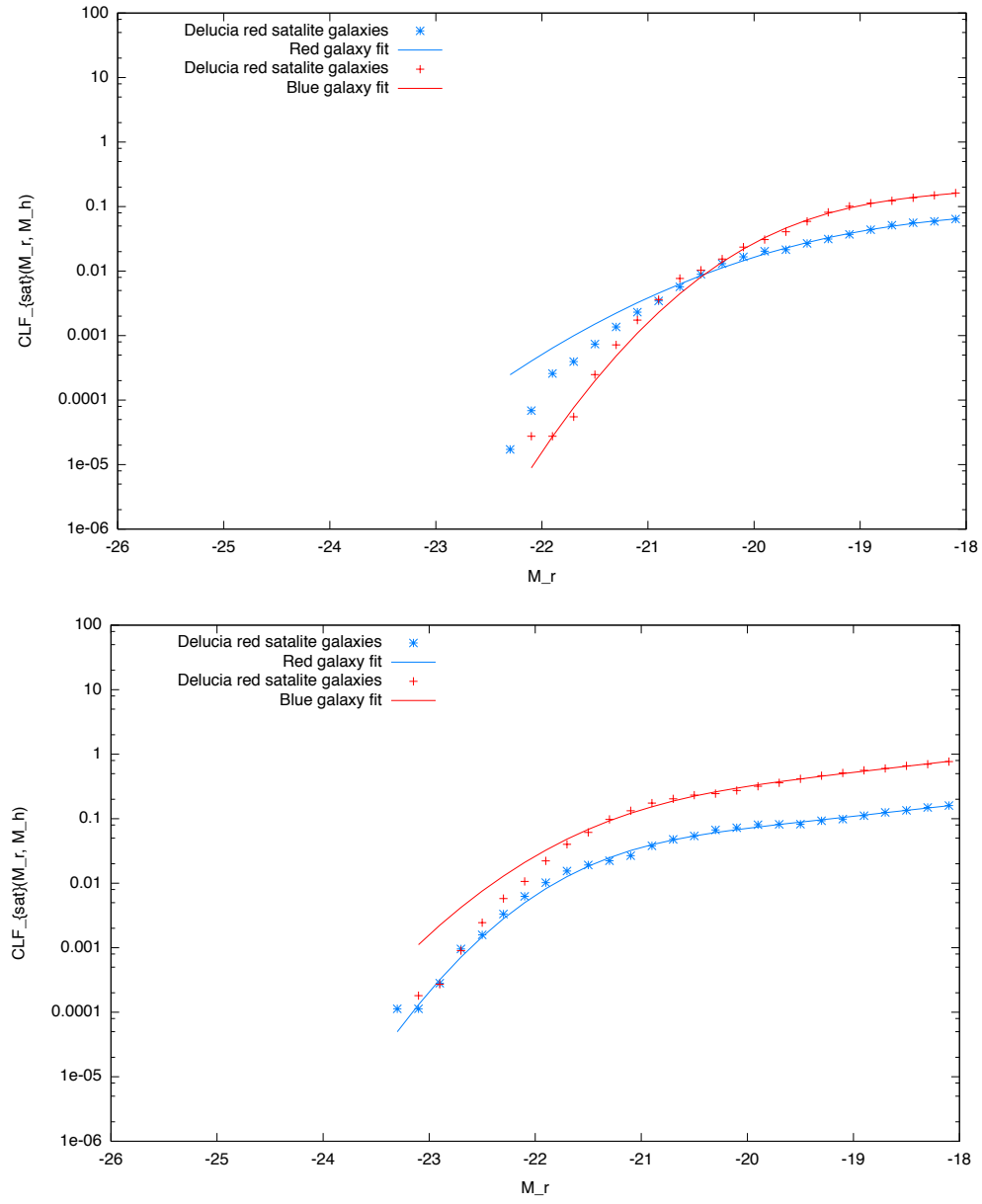

Figure 2.20: The r-band conditional luminosity function for satellite galaxies measured from the DeLucia model. The contribution from red and blue galaxies are shown with their corresponding best fits. The halo masses are $2.58 \times 10^{11} h^{-1} M_{\odot}$, $2.26 \times 10^{12} h^{-1} M_{\odot}$, $1.97 \times 10^{11} h^{-1} M_{\odot}$ and $5.07 \times 10^{14} h^{-1} M_{\odot}$

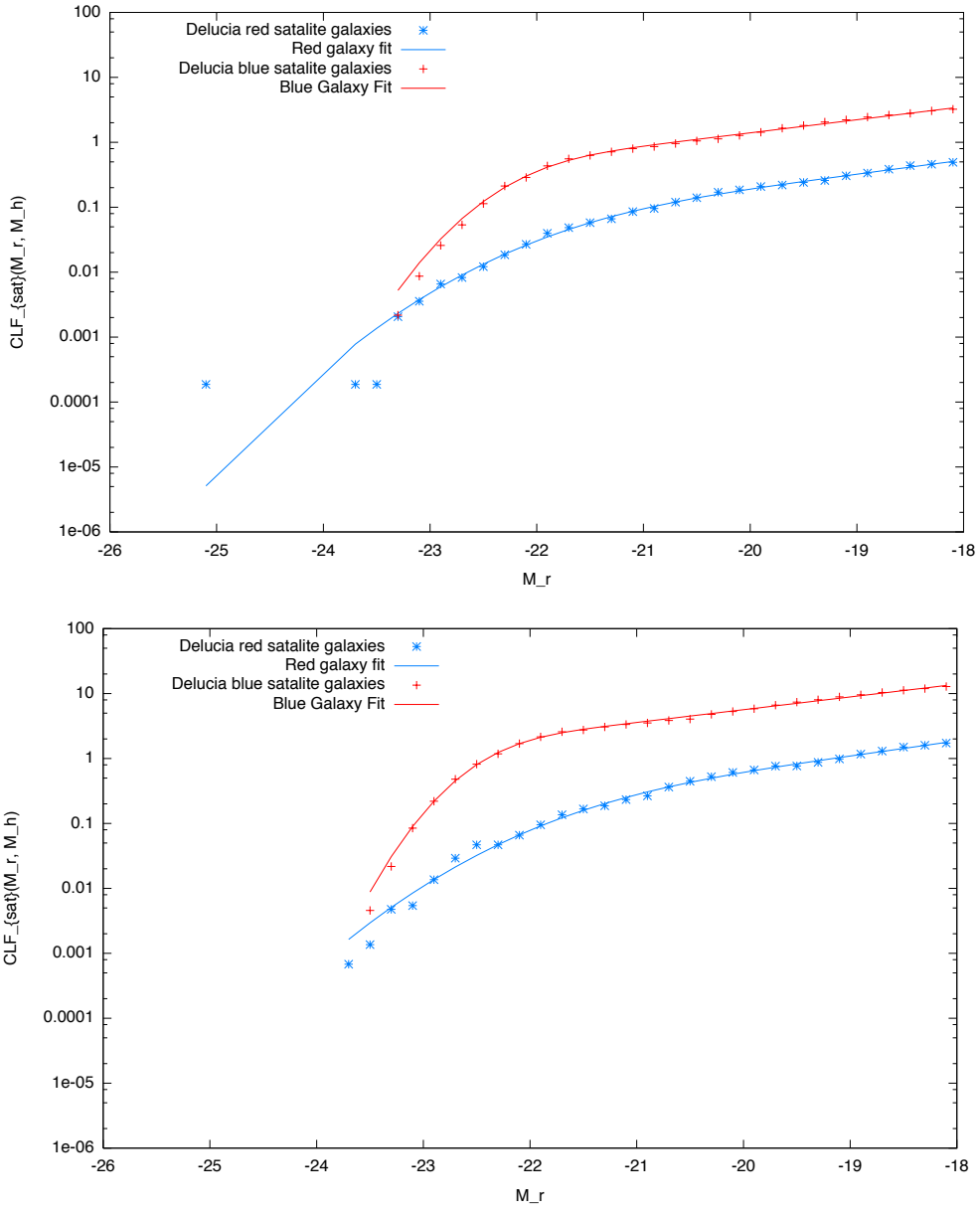

Figure 2.21: Continued from figure 2.20

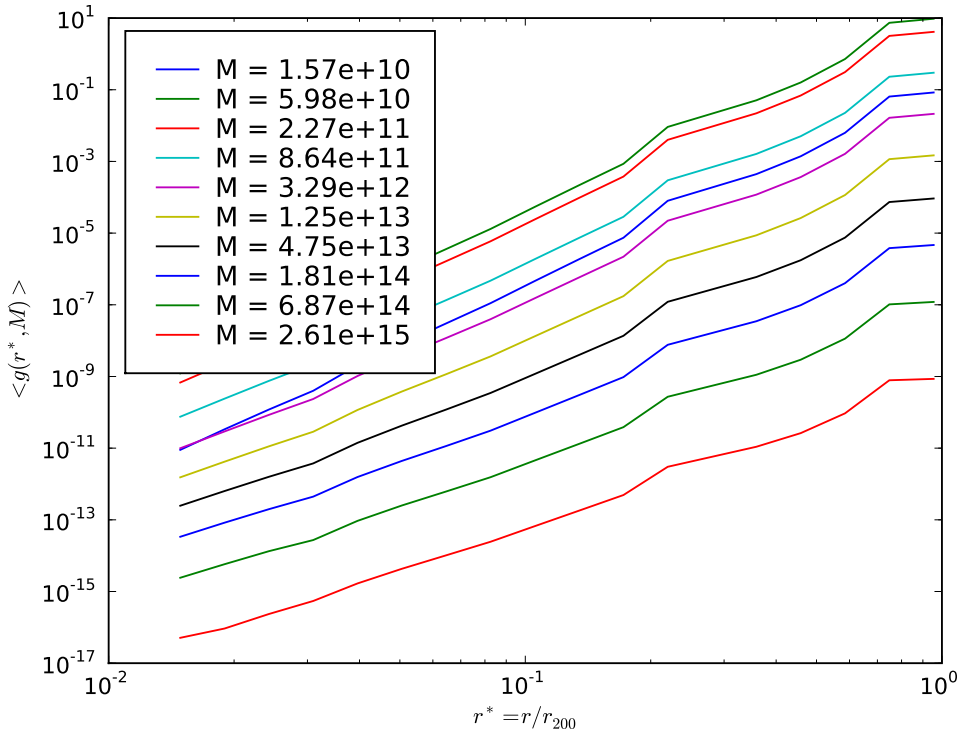

Figure 2.22: The radial bias of galaxies with respect to their haloes matter distribution, masses are in $h^{-1}M_{\odot}$.

2.4.4 Velocity Statistics of galaxies

In galaxy surveys the velocity of galaxies is a quantity almost as important as their positions. The velocity of galaxies contributes to the line of sight redshift, distorting the position of the galaxy which would have otherwise been inferred due to the Hubble flow. This distortion has serious implications for the clustering measured in surveys and must be modeled and accounted for (Kaiser, 1987). We can describe the velocity of galaxies in terms of two components: the velocity of their parent halo plus the velocity relative to the halos' centre of mass. A number of authors have claimed that the distribution of the three components of internal velocities can be described by three independent Gaussians. We want to check this assumption and find the exact mapping between parent halo mass and the variance of the Gaussian distribution. Each component of the velocity of a galaxy is subtracted from that of its parent haloes to give :

$$v_i = v_{g,i} - v_{h,i}, \tag{2.12}$$

which are binned up for each halo and averaged over haloes in $\log_{10} M_h$ bins. As with the radial bias profile we assume that central haloes share the same velocity as their parent halo and so will have the same velocity profile as the haloes discussed previously.

As expected the profiles are Gaussian with a width which is consistent with the relation found by (Bryan & Norman 1998; Sheth & Diaferio 2001) :

$$\sigma_{vir} = 476 f_{vir} [\Delta_{nl} E^2(z)]^{1/6} \left(\frac{m}{10^{15} h^{-1} M_\odot} \right)^{1/3} \text{kms}^{-1} \tag{2.13}$$

where $f_{vir} = 0.9$, $\Delta_{nl} = 18\pi^2 + 60x - 32x^2$ with $x = \Omega(z) - 1$ and $E^2 = \Omega_0(1+z)^3 + \Omega_R(1+z)^2 + \Omega_\Lambda$. This seems to break down at low halo mass where the velocity seems to become roughly constant.

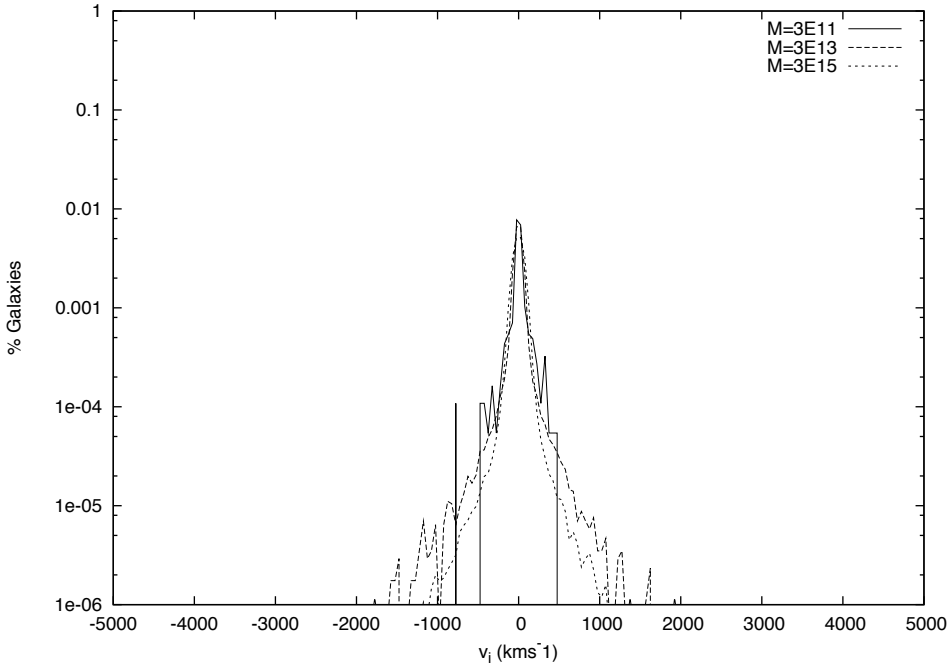

Figure 2.23: The average velocity distribution for galaxies within halos of mass $3 \times 10^{11} h^{-1} M_{\odot}$ (solid line), $3 \times 10^{13} h^{-1} M_{\odot}$ (long dashes) and $3 \times 10^{15} h^{-1} M_{\odot}$ (short dashes)

Chapter 3

Monte Carlo population of the dark matter field

There have been some successful attempts to apply the semi-analytic approach to volumes of $\sim 1\,h^{-1}\,\mathrm{Gpc}$, but the amount of computing power and time is prohibitive, placing limits on the number of realisations which can be produced. Instead we note that the largest haloes contained in the Millennium Simulation volume have masses of order $M_h = 1 \times 10^{16} h^{-1} M_\odot$. At the current epoch these represent the largest, rarest fluctuations of the dark matter density field, corresponding in galaxies to rich systems such as the Virgo Cluster. The sharp fall off in the number count of haloes more massive than this (Figure 2.4) implies that even in a much larger volume the probability of finding more massive objects is negligible. As we increase the size of our simulation volume, we are not probing new structures but rather increasing the number counts of structures already contained within the volume. Having seen and produced statistics which describe such structures in high resolution simulations, it is appealing to consider ways we might use this information to populate larger, lower resolution simulations.

Low resolution simulations are typically undesirable on two counts: the mass resolution of particles within the simulation is such that it is impossible to identify

small haloes using friends of friends methods; even attempting to populate the few large mass haloes using semi-analytic methods fails as we have not resolved the majority of their progenitors. Monte Carlo merger trees can be produced to simulate the merger history of haloes, but this really overcomplicates the situation given the lack of information from the simulation. If we are interested less in the detailed predictions that a semi-analytic prescription can provide (estimates of the galaxy bulge size, star formation rate, metallically, black hole mass, etc) and more interested in producing basic mock galaxy catalogues, then a much simpler and faster prescription can be used.

If we treat a low resolution simulation as a relatively smooth description of the dark matter density field, then we can add the lost resolution by hand by appealing to our knowledge of how haloes populate the background field. Statistically we know from the calibrated models described in the previous chapter that a region of space with mean overdensity δ will on average contain haloes with a distribution given by the conditional mass function $\mathrm{CMF}(M_h|\delta)$. Having measured δ from our simulation, we can then Monte Carlo sample the CMF to obtain a population of haloes for that region and place the haloes within that region as accurately as we can. In this way, we can increase the effective mass resolution of the simulation in regions of high density without incurring a large computational penalty.

Once a catalogue of haloes has been produced, we can in turn populate them with galaxies using the conditional luminosity function, radial bias and velocity statistics we obtained from the Millennium Simulation. One of the major advantages of this approach is that it allows the freedom to "tweak" the realisations produced in a relatively straightforward way and even incorporate halo occupation statistics obtained from surveys to produce mock catalogues which correspond well with what has actually been observed. The downsides are that we cannot follow the evolution, in time, of a halo: at each simulation output we are producing a new population of haloes distinct from the previous outputs. In this chapter we will detail the algorithm we have developed to populate the dark matter field with haloes and in turn the haloes with galaxies. We will then test this algorithm by recreating the Millennium Simulation volume from a low resolution re-simulation and comparing the clustering of haloes and galaxies.

3.1 Probability distributions

The method we detail below is heavily reliant on being able to select random numbers from a given distribution, so we shall summarise a few of the methods used to accomplish this. Generating truly random numbers on computers is virtually impossible due to their deterministic nature. Often we have to settle for pseudo random number generators: given an initial state, these will always produce a sequence of numbers which have the appearance of randomness and will eventually repeat. A good random number generator has a very long period and low correlation between consecutive numbers. We adopt the algorithm of L'Ecuyer with a Bays-Durham shuffle (Park and Miller, 1988) algorithm in this work.

While uniform distributions of numbers are easy to produce, we will also require the ability to generate numbers which conform to a given probability density function $P(x)$. A few select probability density distributions have simple efficient relations which let us take uniformly distributed random numbers and convert them to the desired distribution. One example which we will use often here are Gaussian random numbers which can be generated using the Box-Muller transformation. If x_1 and x_2 are uniformly distributed random numbers then the transform

$$y_1 = \sqrt{(-2\ln(x_1))}\cos(2\pi x_2) \tag{3.1}$$

$$y_2 = \sqrt{(-2\ln(x_1))}\sin(2\pi x_2) \tag{3.2}$$

gives y_1 and y_2 which are Gaussian random numbers. In general however there will not be an easy transform between uniform random numbers and the distribution we desire. Given a probability distribution function we wish to sample, we first construct the cumulative probability distribution:

$$y(x) = \int_0^x P(x^{'})dx^{'}. \tag{3.3}$$

The idea is to then generate a uniform random number R and invert the cumulative distribution:

$$x = y^{-1}(R). \tag{3.4}$$

In practice this is accomplished through the construction and sampling of lookup tables of the cumulative probability distribution. A uniform random number R

is generated and the lookup table is then traversed from lowest to highest value until the criterion $R > y_i$ is met, at which point the values in the table can be linearly interpolated to produce the desired result. For conditional probability distributions of two variables, R and z, a lookup can be performed by identifying the rows of the table which bracket the conditional value of interest ie $z_j < z < z_{j+1}$, performing the lookup operation described above for each row and then interpolating the result between the rows.

To reduce the computational time involved in traversing the lookup table, and to increase its accuracy, it is helpful to choose the variable we are trying to sample to have the highest probability at low values. This ensures that in the majority of cases only a few iterations are required and increases the accuracy of the results by avoiding the addition of many small numbers incurring rounding errors.

3.2 Populating the density field with haloes

We will assume that through some method we already have a low resolution realisation of the density field at a number of outputs; will detail below how this is done for both the MS volume, and larger volumes. The first step in the algorithm is to produce a smoothed density field in cells of a given volume. It is desirable that these cells are larger than the most massive haloes we expect within the volume so as not to split large haloes between a number of cells. However if the cells are too large then we risk smearing out the large scale density fluctuations. In this work the compromise is taken to be cells with a side of $\approx 15\,h^{-1}\,\mathrm{Mpc}$. Halos of a mass $10^{16}h^{-1}M_{\odot}$, which are larger than the limit of structures which have typically been formed in the Universe at the current epoch, have a virial radius of $r_{200} \sim 6\,h^{-1}\,\mathrm{Mpc}$. The choice of $15\,h^{-1}\,\mathrm{Mpc}$ cells then seems a sensible one.

The density within each cell is then calculated by particle binning using the Cloud-In-Cells algorithm discussed previously. This is then used to determine the expected number of haloes $N(M)$ that each cell should contain and the distribution of masses which occupy the cell. We need to be careful not to over-populate a cell with mass, and so as we select each successive halo mass we ensure that

the total mass in haloes for that cell does not exceed the mass contained within the cell.

We first obtain for each cell its expected number of haloes by interpolating the halo occupation function. However, as this is a measure of the expectation number, this neglects the variance within the halo population. Figure 3.1 shows the scatter in the $N(M)$ relationship. We find that this scatter is well approximated by a Poisson distribution when it is greater than 1.5 and a sub-Poisson distribution below this,

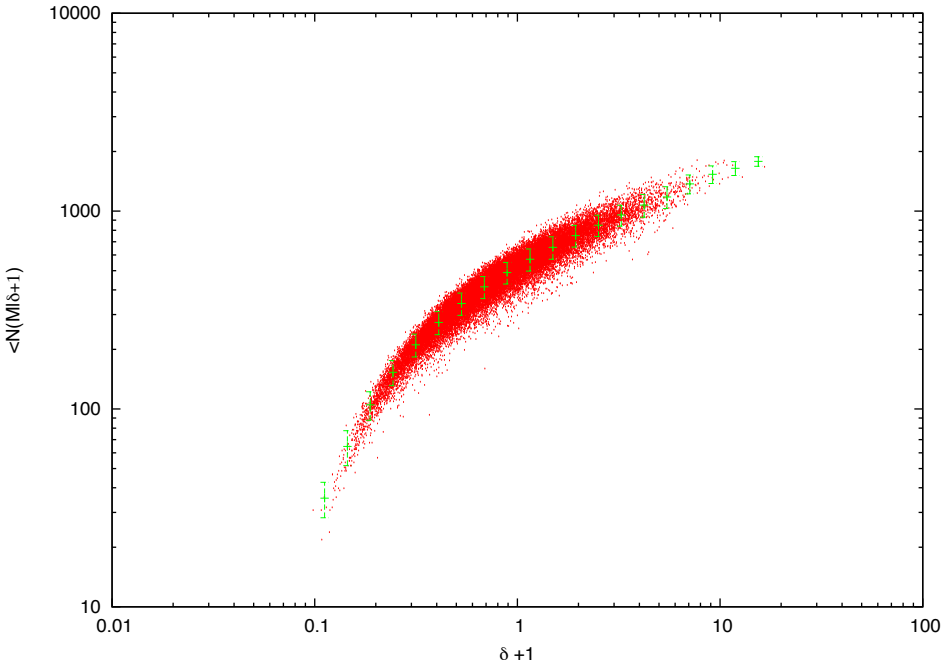

Figure 3.1: Shows the scatter in the number of halos that occupy a cell overdensity $d+1$

$$P(N|\langle N\rangle) = \left\{ \begin{array}{ll} \lfloor\langle N\rangle\rfloor + P(N) & N < 1.5 \\ \exp(-N^2/(2\pi(\langle N\rangle))) & N > 1.5 \end{array} \right\}, \tag{3.5}$$

where the function $P(N)$ is 1 with a probability equal to the decimal component of N: $N - \lfloor N\rfloor$ and zero otherwise.

Once the distribution of masses of haloes is known the next most crucial step is to place the haloes within the cell. Care has to be taken at this stage as there is a danger of producing inaccurate spatial statistics. On the largest scales the

clustering properties are sensitive to the contents of each of the simulations cells much more than the location of objects within that cell; similarly on small scales correlations are dominated by intra-halo contributions. An inaccurate placement of haloes within sub-cells is most likely to cause errors on scales comparable to the cell size.

The most basic approach would be to randomly place haloes within the cell, but this would obviously wipe out all structures below the cell resolution, replacing them with a Poissonian distribution which is undesirable. Scoccimarro and Sheth (2002) tuned the number of perturbation theory particles to the expected number of haloes and then placed each halo centre on a dark matter particle from the original simulation. This places the largest haloes in the regions of largest density within each cell and retains sub cell structure. While this works well with perturbation theory particles, it is impractical when using PM simulations. For a start, the number of haloes will far outnumber the number of particles within a simulation and there is an added complication: in a PM simulation, cells with a large over-density will contain a number of the dark matter particles that belong to the more massive resolved dark matter haloes. Association of a halo centre with a DM particle in this case could cause a number of haloes to artificially clump together as substructure of what should be a larger massive halo. We wish to identify and remove particles within these regions to ensure that the remaining particle distribution is from a smooth background.

An algorithm suitable for placing our haloes therefore needs to 1) use as much information about the subcell geometry as the resolution of the simulation allows while 2) taking care not to place multiple haloes in regions where particles are actually resolving high mass haloes and finally 3) take into account the abundance of haloes compared with DM particles. The general approach will be to first identify the locations of the largest dark matter haloes by associating them with the largest over-densities. We then remove all particles which are located within the virial radius of those halos. This should mostly leave particles which are associated with no massive haloes which can then be used to place the remaining smaller mass haloes. As there is a lack of particles compared with haloes we do this probabilistically, subdividing the cell into a number of subcells and treating the density within each as a probability density function. We associate as many

of the haloes as possible with particles until none are left and then place the haloes randomly within the subcells. The only other care we have to take is with collisions between dark matter haloes: each halo within the simulation should be a distinct object and their virial radii should not overlap. Removing particles within the virial radius in the first stage does not quite ensure this for all other haloes, so with each new halo added we explicitly check that no collisions occur. As long as the cell size is small this amounts to at most checking about 100 pairs of haloes per cell. While this slows down the algorithm it is not prohibitive.

The first step identifies haloes with masses above the expected resolution limit of the simulation. If for example the mass of each particle within the simulations is M_p then we select M_l to be $10M_p$. This typically gives a threshold of around $10^{13}h^{-1}M_\odot$ which corresponds to an R_{200} of around $0.5\,h^{-1}\,\mathrm{Mpc}$. If we use a sub cell resolution comparable to this then we would expect to identify and place them correctly on the resolved regions. Having done so, removing all particles within the R_{200} radius eliminates much of the unwanted small scale clustering.

For clarity each of the steps and logic of the algorithm are detailed below:

- **Stage 0**:
- Rank all haloes contained within the cell from highest to lowest mass.
- Bin all particles in a mesh which divides the current $15\,h^{-1}\,\mathrm{Mpc}$ cell in to a further n_s^3 subdivisions
- **Stage 1**:
- For each halo $> M_l$
- Identify the sub cell with the largest density
- Identify the haloes centre with a particle within this sub cell
- Remove all particles which are within R_{200} of the halo centre
- **Stage 2**:
- For each halo $< M_l$

- Randomly select a cell with probability $P_{cell} \propto n_{cell}$ the number of simulation particles it contains.
- If the cell contains particles which have not been used, associate the halo centre with a randomly selected particle
- If the cell contains no remaining particles, randomly place the particle within the cell
- Subtract the mass of the halo from the density of its current cell and re compute population probability

The other structural property of each halo we need to determine besides its mass is its concentration. Equation (1.44) relates the halo mass to the average concentration with a variance given by equation (1.45). We give each halo a concentration simply by sampling this probability. However we note that a number of authors (Navarro et al., 1997; Neto et al., 2007; Gao and White, 2007) have found correlations between halo concentration and environment for fixed halo masses. In future work we would like to extend the method here to include this dependence.

3.2.1 Limitations to cell and simulations size

Aside from the input statistics, the algorithm has a few adjustable values which need to be chosen. The two main inputs we have to consider are the size of the cells used to evaluate the conditional mass functions on and the sub-cell size used to place the dark matter haloes once they have been assigned a mass. There are a number of subtitles here which we feel it is useful to discuss in more detail.

Firstly the issue of the size of the box used for the evaluation of the conditional mass function. This would seem to be arbitrary but as we take the size of this box too small or too large we can run in to problems. As the box size gets larger and larger the mass functions of cells of a different δ tend towards the universal mass function. This reduces the accuracy of the halo mass selection as we are considering too general a region and also means its harder to determine the correct location for the dark matter halos in the density field. If we use a

cell of side of say $30\,h^{-1}$ Mpc to calculate the masses of the haloes and place them, then subsequently re-measured the conditional mass function using cells of $15\,h^{-1}$ Mpc, we would not recover the correct answer as we have over generalised by using the $30\,h^{-1}$ Mpc cells.

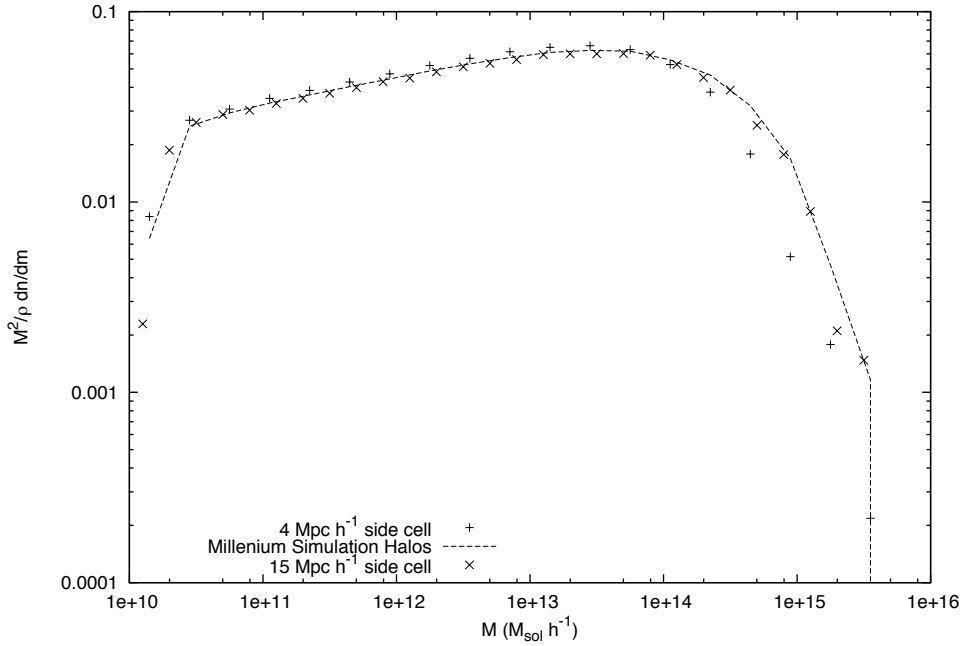

Figure 3.2: The halo mass function constructed using the method described in this chapter using cells of side $4\,h^{-1}$ Mpc and $15\,h^{-1}$ Mpc respectively. The mass function of the Millennium Simulation halos is shown for comparison. The splitting of high mass haloes in to many smaller haloes can clearly be seen for the smaller cell size

.

To avoid this loss of detail in the conditional mass function we may be tempted to make our cells as small as possible. However this too has problems when we consider the physical size of dark matter halos. The largest haloes' in the Universe today are of the order $1 \times 10^{16} h^{-1} M_{\odot}$ which gives them a r_{200} radius of approximately $6\,h^{-1}$ Mpc. If we use a cell size of $4\,h^{-1}$ Mpc then we can imagine a situation where a region of space which should contain a single $1 \times 10^{16} h^{-1} M_{\odot}$ halo gets covered by 3 or 4 cells. When we calculate the density within these cells, through the binning of simulation particles, we will tend to get densities which are lower than a larger cell which fully contained the mass. These cells will be

mistaken by the algorithm as lower density cells which will therefore be populated with smaller mass haloes instead of the single large mass halo. This effects the halo mass function as can be seen in Figure 3.2. A happy medium between these two extremes, is a cell size of around $15\,h^{-1}\,\mathrm{Mpc}$. Figure 3.6 shows this choice of cell size recovers the mass function accurately. As we will be working a lot with the Millennium Simulation which has a side of $500\,h^{-1}\,\mathrm{Mpc}$ we will actually use a cell size of $15.625\,h^{-1}\,\mathrm{Mpc}$ which tiles the simulation volume with 32^3 cells.

It is interesting to ask how large we can make a simulation volume and still use this algorithm to populate it. The algorithm we have described is essentially a repeated set of calculations which operate on each cell. One of the nice features of this algorithm is that the individual cells do not need to interact with each other. This means if we simply double the number of cells we are required to generate objects for, then our workload only increases by a factor of only 2. So in principle given a background density field the algorithm is boundless in the size of mock catalogue it can generate. The bottleneck in performance for the algorithm is therefore not the algorithm its-self but the N-Body code used to generate the underlying density field. We have found that the algorithm will run well for a simulation with mean particle density of ≈ 0.5 particles per $(\,h^{-1}\,\mathrm{Mpc})^3$. Below this there is little information on small scales that we can use to place the haloes and the algorithm fails. We also start to find regions in the voids where the cells we use for sampling contain only 1 particle. When this happens we are not sufficiently recovering the distribution of cell densities which can have adverse effects on the recovered halo mass function.

3.2.2 Halo centre of mass velocity

As we wish to eventually examine clustering in both real and redshift space, we also need to also assign each halo a centre of mass velocity. In the previous chapter we illustrated the fact that the velocity distribution of haloes is relatively insensitive to the halo's mass, so we will treat each halo as simply a point which samples the velocity field in which it is embedded. In PTHALOS the velocity of each halo is taken to be the velocity of the perturbation theory particle on which it is located. But in assigning a halo a centre of mass velocity using PM particles

we face a similar problem as with placement of the halo within the density field: that of particles belonging to resolved massive haloes. Each of these particles will have two components of velocity, $v = v_h + v_i$, where v_h is the centre of mass (COM) velocity of their parent halo and v_i is the particles' internal velocity within the halo. If unchecked this will mean that in general massive haloes receive an added dispersion of $\sigma_v(M_h)$, broadening the distribution of halo and galaxy velocities and ultimately leading to inaccurate redshift space distortion effects. An alternative approach would be to calculate an averaged velocity field within the cell of interest and then interpolate the velocity field in v_x, v_y and v_z to the location of a given halo. Computing the average velocity inside each cell is attractive as it helps solve the issue of resolved haloes, but it will not give the correct velocity dispersion in the cell. In addition to the average velocity in the cell we therefore also calculate the variance in each of the three Cartesian co-ordinates. As we know the masses of each halo within the cell, we also know the contribution of the internal velocity variance from each which can then be subtracted from the total variance of the cell:

$$V_{cell}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i}} \mathbf{v_i} \tag{3.6}$$

$$\sigma_{cell}(\mathbf{x}) = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i}} (\mathbf{v_i} - \mathbf{V_{cell}})^{\mathbf{2}} \tag{3.7}$$

As we know the distribution of masses of haloes within this cell, we can estimate the contribution from each halo to the average internal particle velocity dispersion. We can then subtract this from the variance in the particle velocity field to obtain the scatter of halo velocities. We interpolate all these fields to the position of the halo and assign the halo a velocity with a Gaussian variance about the mean for the cell.

3.2.3 From haloes to galaxies

At this stage we have a catalogue of halo positions, velocities, masses and concentrations, and we now turn our attention to the population of these haloes with galaxies. For each halo we treat separately central galaxies, which reside at

the bottom of the halo potential well, and satellite galaxies which orbit around the halo as independent populations. In turn we split both these populations by colour into red and blue subsets. For each dark matter halo we obtain the number of satellite galaxies by interpolating the occupation number $N(M)$ for each galaxy colour. The central galaxy is chosen to be red or blue by treating the central galaxy occupation number for each population as a probability. A cumulative probability function is then constructed from the conditional luminosity functions for each population and the magnitudes of each are obtained by sampling the probability distribution. The central galaxy in each halo is taken to be the brightest galaxy, of the selected colour, which is produced by sampling the probability function. This ensures that we do not end up with the case of a halo having satellite galaxies which are more massive/luminous than their central galaxy which is consistent with both observations, theoretical models and the semi-analytic prescriptions.

Galaxy location and velocity

Once we have a list of galaxy properties, we need to place them within the halo. The central galaxy is assumed to sit at the centre of mass of the halo: it is simply assigned the same position as the halo centre and also inherits the centre of ma ss velocity of its parent halo. Satellite galaxies are a little trickier. The simplest approach to take would be to assume that the galaxies sample the density run around the halo, in this case an NFW profile. In other words, $p_{gal}(r)dr \propto \rho_{\rm NFW}(r, M_h, c)r^2dr$: the probability of finding a galaxy between r and $r + dr$ from the halo centre is simply proportional to the fraction of the mass of the halo contained within that shell. If all galaxies formed out of the gas in their present day parent halo, this would be a sensible, approach as where there is more gas it is more likely a galaxy would form. However, in the hierarchical view galaxies form in isolated haloes which eventually merge to form their parent halo. In this view we would like to have a way to introduce a radial bias on the galaxies to encapsulate some of this complicated process and so modify the probability to:

$$p_{gal}(r)dr \propto \rho_{\rm NFW}(r, M_h, c)b(r, M)r^2dr, \tag{3.8}$$

where $b(r, M)$ is a radial bias function for a halo of mass M which we discussed in chapter 2. Having drawn a radius for a galaxy randomly from this distribution, we then assign its angular co-ordinates consistent with a spherically symmetric distribution using

$$\theta = 2\pi R \tag{3.9}$$

$$\phi = \cos^{-1}(2R - 1), \tag{3.10}$$

where R is a uniformly distributed random number.

As we saw in the previous chapter, the distribution of velocities within a halo is well approximated by three independent Gaussian each with a variance proportional to $M^{2/3}$.

$$v_{gal} = v_{halo} + v_{vir}. \tag{3.11}$$

A number of studies (Sheth and Diaferio, 2001; Bryan and Norman, 1998) have found the relative velocities of galaxies in haloes to be well approximated by an isothermal profile, with each of the Cartesian components of velocity independently drawn form a Gaussian whose rms depends on the mass of the halo as $\sigma^2_{vir} \propto M^{2/3}$. If this were accurate for all locations in the halo, then the equations of hydrostatic equilibrium would force all haloes to have isothermal sphere profiles. The assumption of an isotropic velocity profile fails at large radii within the halo, becoming more radial as material falls into the halo from the surrounding density field. For ease in this work we will ignore this complication, although one could imagine using a velocity profile dynamically consistent with an NFW profile.

In detail the $\sigma_{vir}(M)$ relation can be calibrated from semi-analytic models as is done in the next chapter, or can be obtained from theoretical considerations. From Sheth and Diaferio (2001) and Bryan and Norman (1998) we again have:

$$\sigma_{vir} = 476 f_{vir} [\triangle_{nl} E^2(z)]^{1/6} \left(\frac{m}{10^{15} h^{-1} M_\odot} \right)^{1/3} \mathrm{kms}^{-1} \tag{3.12}$$

where $f_{vir} = 0.9$, $E^2(z) = \Omega_0(1+z)^3 + \Omega_\Lambda$, $\triangle_{nl} = 18\pi^2 + 60x - 32x^2$ with $x = \Omega(z) - 1$ and $\Omega(z) = \Omega_0(1+z)^3/E^2(z)$.

Secondary galaxy properties

Beyond the conditional mass function and conditional luminosity functions, we can envisage adding more information into the galaxy and halo populations. A number of secondary properties such as galaxy metallicity, morphology and stellar mass could potentially be measured statistically from semi-analytic runs like the Millennium Simulation and used as conditional probabilities within the algorithm. Perhaps one of the most interesting, and the one we will concentrate on here is to add galaxy colour to the simulation. The most sophisticated approach would be to model the dependency of say the $b-v$ colour as a property of both halo mass and galaxy luminosity, producing a conditional probability function $P_{col}(b-v, mag, M_{halo})$. Using this function, each galaxy could be assigned a colour using a Monte Carlo sampling. Although this would be largely straightforward, here we will take a slightly simpler approach as a proof of concept and simply define a galaxy to be red or blue using a rest frame colour cut of $b-v=0.7$. Treating red and blue galaxies as separate independent populations, we then construct from the Millennium Simulation two conditional luminosity functions $\mathrm{CLF}_{red}(mag, M_{halo})$ and $\mathrm{CLF}_{blue}(mag, M_{halo})$, and their corresponding halo occupation numbers $N_{red}(M_{halo})$ and $N_{blue}(M_{halo})$. The only place the assumption of treating the two galaxy populations independently needs to be modified is when we consider central galaxies. As a central galaxy can only be red or blue the probabilities are mutually exclusive.

3.3 Repopulation of the Millennium Simulation

Armed with a fully determined set of statistics for the halo population and galaxy population, we want to ensure that the algorithm works and that any systematic effects are identified. We use the halo statistics measured in sections 2.2.2 and 2.2.3 and the functional fits obtained to the galaxy properties we measured from the DeLuicia model in sections 2.4 and 2.4.4 as the required inputs in our algorithm. We construct a new catalogue of galaxies and haloes within the Millennium Simulation volume using our algorithm. We will compare the mass function, luminosity function, colour and clustering of the haloes and galaxies produced against those found in the Millennium Simulation and its associated semi-analytics.

We will also require a density field as the basis for the algorithm. It would be possible to bin the original Millennium Simulation particles and calculate a density field to use as an input to the algorithm however in subsequent applications of this method we will not have access to such a high resolution simulation. Instead to keep this as fair a test as possible, we re-simulate the Millennium Simulation volume with a much reduced particle density.

In addition to a full reconstruction of volume we will also run two more tests where individual parts of the algorithm, the halo placement algorithm and the halo population algorithm, are compared with information from the original MS to illustrate the errors which arise from each process.

3.3.1 Dark Matter properties

As mentioned above, we do not want to rely on the high resolution density fields of the original simulation as all other realisations will have to utilise much coarser density fields. We could sparse sample the existing snapshot files from the MS to obtain the kinds of particle numbers we expect to use in our smaller simulations but this would not address the lower force resolution which is also used in the simulations. Instead we produce coarse versions of the volume by running new PM simulations with the same initial conditions but greatly reduced particle number. The original MS contained $N = 2160^3$ giving a spatial sampling, in the initial conditions, of the order of $0.1\,h^{-1}\,\mathrm{Mpc}$. As we have already explained, for this work we simply require the large scale fluctuations and so we re-simulate with a particle load of $N = 256^3$. The initial conditions for the re-simulation are obtained by binning the particles of an early ($z = 12.9$) output of the Millennium Simulation on a 256^3 mesh, obtaining three velocity component fields and a density field. The Zeldovich approximation discussed previously was then used to evolve the particles of the re-simulation to their starting points and the simulation was then run using the Millennium Simulation cosmological parameters to the present day.

Figure 3.3 shows the resulting density field of the re-simulation in comparison to the original Millennium Simulation. As expected the lower particle number has decreased the definition of small scale structures, but the large scale distribution

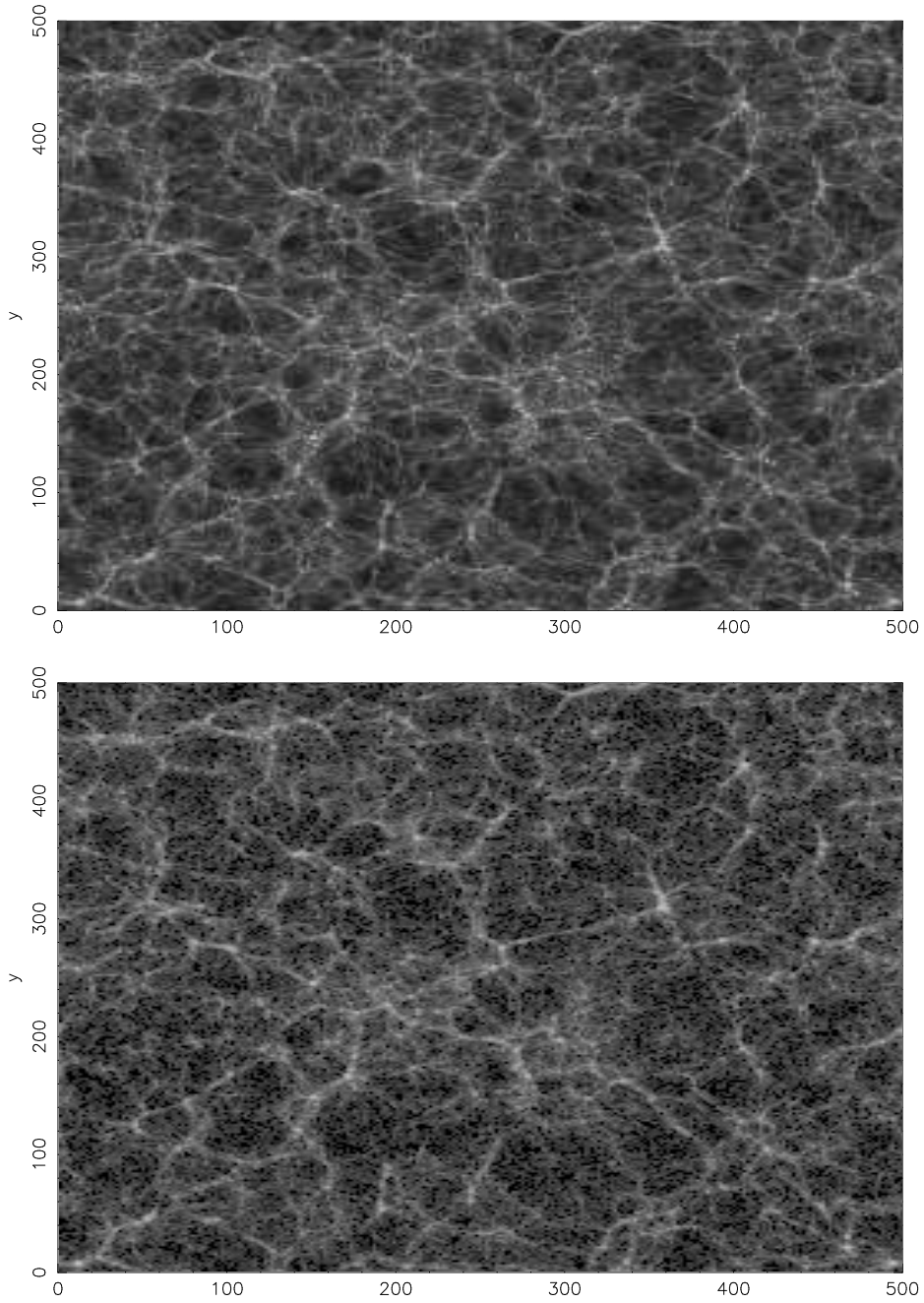

Figure 3.3: A 500 by 500 by 15.6 h^{-1} Mpc slice of the Millennium Simulation (top) and the 256^3 particle re-simulation (bottom). The large scale structure remains intact while small scale features are less defined. The grey scale is the logarithm of the density in the range $0.1 < \delta < 15$

.

of matter is largely intact. To determine exact effects of the decreased resolution we calculate the two point correlation function and compare with that of the Millenimum Simulation. As can be seen in figure 3.4, we recover the correct clustering on large scales but at small scales particles are much less clustered as expected. In terms of the halo model, this is because we still sample some of the contribution of larger haloes to the single halo term, but the lack of resolved low mass haloes significantly deteriorates correlations below a few $h^{-1}\,\mathrm{Mpc}$. This is however the ideal starting point for our algorithm: as we will see, by populating the density field with haloes and then reconstructing the density field from them, we can reconstruct the full MS clustering measurements.

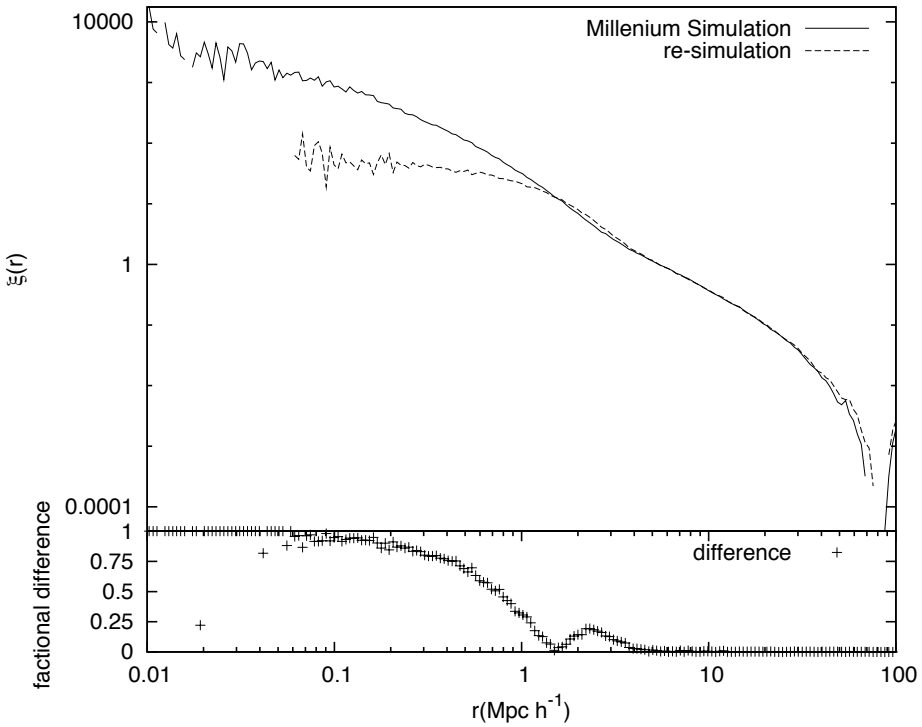

Figure 3.4: Comparison of the clustering of dark matter in the coarse density field and the original MS density field.

Algorithmically one of the main concerns we have is that the distribution of cell densities should be the same in both cases, as this is the quantity which when combined with the conditional mass function will give rise to the halo population. For cells of $15.6\,h^{-1}\,\mathrm{Mpc}$, which we can see from , Figure 3.5 that the agreement is extremely good.

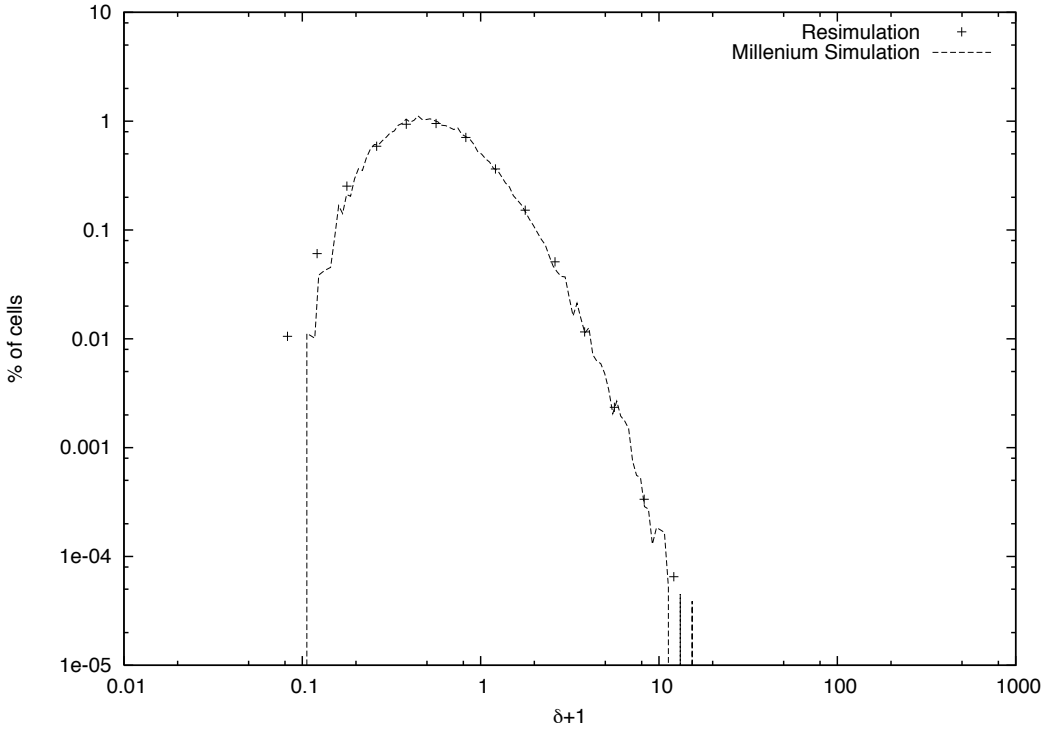

Figure 3.5: Comparison of the distribution of cell densities between the 256^3 particle re-simulation and the original Millennium simulation on a scale of $15.625\,h^{-1}\,\mathrm{Mpc}$

Confident that we have a density field which is accurate on large scales, we apply the algorithm to produce galaxy and halo populations in our volume.

3.3.2 Comparison of halo properties

As the positioning algorithm is an important part of the algorithm, affecting not only the clustering of haloes, but later galaxies on small scales we want to be sure it works well. As a test we take the MS halo catalogues, identify which $15\,h^{-1}\,\mathrm{Mpc}$ cell each halo inhabits and discard their positions. We then apply the positioning algorithm using a particle field from a low resolution re-simulation of MS (discussed in more detail below) to assess how well the algorithm copes. Figure 3.6 shows the correlation functions which result from repositioning all halos with a mass $M_h > 10^{11}h^{-1}M_\odot$: we see that on scales $r < 2\,h^{-1}\,\mathrm{Mpc}$ the repositioned haloes are overly clustered. The placement algorithm therefore tends to place haloes too close together within the subcells of the method described above. The upper two lines in the plot show the correlation function measured after each halo is individually dressed with an NFW profile by dividing the haloes mass into an appropriate number of particles with mass $M_p = 10^{11}h^{-1}M_\odot$. On small scales, as we would expect given that the correlation function here is dominated by the 1-halo term, the dressed correlation function gives a good match to what is observed. On large scales as well both correlation functions agree well with the minor deviations being seen at the transition from the 1 halo to 2 halo term. This extra clustering can be seen in Figure 3.6 .

These discrepancies may be worrying, but there are a number of simple modifications which could be applied to the halo placement algorithm to account for the extra small scale clumping. The simplest of these would simply be to manually increase the pair separation of haloes within a cell, compensating for the close pairs. However, as we have stated before, much of the utility of the type of algorithm we present here is in testing methods for statistical recovery of key quantities. It is therefore sufficient to produce a realisation which is broadly correct and whose underlying statistics we know well.

With an understanding of the limitations of the halo placing algorithm we move on to a full recreation of the halo catalogue with halos masses sampled from

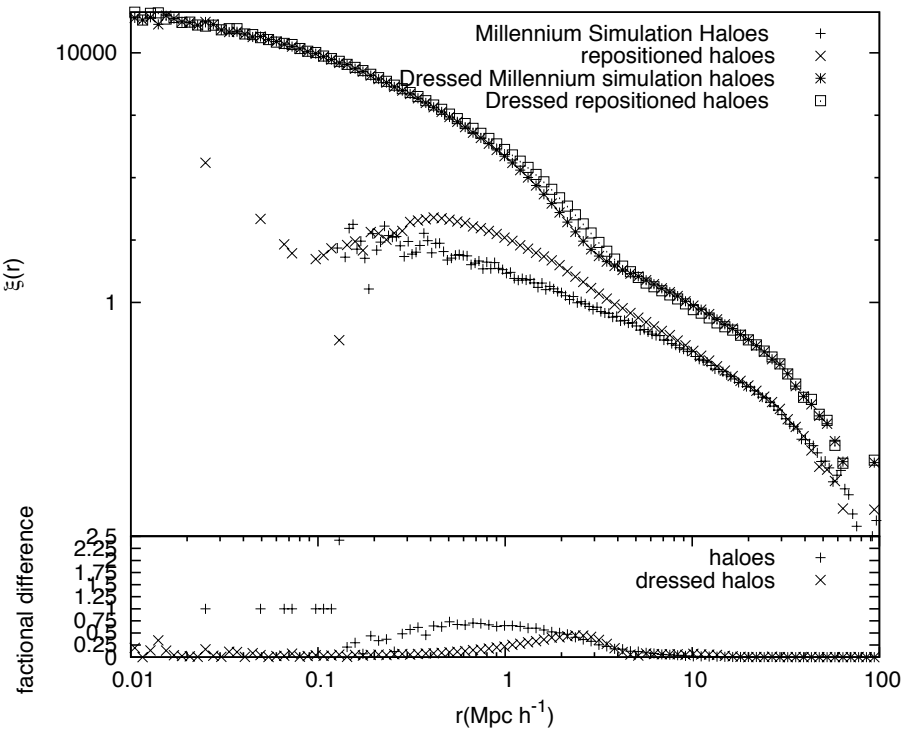

Figure 3.6: The clustering of repositioned haloes compared with the original haloes found in the MS and both populations dressed with particles in an NFW profile. There is a tendency for haloes to cluster more on small scales which has an effect on the intermediate scale correlation function. As expected however the small scale power, which has its main contribution from the 1 halo term, is well recovered as is the large scale correlation function.

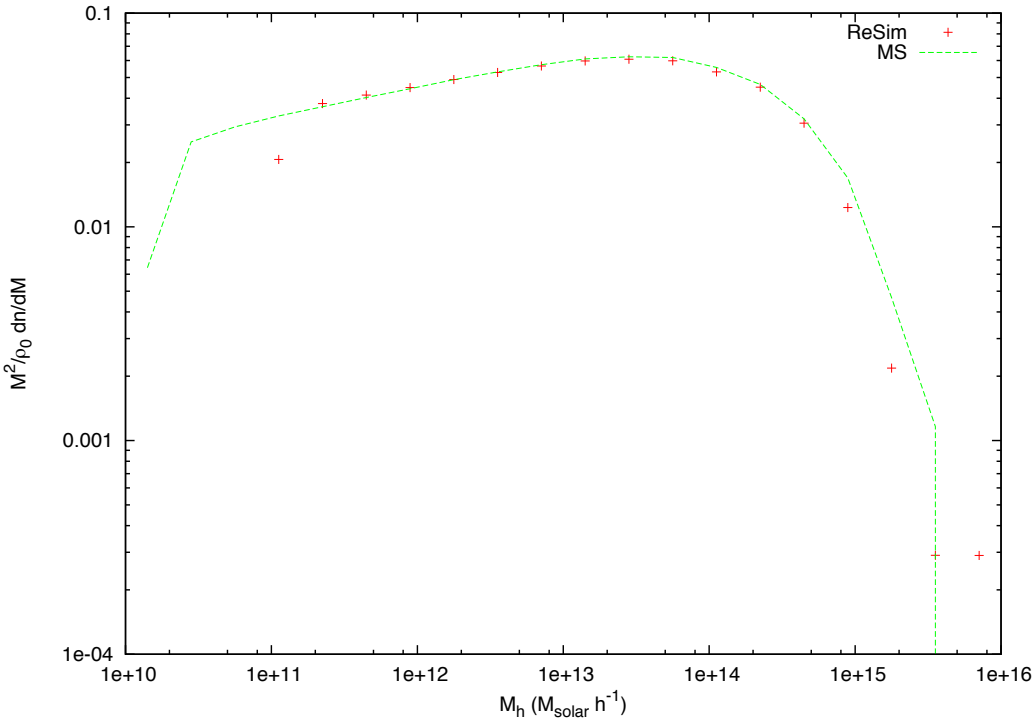

Figure 3.7: Comparison of the halo mass function generated by the Monte Carlo algorithm compared with the haloes in the original MS. We recover the global distribution of halo masses well.

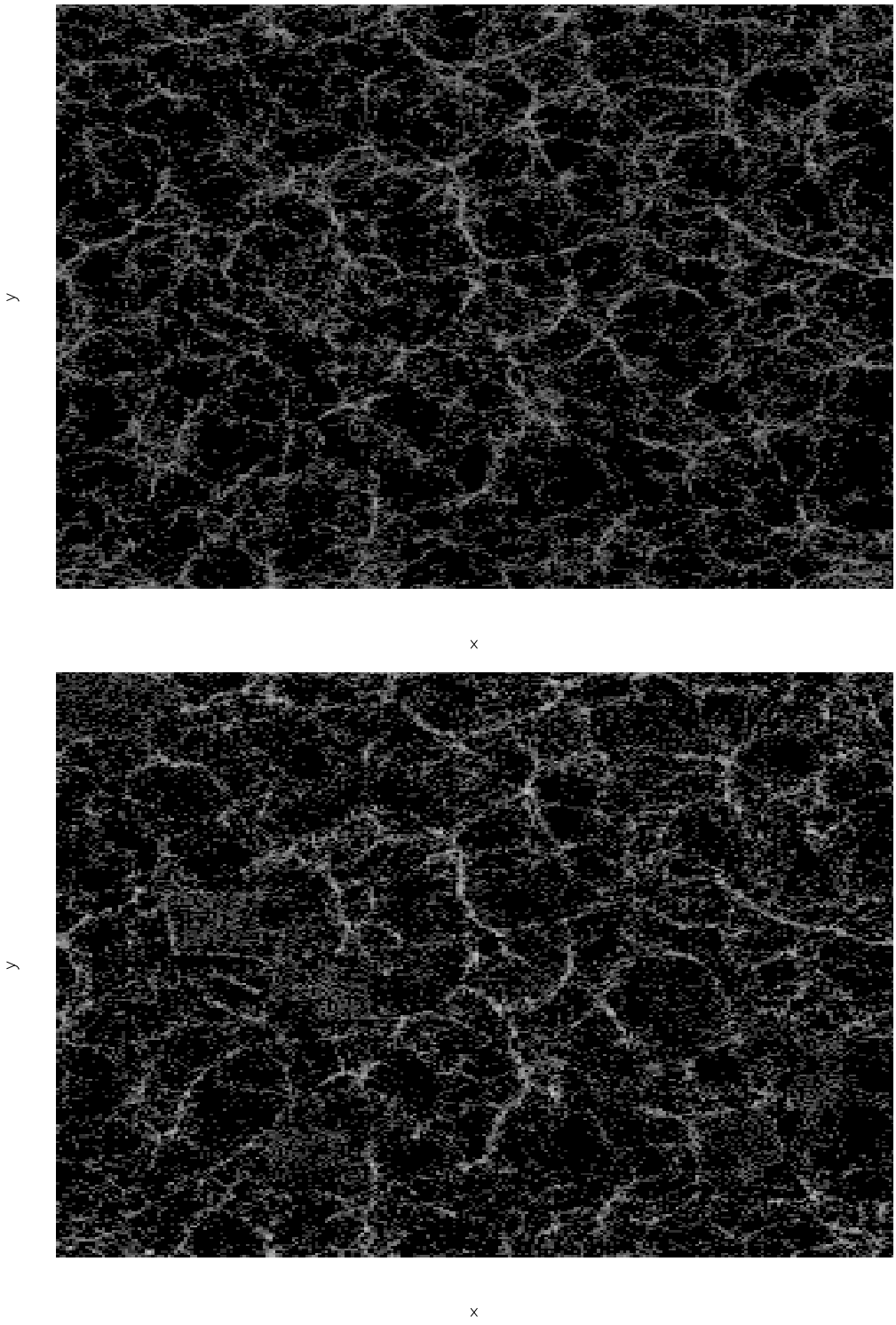

Figure 3.8: A 500 by 500 by 15.6 h^{-1} Mpc slice of the number density of haloes Millennium simulation (top) and the repositioned Millennium simulation haloes within the 256^3 particle resimulation (bottom). The pixel resolution of these images is of order $1\,h^{-1}$ Mpc and the greyscale scheme is logarithmic in the overdensity ($0.1 < \delta < 200$).

the conditional mass function. The first quantity we wish to check we recover is the mass function which is compared with the MS mass function in Figure 3.7. The global mass distribution is recovered well over the range of masses we consider. While the MS mass function continues to haloes with masses of the order of $10^{10}h^{-1}M_{\odot}$, the halo catalogues are only complete down to around 10^{11}, so we only plot our comparison to this point.

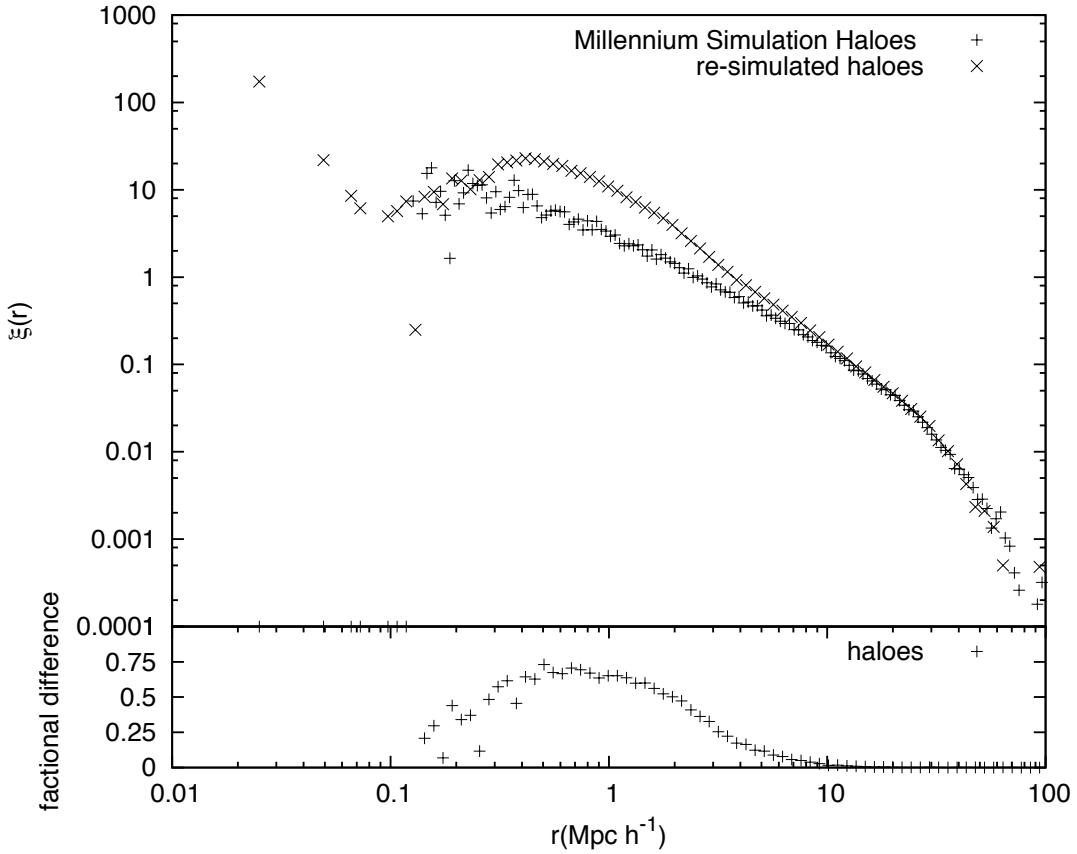

Figure 3.9: The correlation function of the re-simulated halo distribution compared with that of the original MS haloes. The same over-clustering at small scales as seen in Figure 3.6 can be seen here, but outwith this range at larger separations we recover the halo clustering well.

3.3.3 Comparison of galaxy properties

In a similar vein as with the halos, to obtain a better appreciation of the errors introduced with each step in the algorithm, we will approach the comparison

of galaxy clustering in two steps. Initially we will apply the galaxy population algorithm to the original MS haloes and compare the clustering of both red and blue galaxies before applying the same algorithm to the halos generated with the halo algorithm.

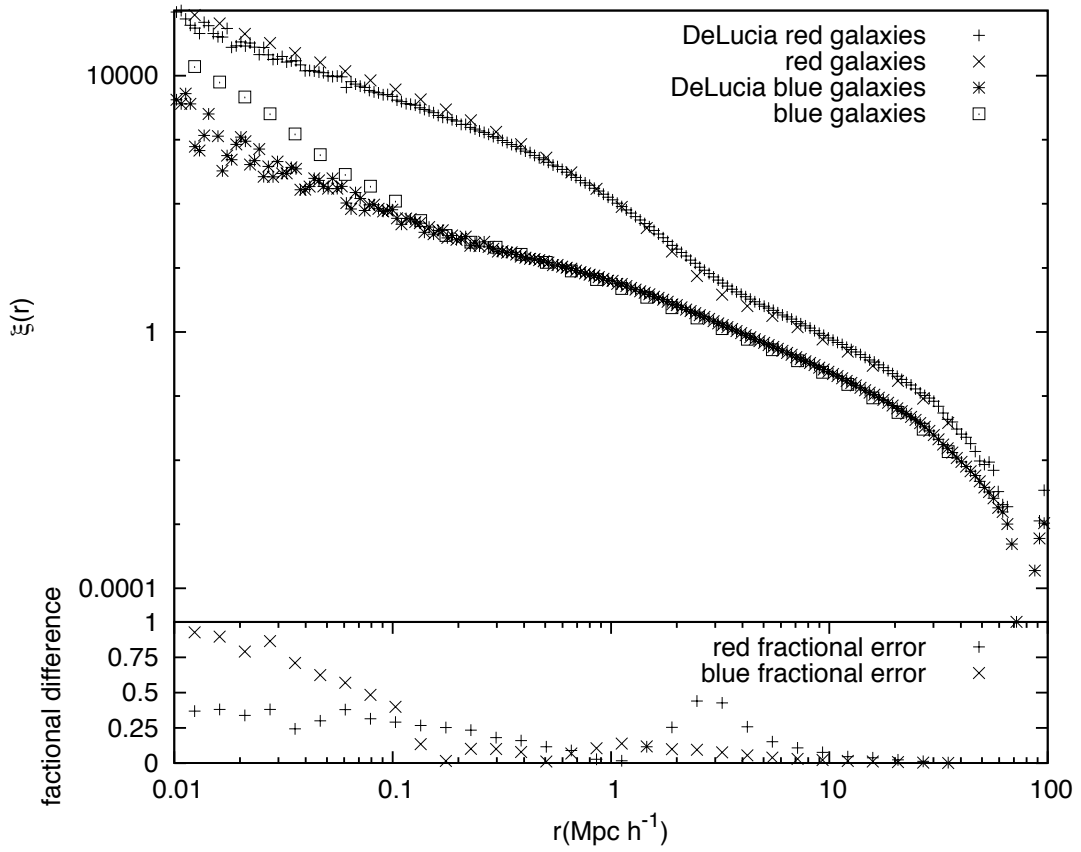

Figure 3.10: The galaxy population algorithm applied to the MS halos split by colour. The split is taken to be at the boundary $M_b - M_v = 0.7$. The clustering shows good agreement with the exception that blue galaxies are more densely clustered at small radii.

Figure 3.10 shows the correlation function of the red and blue galaxies compared with the original MS galaxies. The results show a good agreement with the model with the notable exception of the clustering of blue galaxies at separations $r < 0.1\,h^{-1}\,\mathrm{Mpc}^{-1}$. The most likely explanation of this discrepancy may be because we have not treated the radial bias of the two galaxy populations separately, and the higher clustering of red galaxies at small radii is biasing the blue galaxies. When we consider the total galaxy population as is shown in Figure 3.11

and compare with the clustering of the DeLucia and Bower models, we see that the agreement is excellent and any variations are within the scatter between the two different semi-analytic models. We are confident then that given an accurate halo population we can successfully produce an accurate galaxy population.

We now bring both methods together with the understanding that the major source of imperfection is likely to be over clustering of halos introduced by the halo placement algorithm.

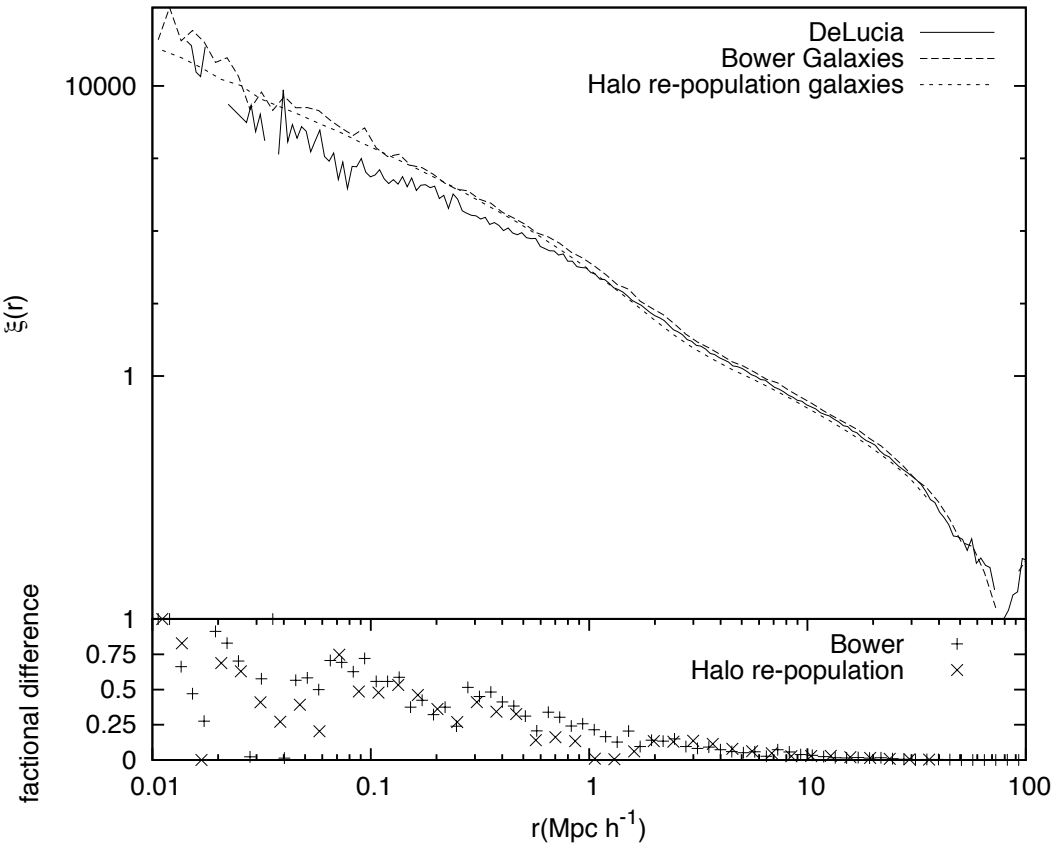

Figure 3.11: Clustering of all galaxies obtained from the halo re-population algorithm compared with the Bower and DeLucia models. The bottom panel shows the fractional difference of both the Bower and re-population galaxies when compared with the Delucia curve.

Having confirmed that we recover the correlation function of galaxies accurately we also want to check that we get the correct results for the luminosity functions of galaxies. Figure 3.12 shows the comparison between the galaxy luminosity functions in the re-simulated galaxy sample and the original DeLuica

galaxies. We recover both the full luminosity function and the luminosity function cut by colour as well. There are small deviations at high luminosity in the full galaxy luminosity function but these are on the order of the deviations we saw between semi-analytics in Figure 2.11.

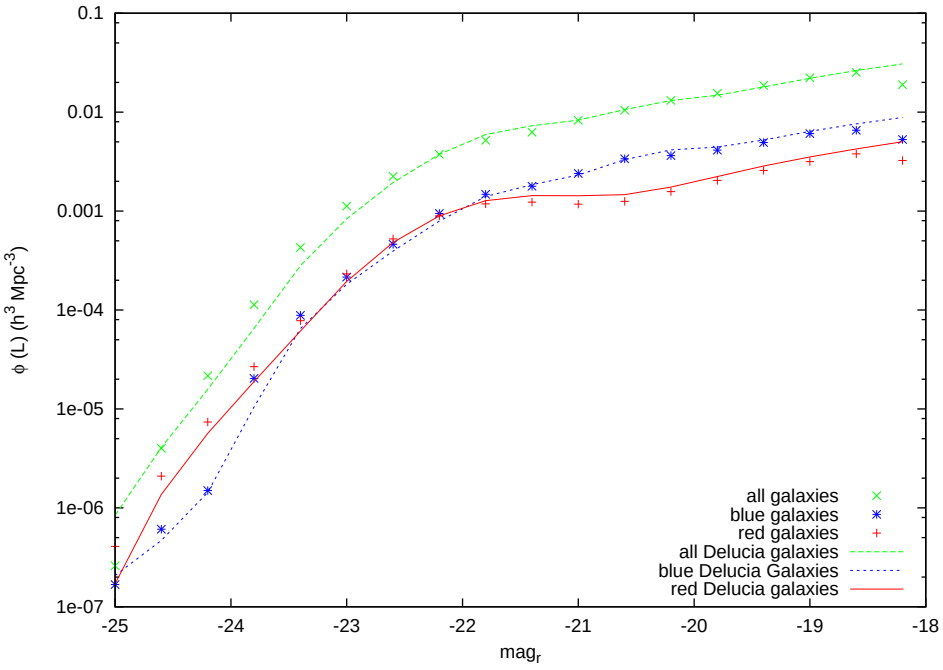

Figure 3.12: The luminosity function comparison between the DeLucia model (lines) and the re-simulated galaxies (points). The total galaxy population, and population split by colour, is presented.

3.3.4 Velocity statistics and redshift space

Until now we have only examined the positional and primary physical properties (halo mass, luminosity and colour) of the mock galaxies and halo populations, however if we are to successfully model galaxy surveys we also need to assign each halo and galaxy a velocity.

In Chapter 5 I will explain how the algorithm can be modified to produce lightcones in which the evolution of the Universe and its halo and galaxy prop-

erties are interpolated in time to produce a realistic sky. When surveys of real galaxies are carried out the quantity measured is not the true distance to the galaxy but rather the redshift. As I discussed in Chapter 1, this is a biased distance measurement, affected by the peculiar velocity of the galaxy. Any simulated mock survey must take these effects in to account.

The first thing to check is that the velocity statistics of the haloes and galaxies within the reproduced MS match those of the original. As there should be no preferred direction in the simulation, we present the distribution of the components of velocity, regardless of Cartesian dimension for the halo centres and the galaxy velocities. Figure 3.13 shows the comparison between the original MS halos and galaxies, and the reconstructed populations. We obtain an excellent agreement between the velocity distribution of both halos and galaxies.

From these velocities we can calculate the redshift space co-ordinates of each of the objects. The boost each receives is:

$$\mathbf{r}_{\rm s} = \mathbf{r} + \frac{\hat{\mathbf{r}} \cdot \mathbf{u}}{H}\hat{\mathbf{r}}. \tag{3.13}$$

When constructing lightcones, this full prescription is necessary but as a test of the algorithm we will appeal to the fact that the Universe is isotropic and simply pick one Cartesian direction as the line of sight. In this case the apparent position of a particle is only modified in one direction (here the z direction) giving :

$$r_{zs} = r_z + \frac{u_z}{H}. \tag{3.14}$$

Figure 3.14 shows this calculated for galaxies in our re-simulation and we can clearly see the characteristics of small scale suppression of power.

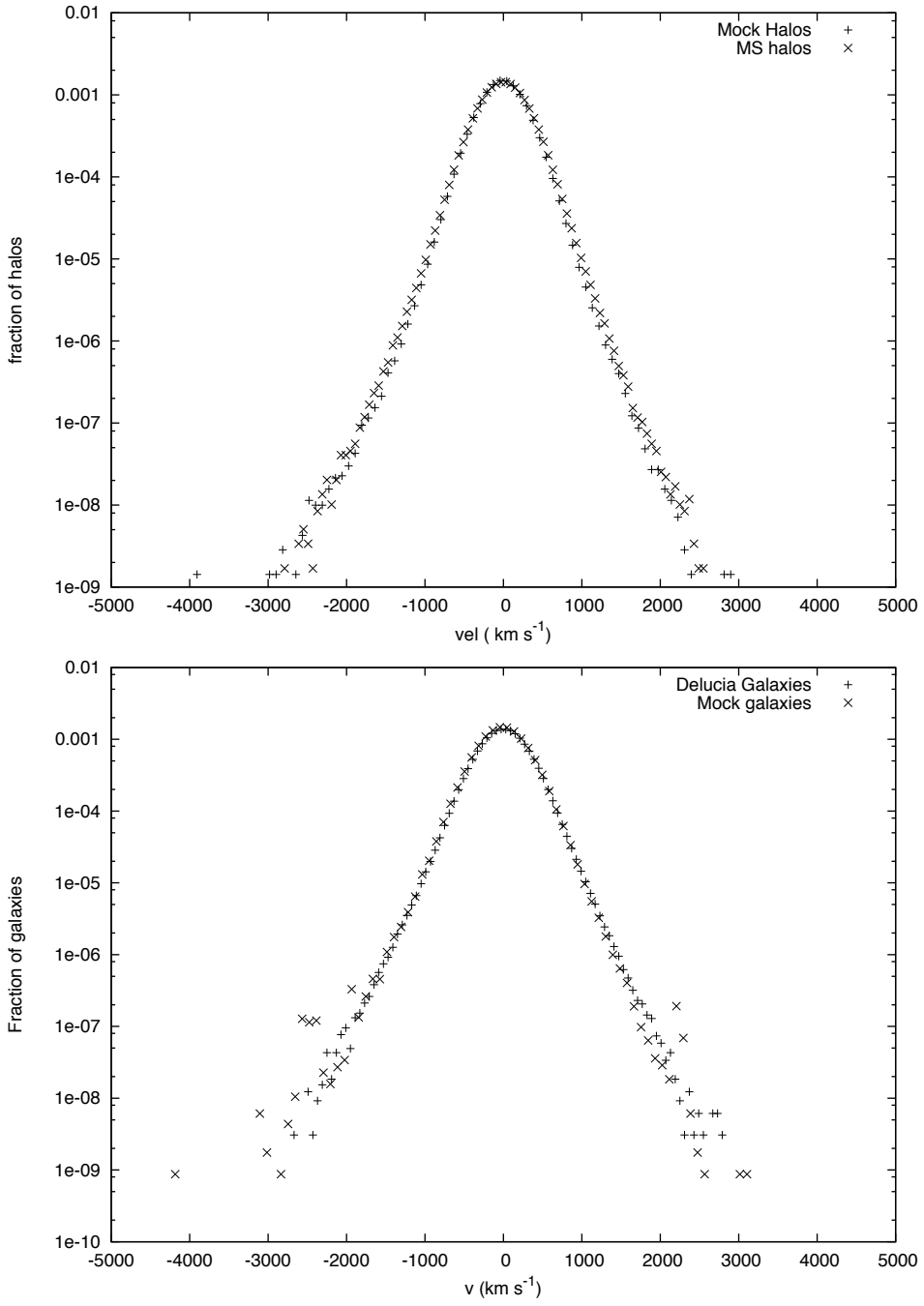

Figure 3.13: A comparison of the velocity distribution of halos (top) and galaxies (bottom) in the millennium simulation and our re-simulation. The algorithm for producing halo velocities from the underlying velocity field and the assignment of velocities within halos reproduces the velocity distribution well.

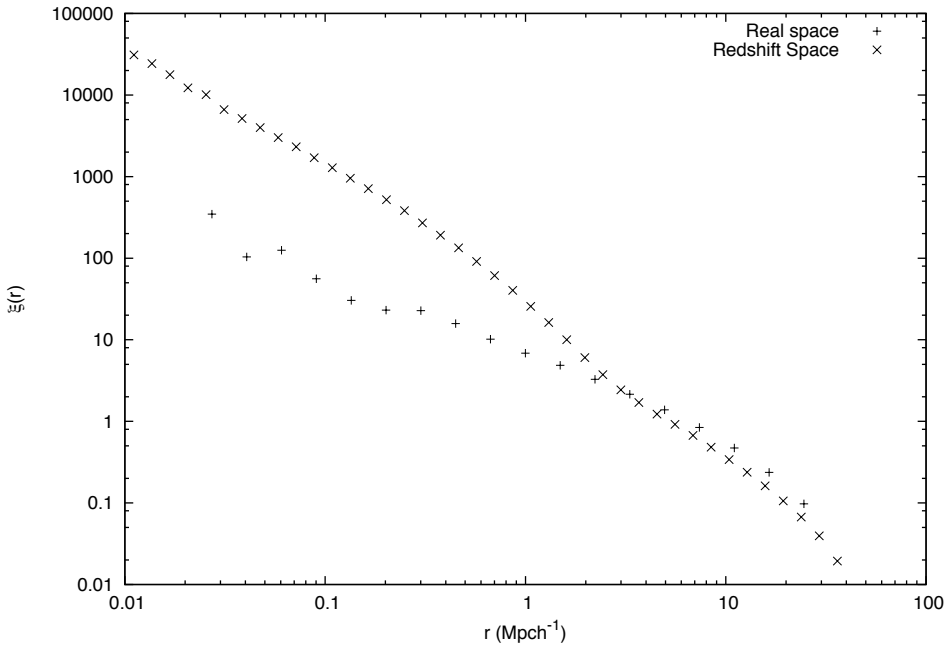

Figure 3.14: A comparison of the redshift space correlation function and the real space correlation function for galaxies in our re-simulation. This has been computed buy taking the z direction and redshifting each particle along it

Chapter 4

Application to baryon acoustic oscillations

One of the most exciting prospects for probing the equation of state and evolution of dark energy is that of Baryon Acoustic Oscillations. As discussed in Chapter 1, sound waves in the primordial plasma imprint a physical scale on the power spectrum of the CMB and also the matter distribution, which is translated into the galaxy power spectrum. Seen as an excess of clustering at around $100\,h^{-1}$ Mpc separation, this feature provides a standard ruler against which the expansion history of the Universe can be determined. As the Universe expands, the angle subtended at $z = 0$ by the BAO scale, is scaled by the angular diameter distance which depends on cosmology, therefore by measuring the BAO scale at two separate epochs we can obtain a measure of the expansion between the given epochs with the precise relation being determined by the Hubble factor (for a flat $\Omega_k = 0$ Universe):

$$H^2(a) = H_0^2(\Omega_m a^{-3} + \Omega_v a^{-3(w+1)}) \tag{4.1}$$

To determine the value of w to 5% requires a determination of the peak of the BAO to 1% (Seo and Eisenstein, 2005). At first inspection to reach this accuracy simply requires us to obtain a large enough sample of galaxies in a large

enough volume. The accuracy with which the power spectrum can be measured is commonly approximated by the expression due to Feldman et al. (1994) :

$$\left(\frac{\sigma}{P}\right)^2 = \frac{2}{n_{modes}}\left(1 + \frac{1}{P\overline{n}}\right), \tag{4.2}$$

where n_{modes} is the expected number of modes in $|k|$ space. The first term in this expression accounts for the variance in the measurement and decreases with the square root of the surveys volume. The second term accounts for the discreteness of the population with the term $P\overline{n}$ measuring the amplitude of the power spectrum in units of Poisson shot noise.

In an ideal world this would allow us to determine the volume and depth of a survey required to accurately measure w. However, we are assuming that the shape of the power spectrum (or correlation function) is unchanged by physical processes unrelated to the expansion. Such systematic errors need to be understood and controlled to a high level to enssure an accurate measurement of the dark energy equation of state. In linear theory the shape of the power-spectrum is fixed between epochs, but this is not true when one begins to consider non-linear structure formation, the process of galaxy formation and redshift space distortions. Each of these physical effects has the ability to distort the BAO signal and shift its peak away from the original value (Angulo et al., 2008), introducing a bias in our measurement of the expansion.

A number of authors have tackled this problem through both theory and simulation. Eisenstein et al. (2007) argued that the dominant effect on the shape of the BAO feature comes from the differential motion of particles originally separated by $100\,h^{-1}$ Mpc. Despite the BAO feature appearing at relatively large separations, the distortions to the signal by the differential motion of particle pairs, is on much smaller scales where the dominant physical forces are those of cluster formation and bulk flows at scales almost 10 times smaller. They follow the displacement of particles separated by $100\,h^{-1}$ Mpc in numerical simulations to determine their final displacements and determine that a good model for the modification to the BAO scale is a suppression to the high k oscillations in the power spectrum :

$$P(\mathbf{k}) = P_{linear}(\mathbf{k}) \exp\left(-\frac{k_\parallel^2}{2\sigma_\parallel^2} - \frac{k_\perp^2}{2\sigma_\perp^2}\right). \tag{4.3}$$

Other authors have examined the distortion through large suites of numerical simulations. Smith et al. (2008) ran a number of numerical simulations with a volume totalling $\sim 100\mathrm{Gpc}^3h^{-3}$ and examined the effect on the correlation function BAO peak. Extending the examination into the realm of galaxies, Angulo et al. (2008) have used semi-analytic methods applied to a very large simulation volume of $2.41h^{-3}\mathrm{Gpc}^3$ to probe the effect on the BAO scale, finding that the typical shift in the galaxy power spectrum is of the order of a few percent.

In this chapter we apply the mock galaxy algorithm to the study of these systematic effects. We produce a number of large $h^{-1}\,\mathrm{Gpc}$ simulations and populate them with halos and galaxies. For each population we measure the power spectrum and correlation function to determine the location of the BAO signal.

4.1 Correlation function and Power Spectrum estimation

In what follows we will be trying to locate the BAO scale feature in our simulations. To accomplish this we need to be confident that we are accurately recovering clustering information from the simulations. Before we proceed with an analysis of our simulations we outline how we calculate both the power spectrum and correlation function.

4.1.1 Correlation Function Estimation

The estimation of the correlation function is the task of counting pairs of particles with a given separation. This can be achieved in a straightforward manner by simply calculating the distances between each pair of particles then binning up the results. This brute force method, while useful for small numbers of particles unfortunately scales in computational power as $O(N^2)$. Even with modern machines this quickly becomes computationally prohibitively expensive.

There are however four alternatives we can appeal to aid us: 1) a gridded correlation function, 2) KDTrees, 3) sparse sampling of the population and 4) Fourier transforms. Each of these methods are able to speed things up by, in the first two methods, treating not individual particles but regions of space, and in

the final method, discarding particles. We examine each of these methods and evaluate their merits and drawbacks.

Gridded Correlation Function

In a gridded correlation function approach we divide the density field into a number of evenly spaced points and assign the mass of each particle in the simulation to this grid using a CIC algorithm. Then we simply calculate the distance between each pair of grid points to obtain the correlation function.

Aside from simply the reduction of the number of distances that are needed to be computed (down from N_{part}^2 to N_{grid}^2) we can achieve further speed gains by taking advantage of the regular nature of the grid. Given a maximum r_{max} scale over which we are interested in, it is easy to identify large number of pairs of cells which will not contribute to the correlation function. This is achieved by noting that the squared distance between to pairs of cells with index (i_1, j_1, k_1) and (i_2, j_2, k_2) is $r_{12} = L_{cell}^2((i_1 - i_2)^2 + (j_1 - j_2)^2 + (k_1 - k_2)^2)$, where L_{cell}^2 is the length of an individual cell. If we hold the first cell constant and iterate over all other cells in the simulation then we only have to consider cells over a range in which each individual cartesian contribution to the separation is less than r_{max}.

The main disadvantage of this method is it is not exact. In binning the density field we are convolving it with an unknown function which introduces a smearing of the correlation function as seen in Figure 4.1. The line in this figure is an exact determination (calculated by the brute force method) of the clustering around the BAO peak while the points represent the clustering calculated by a binning method with 32^2 cells over a $1\,h^{-1}$ Gpc volume.

If we Fourier transform the correlation function we obtained from the brute force method to obtain $FT[\xi(r)]$, and Fourier transform obtained from the grid approach to obtain $FT[\xi_{grid}]$, then the two are related by

$$FT[\xi_{grid}(r)] = FT[\xi(r)]FT[f(r)], \quad f(r) = FT^{-1}[FT[\xi_{grid}(r)]/FT[\xi(r)]], \quad (4.4)$$

rearranging we find an expression for the convolving function which is shown in Figure 4.2 . One approach to accurately measuring the correlation function then would be to fit this function and deconvolve with the result from the grid algorithm. As we will see, however, there are better methods.

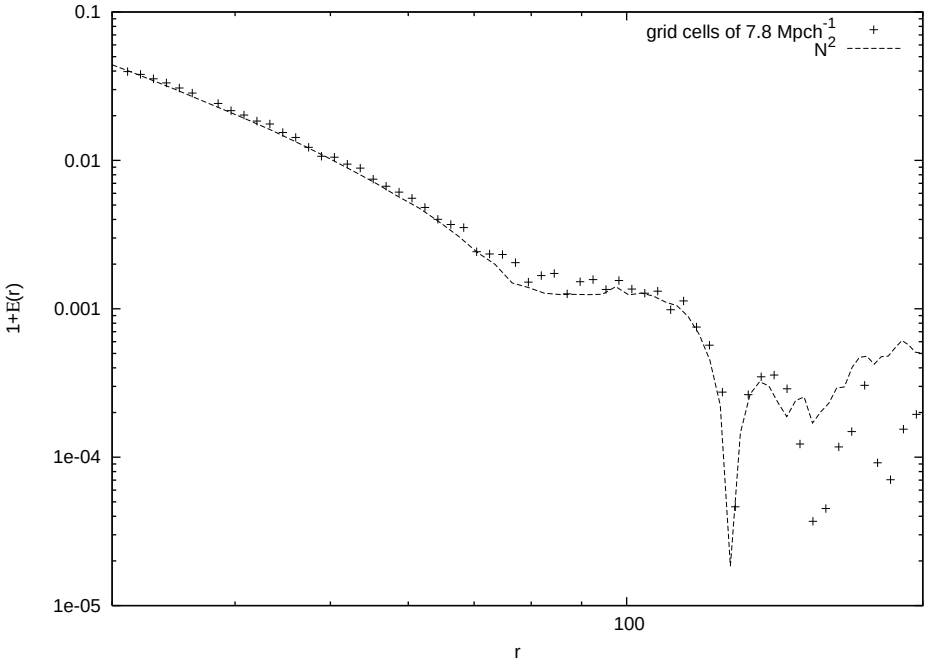

Figure 4.1: A comparison of the correlation function obtained by an exact brute force calculation (the line) and a binning method (the points). The gridded method bins particles in to 32^2 cells over a $1\,h^{-1}\,\mathrm{Gpc}$ volume and then calculates the correlation between the mass weighted grid points. The smearing effect caused by the binning process can be clearly seen as the increase in correlation around $100\,h^{-1}\,\mathrm{Mpc}$.

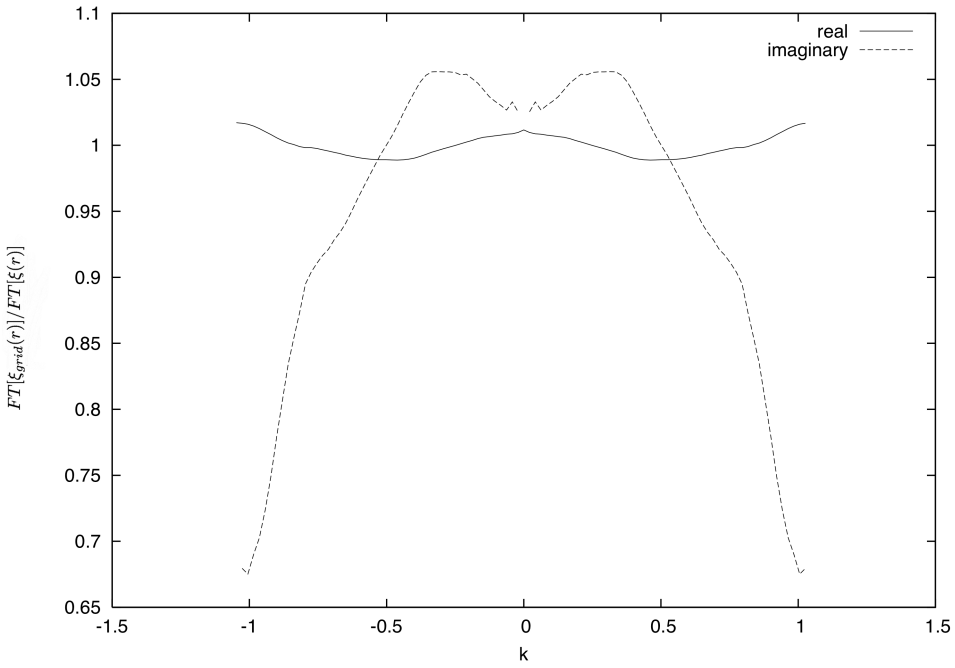

Figure 4.2: Fourier transform of the convolving function in Equation (4.4) obtained by dividing the FT of the correlation function obtained from the gridding algorithm by the Fourier transform of the correlation function obtained with a brute force calculation applied to the same density field.

KDTrees

The KDTrees algorithm and similar space splitting algorithms have been developed and applied to a number of problems. Gray et al. (2004) apply this approach to clustering. The main idea is to reduce the complexity of the problem by geometrically dividing the volume into nested sub-volumes. Each of these nested sub volumes can then be examined to see if they contribute significantly to the correlation function. If they do, the particles the cell contains can then be accounted for without explicitly calculating each particle separation; if not, then the sub cells of the current cell are considered. In detail the method proceeds as follows:

The tree is constructed by defining two threshold values N_t and L_t, we then continually subdivide the simulation volume until each sub region either contains $N < N_t$ particles or has a principle axis of length $L < L_t$. The volumes which satisfy these criteria are known as leaf nodes. We end up with a nested hierarchy of volumes going from the entire simulation volume down to the leaf nodes. In regions where there is a higher particle density the tree with have a larger depth and the volume will be represented by smaller leaf nodes. The recursive function, Split(s, N_t, L_t) which splits the space down to the lead nodes is as follows in pseudocode:

Split(s, N_t, L_t)

- Let N be the current number of particles in the current sub-volume
- Let L be the largest dimension of the current sub-volume
- if ($N < N_t$ OR $L < L_t$)
- Sub-volume is a leaf node return;
- else
- sub-volume needs to be divided further: Bisect the box across its longest Cartesian axis.
- Let s_{left} and s_{right} denote the new sub-volumes
- Split (s_{left} , N_t , L_t)

- Split (s_{right}, N_t,L_t)

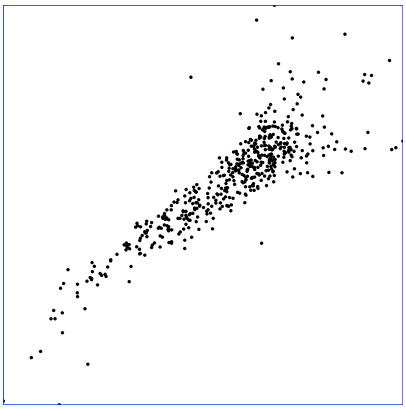

Figure 1a: The top node of a kdtree is simply a hyper-rectangle surrounding the data-points.

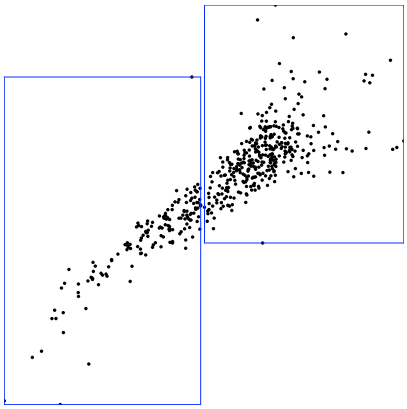

Figure 1b: The second level contains two nodes.

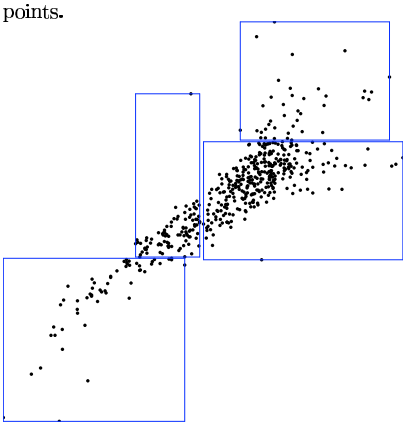

Figure 1c: The third level contains four nodes. Note how a parent node creates its two children by splitting in the centers of its widest dimension

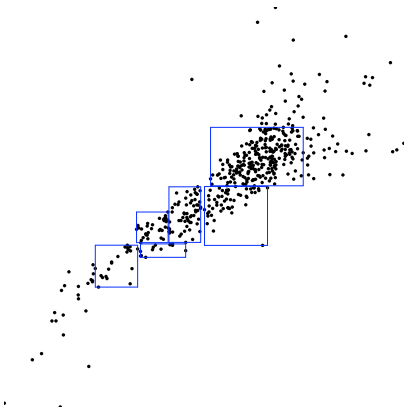

Figure 1d: The set of nodes in the sixth level of the tree.

Figure 4.3: An illustration of the production of a KDtree taken from Gray et al. (2004)

This produces a tree like the one in Figure 4.3 which naturally traces the density field: regions of high particle number are subdivided into smaller sub-volumes than those of relatively low particle number. Once a tree has been constructed we can calculate the number of particles which lie in a distance range, $r_e < r < r_s$, from a particle P by walking the tree: descending down through successive sub-volumes. We can save time on the calculation if any one of the following criteria are met:

1. A given sub-volume which has closest and furthest distance to P, D_{min} and D_{max} respectively. Figure 4.5

2. If the sub-volume is completely within our search radius (Figure 4.4), $r_s < D_{min} < D_{max} < r_e$, we simply count the number of particles within that volume.

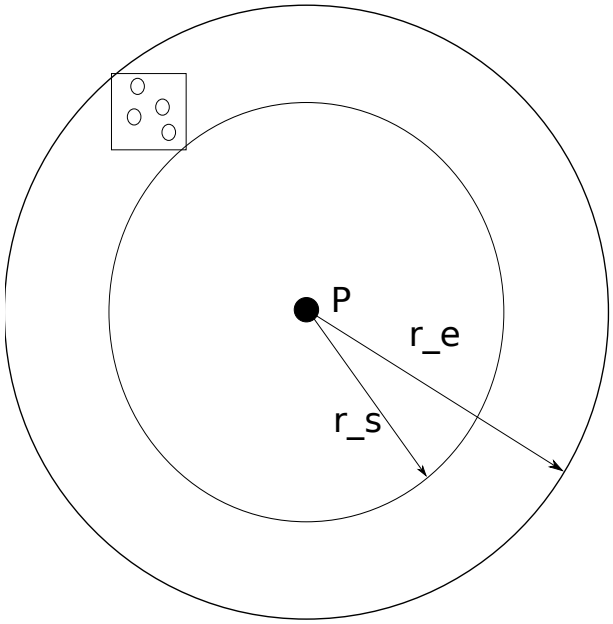

Figure 4.4: A node entirely within the search radius. We know that every single one of these particles contributes to the count. We don't need to individually check the distance. Instead we can just add the number of particles in this node directly to the counts running total.

3. If the entire sub-volume is closer to P than D_{min} (Figure 4.5) or more distant than D_{max} (Figure 4.6, we simply ignore all the particles it contains

For volumes which meet neither of these criteria ie those that straddle the borders of the search radii, we descend another level down the tree and apply the above criteria to the lower sub-volumes until they are met or we reach a leaf volume in which case we have to give up and calculate the correlation function using the brute force method.

The full algorithm is as follows, we define rangeCount(node) as the following recursive algorithm :

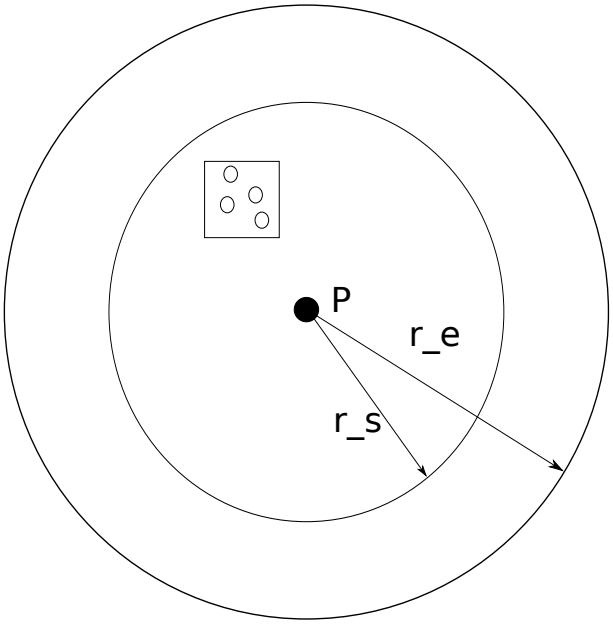

Figure 4.5: A node entirely inside the search radius. None of these particles contribute to the count and so can be ignored without explicitly calculating the distance to each

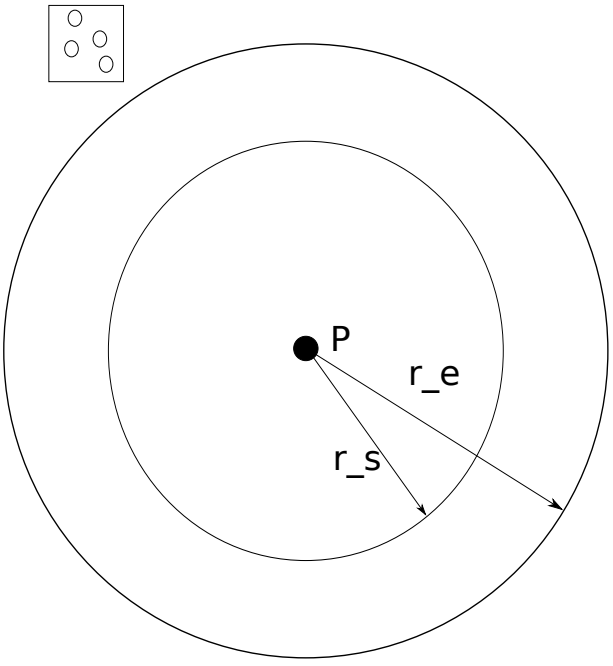

Figure 4.6: A node entirely outside the search radius. None of these particles contribute to the count and so can be ignored without explicitly calculating the distance to each

- Let n be the cell we are considering which has sub cells $n \rightarrow left$ and $n \rightarrow right$.
- Let P be the centre of our search
- Let the range we wish to count particles in be $r_s < r < r_e$
- Let D_{max} be the largest distance between P and n
- Let D_{min} be the smallest distance between P and n
- if $r_s > D_{max}$ then all possible particles in this cell lie closer to P than we are interested. return 0
- if $r_e < D_{min}$ then all possible particles in this cell lie further away from P than we are interested. return 0
- if $D_{min} < r_s < r_e < D_{max}$ then all particles must lie within the range we are interested in. return n_{count}
- if n is a leaf node: calculate by brute force the number of particles in this cell which lie between r_s and r_e from P.
- else return $rangeCount(n \rightarrow left) + rangeCount(n \rightarrow right)$

The rangeCount() algorithm is them called on the root note to construct the tree. As we can see from Figure 4.7, by implementing this pruning of the tree when sub-cells are not relevant to our search and subsuming cells which fall entirely within our search criteria we drastically reduce the number of N^2 pairs we have to calculate and so dramatically speed up the correlation function calculation.

Further speed gains can be obtained by using two identical trees and comparing cell with cell rather than particle with cell. While this approach is useful for speeding up the calculation of a correlation function over a small scale, we found that when applied to the large ranges, r_{min} to r_{max}, over which we wanted results, it was hard to maintain the speed boost. While we still obtain the speed up from discarding regions which lie completely outside of the range in which we are interested, to obtain the speed boosts for subsuming particles in cells which are completely contained within a small distance bin requires navigating down

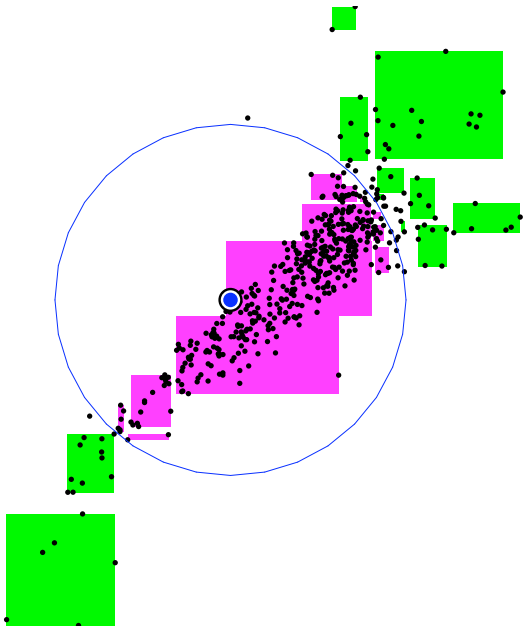

Figure 4.7: An illustration of the range search performed on the tree for a position centred at the blue circle. For an annulus where $r_s = 0$ and r_e is some constant, the tree nodes which are entirely within the annulus (purple nodes) can be completely subsumed, adding the number of particles that node contains to running totally without the need to perform a distance check. Nodes which lie entirely outside the annulus (green nodes) are ignored.

almost to the leaves of the tree. In our experience for small bins, the overhead of doing this (calculating the maximum and minimum distances between cells) overwhelms the benefit of subsuming particles.

Hybrid grid / sparse sampling

For the results presented here we settled on a hybrid of both of these methods, which is simple but effective. In a similar way to the KDTree method we assign particles to cells, but in this case we simply use a single level regular grid. Then we examine pairs of cells, calculate their maximum and minimum separations and discard them if they lie completely outside of our range in much the same way as the KDTree algorithm. For the remaining cells we calculate the particle distances directly.

While this still requires some sparse sampling of the population in order to run in a reasonable timescale, the level of sampling in question is at about the 10% level for the simulation sizes present here. We have confirmed that this is more than enough particles to accurately determine the large scale features of the correlation function.

4.1.2 Power Spectrum Estimation

The calculation of the power spectrum is a two stage process: the particles are binned onto a regular 3D grid to obtain an over-density field; this field is then Fourier transformed and the resulting modes are spherically averaged in annuli spanning k to $k + dk$. There are a number of methods of assigning particle mass to the 3D grid. In the simplest case the total mass of the particle is assigned to the grid point nearest to it, but this approach neglects the fact that each simulation particle actually represents an extended element of the smooth field we are trying to model. There are a number of schemes which aim to correct for this by assigning a fraction of the particle mass to each surrounding grid point. One of these, the cloud in cell (CIC) scheme is the one we adopt here.

There are a number of corrections which need to be made to the resulting power spectrum:

- Expected number of modes: Depending on the coarseness of our binning in

shells of $|k|$, a given grid point in Fourier space will lie either inside of a shell or outside while the power associated with that point is actually distributed smoothly between these two extremes. Because of this we correct for the expected number of modes in a k bin in the continuum limit :

$$N(k) = \left(\frac{L}{2\pi}\right)^3 4\pi k^3, \tag{4.5}$$

where L is the length of the side of the simulation box.

- The act of binning particles onto a density field means that what we are actually measuring when we measure the power spectrum is not the true power of the particle distribution but rather the density field convolved with a grid of top hat functions. In Fourier space this becomes a multiplication of sinc function which we correct for with the factor :

$$P(k) = \grave{P}(k) * (1 + \frac{L^2}{12n^2}k^2), \tag{4.6}$$

where the n is the number of bins (in each dimension) we are using to produce the density field and L is the length of the side of the simulation box.

- Shot noise: Finally the discreteness of particles means that on small scales the power spectrum is affected by shot noise. For a population of particles with a spatial density of $\overline{n}$, the shot noise is subtracted from the power spectrum as:

$$P(k) = P_{measured} - \frac{1}{\overline{n}}. \tag{4.7}$$

4.1.3 Picking out the BAO signal

Once we have obtained a measure of the power spectrum and correlation function we are required to model the position of the BAO feature. This is done most easily in Fourier space for a number of reasons; although the signal is more complicated there, there are a number of oscillations rather than a single peak and the background contribution from a smooth spectrum is more easily dealt with. Our approach is to produce a smoothed version of the underlying power spectrum with no BAOs and then divide the measured spectrum by the smoothed

one, leaving a relatively clean BAO signal which can then be fit (Parkinson et al., 2007).

There are a number of possible ways of producing a smooth reference spectrum. We can use a linear zero baryon power spectrum, where the baryons can either be omitted leaving Ω_m less than in the simulation, or the baryons can be converted to dark matter keeping Ω_m the same (Blake and Glazebrook, 2003). This method has the drawback that it does not account for any large scale non-linear evolution of the power spectrum. Even when applied to a linear power spectrum there is a tendency for this approach to introduce a tilt in the resulting BAO signal. An alternative method is to coarsely re-bin the spectrum into bins which are large compared with the oscillations. A cubic spline can then be used to fit the general shape of the spectrum (Percival et al., 2010). The benefit of this approach is that it produces a level representation of the BAOs and also accounts for some of the non-linear features in the measured spectrum. In this work we calculate the coarse power spectrum by resampling the power spectrum into 10 bins from $k = 0.005$ to $k = 1.5$. The cubic spline fit is then constrained to pass through each of these points. Figure 4.8 shows a comparison of both methods clearly showing the residual tilt in the zero baryon model.

After this reference spectrum has been calculated we are left with the quantity:

$$P^{BAO} = \frac{P(k)}{P_{\rm ref}(k)}. \tag{4.8}$$

Regardless of how this is done we also require a theoretical model for the expected BAO signature with some variable shift α. To construct this we take a linear power spectrum constructed from the fitting formula presented in Eisenstein and Hu (1998) and apply the same spline fitting method of producing a smoothed spectrum. We then modify the reference spectrum as :

$$P^{BAO}_{ref}(k) = \lambda P^{BAO}_{lin}(k\alpha)\exp(-k^2/k^2_{nl}) \tag{4.9}$$

where α is the variable which will measure the shift in the BAO scale away from the expected value $\alpha = k_{app}/k_{true}$, Λ accounts for overall normalisation and k^2_{nl} is a dampening factor to dampen the large k oscillations. The motivation for using a linear spectrum to map the non-linear P^{BAO}_{ref} as opposed to fitting a shift to the evolved dark matter power spectrum, is that we are trying to compare, for

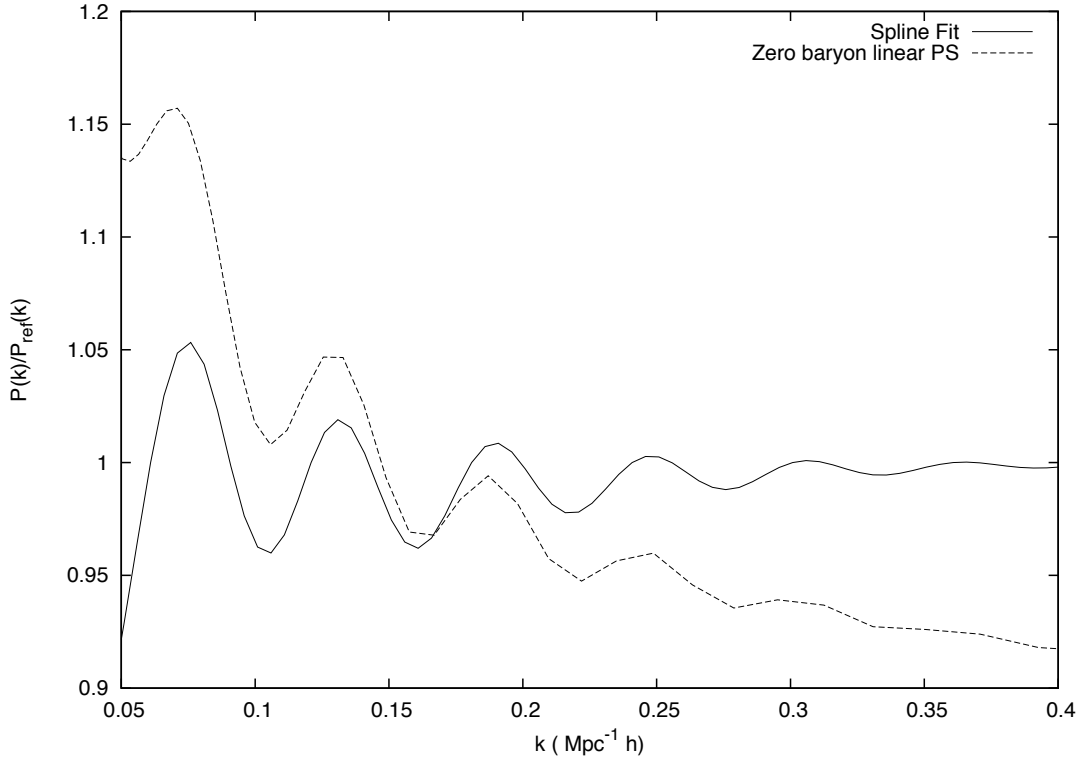

Figure 4.8: A comparison of the effect of using a zero baryon model and a spline fit as the reference spectrum for picking out the BAO signal. The spline fit does a much better job of producing a BAO signal without imposing an overall gradient and is therefore much more appealing.

each of the species of interest, not to the location of the BAO signature in the underlying evolved dark matter field, but it to the original location of the BAO signature at high redshift. It is this shift that will ultimately be measured from surveys as it is hard to probe the power spectrum of the dark matter field.

We define a likelihood function for each combination of λ, α and k_{nl}:

$$-2\ln L = \chi^2 = \sum_i \left(\frac{P^{BAO}(k_i) - P^{BAO}_{ref}(k_i)}{\sigma^i / P^i} \right)^2 , \tag{4.10}$$

where the error term σ^i/P^i is taken to be the same as in Equation (4.2) and this equation is valid only in the absence of mode coupling.

4.1.4 Realisations

The Millennium Simulation volume is small in comparison with the BAO scale of $\sim 100\,h^{-1}$ Mpc containing only a small number of Fourier modes which probe it. To be able to properly examine the BAO signal we require larger simulation volumes. To this end we have run a suite of 5 simulations of $1\,h^{-1}$ Gpc and populate each using the algorithm described previously. The cosmology used in each simulation is identical to that of the MS and the accuracy characteristics are identical to those used to create the coarse MS simulation in Chapter 3. 512^3 particles are used in each simulation box and the force grid used by the PM code is also increased to 512^3 points. Each realisation is started with a different random seed and the initial conditions are produced in an identical manner to the coarse MS simulation. Figure 4.9 shows the mass distribution obtained from one of the simulation volumes. For each of the 5 simulations we calculate the power spectrum and divide by a spline curve as explained above. We calculate the likelihood function on a grid of 1000 by 200 different values of α and λ in the ranges [0.8,1.2] and [0.5,1.5] respectively. As we appear to have some systematic effects at large k we only fit over the first two BAO peaks in the range $0.1 < k < 0.15$. Figure 4.10 shows the BAO power spectrum with reference to the linear theory power spectrum while Figure 4.11 shows the same results with respect to their best fitting curves. The parameters for each best fit is shown in Table 4.1.5. The recovered values for k_{nl} are all at the high end of the fitting range. This is not surprising as we are fitting only over the first two peaks where

damping is negligible. The shifts measured in the BAO are in line with what others (Angulo et al., 2008) have found, a shift of a few per cent. However, the value for the halo shift is much greater than we would expect. This suggests that our fitting algorithm may be failing to accurately recover the power spectrum in this case. Both the red and blue galaxies are shifted at the level of around 2% and 3% respectively in line with what we would expect. The errors on the values of the shift are however large and its hard to tell if these results are significant at all.We note that with a better determination of the power spectrum, we could fit to higher k which would improve our results. Similarly a larger number of larger volume simulations is probably required to reduce the cosmic variance on the largest scales, allowing a better determination of the first BAO peak.

4.1.5 Implications for future studies

Should we be worried about a 1% shift in the position of the acoustic scale in the local Universe power spectrum? In terms of the corresponding shift in w this could be a large source of error. If unaccounted for the estimates obtained from future surveys could be heavily systematically biased by such a shift.

They would be wrong because they have assumed the linear value for the BAO scale. In practice if we have a good enough model for the shift in α with redshift and galaxy class we can fit instead to the model prediction of $\alpha \neq 1$.

Population	α	k_{nl}	σ_α
Dark Matter	0.98	0.38	0.02
Red Galaxies	0.98	0.39	0.015
Blue Galaxies	0.97	0.39	0.02
Halos	0.96	0.398	0.03

Table 4.1: The best fitting parameters for the BAO scale for the different populations. The first column is the apparent shift in the BAO scale, the second if the fit parameter for the dampening of the large k oscillations and σ_α is the standard deviation of the shift calculated from 5 realisations. There is no statistically significant shift found but the general trend seems encouraging.

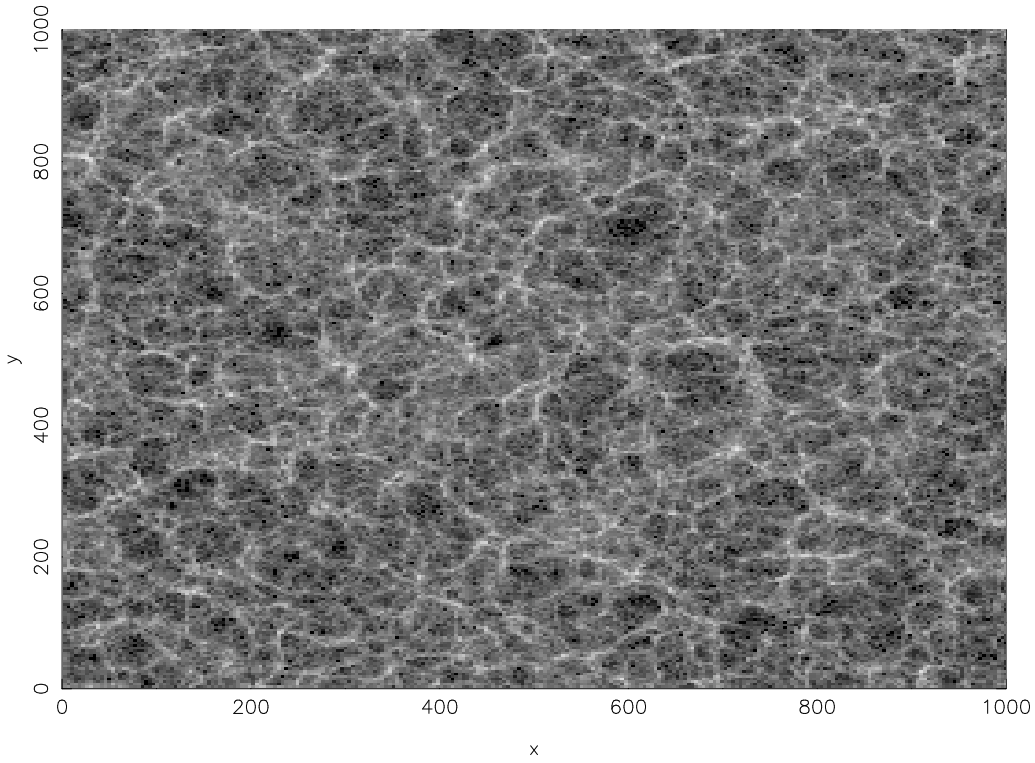

Figure 4.9: A $1000\,h^{-1}$ Mpc squared $15\,h^{-1}$ Mpc slice from a $1\,h^{-1}$ Gpc simulation.

4.2 Conclusions

We have used out method to probe the BAO signal and have tried to determine how non-linearities effect the location of the BAO's for the four populations : red galaxies, blue galaxies, haloes and dark matter. While it is disappointing that the simulation sizes we have run struggle to give us good enough statistical results to say conclusively the magnitude of these shifts, we are encouraged that this method seems to be capable of modelling the BAO signal and has produced results which are broadly what other authors have found.

In the future we plan to run a great deal many more simulations to improve the statistics of the current box size's and run a much larger boxes up to $3000\,h^{-1}$ Mpc on a side. This should hopefully let us examine even closer the results of non-linear structure growth and galaxy formation on the BAO scale. We also hope to explore other methods for extracting the BAO signal.

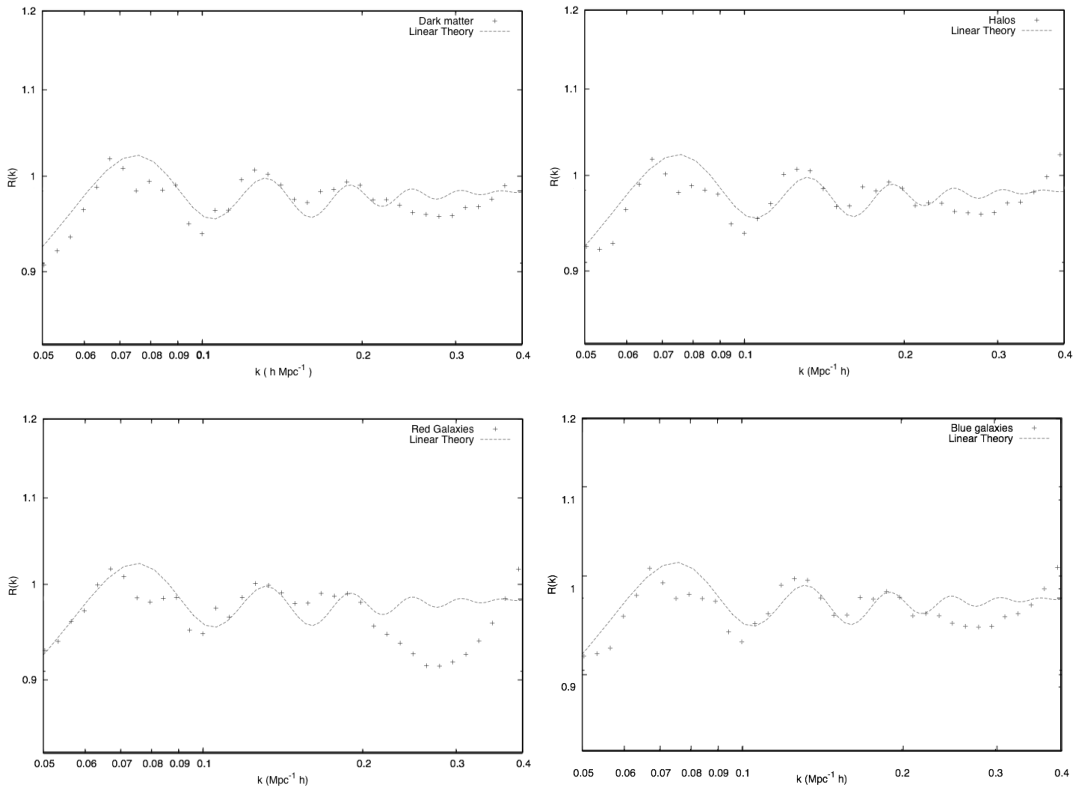

Figure 4.10: The points are a measure of the average power spectrum from 5, $1\,h^{-1}\,\mathrm{Gpc}$ simulations at $z = 0$ for each population. The power spectrum has been divided by a smooth power spectrum obtained by fitting with a cubic spline curve. This was obtained by fitting to a coarse re-binning of the power spectrum into 10 bins over the range $k = 0.005$ to $k = 1.5$. The line is the linear theory prediction for the simulation which has been divided by a reference spectrum calculated in the same way.

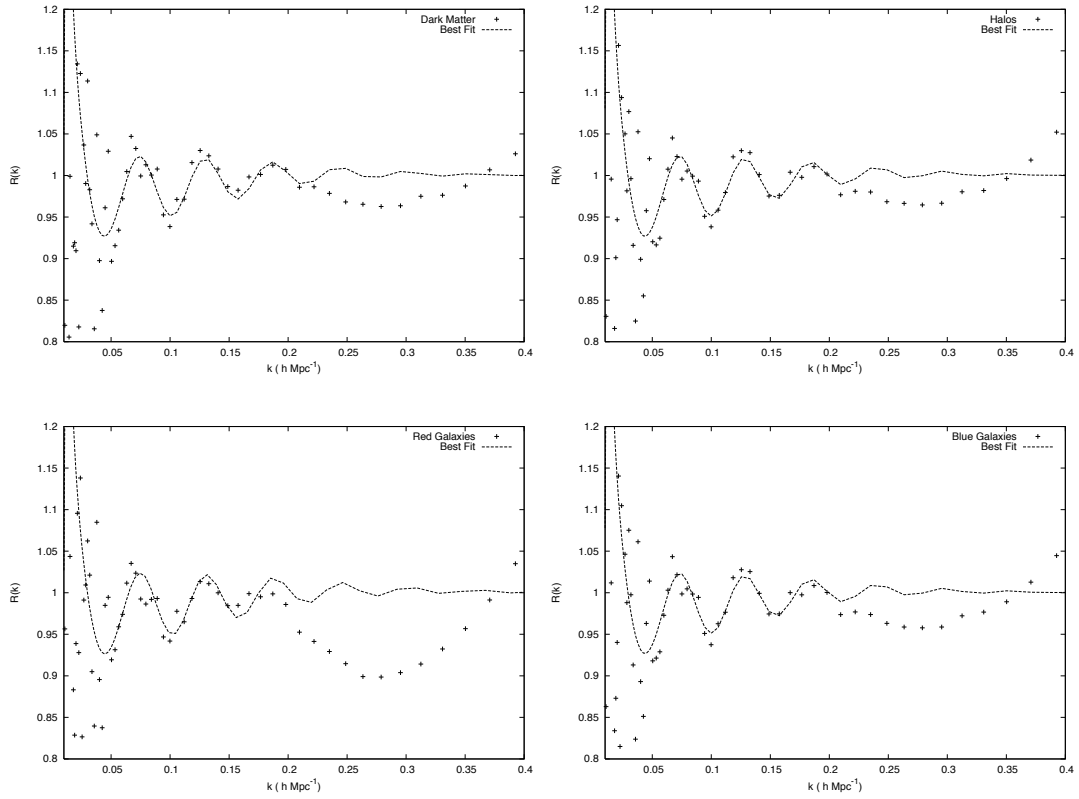

Figure 4.11: As with Figure 4.10, but now the line represents the best fit. The parameters for each best fit (over the range $0.1 < k < 0.15$) value of α and k_{nl} can be found in table 4.1.5.

Chapter 5

Lightcones

Up till now we have applied the mock algorithm to snapshots of the evolution of the Universe. This has allowed us to compare with semi-analytic and theoretical predictions for clustering. However in the real world observers do not simply view the Universe at a single timestep but rather along a lightcone. In this chapter I will discuss how the algorithm can be modified to produce realistic skies which incorporate evolution.

5.1 Box Stacking

If the evolution of the Universe is ignored then the next most immediate problem with producing lightcones which are comparable in scale and depth to real surveys is that our boxes do not have the required volume. This becomes even more of a problem when one considers that future surveys will cover almost the entire sky to redshifts $z > 1$.

A single simulation box can therefore not be used to construct the desired lightcone and we need to stack boxes along the line of sight. This however can lead to structures repeating along the light of sight which will adversely affect clustering and lensing statistics as well as producing visible artefacts. To combat this problem some authors (Kitzbichler and White, 2007; Blaizot et al., 2005)

adopt the approach of applying a number of transforms to boxes along the line of site. Both random translations to the centroid, Equation (5.1), and rotations, Equation (5.2), of the box are designed to ensure that along the light of sight objects are not repeated:

$$\acute{\mathbf{x}} = \mathbf{x} + \mathbf{a} \tag{5.1}$$

$$\acute{\mathbf{x}} = \mathbf{R}\mathbf{x} \tag{5.2}$$

This approach however has its own complications. If large haloes or structures such as filaments span the edge of the box, rotations and translations will break up the mass density producing sharp discontinuities. In full particle simulations there is also the added complication of determining the red-shift at which a halo comes into being or the redshift at which a merger occurs. As haloes are only identified at given time-steps we require a high time resolution to be accurate in determining the state of a halo.

5.2 Constructing a light cone

Our approach to constructing a lightcone is similar to that of populating a single snapshot, but we need to take into account both the larger volume that is required and the evolution of structures in the Universe. We start by running a $1\,h^{-1}\,\mathrm{Gpc}$ simulation, outputting particle positions and velocities every $\Delta z = 0.3$ from a redshift of $z = 3$. Given a lightcone geometry: $\theta_{min}<\theta_{max}$, $\phi_{min}<\phi<\phi_{max}$ and $z<z_{max}$, we determine the maximum number of boxes required to tile the volume. For each box we then subdivide the volume into subcells of the same volume as used in Chapter 3 and determine the distance from the observer to the centre of that box r_b. The redshift at which the observer would view that box, z_b, is then calculated from the distance redshift relation discussed in Chapter 1. This redshift is used in the subsequent determination of the properties of the halo and galaxy contents of this region.

(5.3)

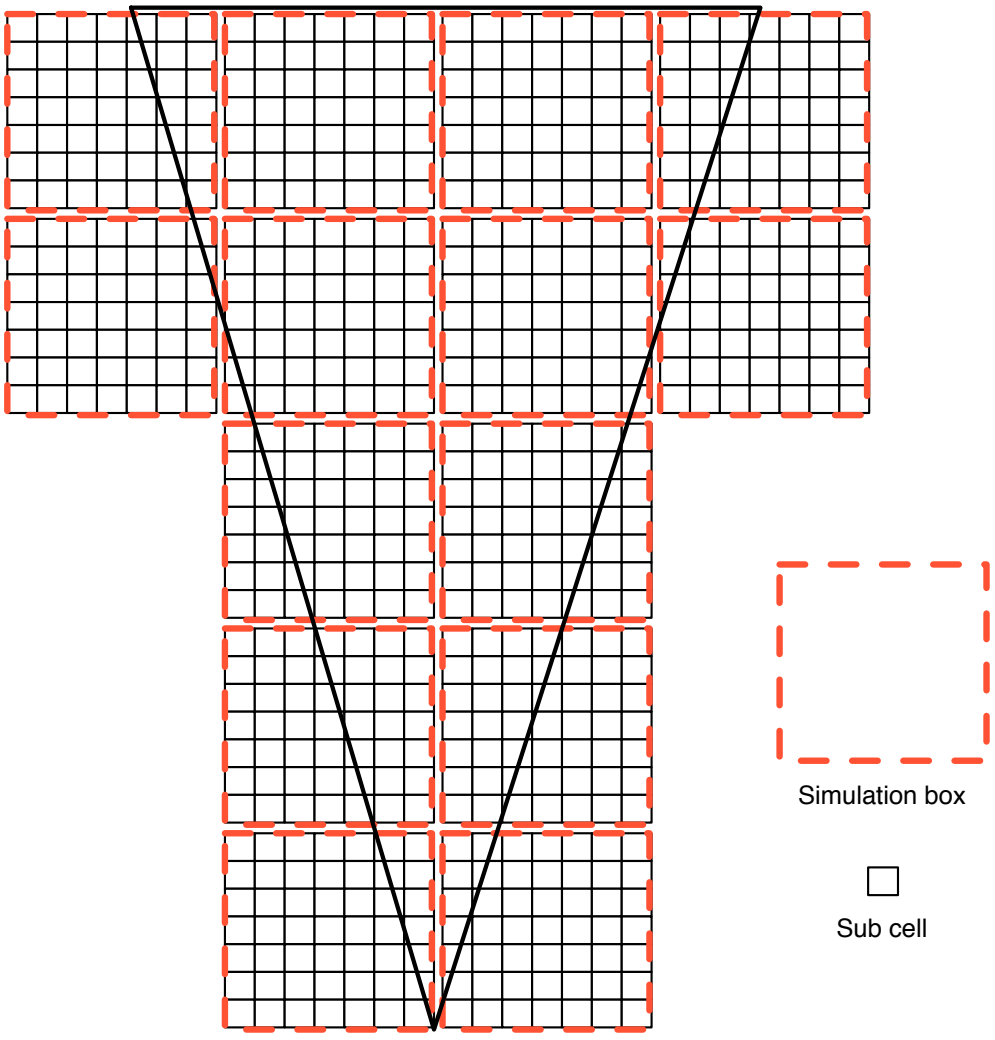

Figure 5.1: Schematic of the scheme used for light-cone production. Boxes of the size of the simulation box are repeated to span the extent of the light-cone. Each of these simulations is divided into a number of sub cells.

Our cells are small enough that the change in redshift between the point nearest to the observer and the point furthest from the observer is much smaller than the output frequency of the simulation. For each cell in the simulation we next identify the two snapshots that bracket the subcell. We now have the density and particle distribution at z_{s2} and z_{s1} such that $z_{s1}<z<z_{s2}$. The idea is to interpolate both of these quantities to the cell redshift and combine them with interpolated conditional mass function statistics using the same algorithm prescribed in Chapter 3 to produce a halo catalogue for that cell.

We interpolate the density linearly in the expansion factor between the two outputs to give:

$$\delta_b = \delta_2 + \frac{\delta_1 - \delta_2}{a_2 - a_1}(a_2 - a_b) \tag{5.4}$$

We need to be more careful however in interpolating the particle positions as between the initial and final timesteps not all the particles contained within the cell will be the same at both timesteps. Simulation particles can both move in and out of the cell during the time step and we need to be aware of this. We could interpolate the particles contained in 16 surrounding cells to ensure that any particles entering the cell between $z = 2$ and z_b are accounted for, but we take the philosophy here that as we are not trying to directly identify haloes from the particles and instead are only using the particles in the placement algorithm, we can afford to lose a few of the particles.

Each particle from the simulation is assigned a unique id which allows us to track them from timestep to timestep, and so for each particle which exists in the box at both redshifts we have 12 degrees of freedom: 6 Cartesian position co-ordinates and 6 velocity components. This leads naturally to using a 3rd order quadratic to interpolate the particle positions. We use equation 5.5 then to interpolate our particles to the correct redshift for the cell. For each Cartesian coordinate we set

$$x(t) = a + ba(t) + ca(t)^2 + da(t)^3 \quad v(t) = b + 2ca(t) + 3da(t)^2 \tag{5.5}$$

where we have differentiated $v_i = \dot{x}_i$. This must hold at both timestep $a(t_1) = a_1$ for x_1 and v_1 and timestep $a(t_2) = a_2$ for x_2 and v_2, so we need to solve for

the four coefficients a, b, c, d :

$$\begin{pmatrix} x_1 \\ x_2 \\ v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 & a_1 & a_1^2 & a_1^3 \\ 0 & 1 & 2a_1 & 3a_1^2 \\ 1 & a_2 & a_2^2 & a_2^3 \\ 0 & 1 & 2a_2 & 3a_2^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \tag{5.6}$$

We can then invert the matrix to obtain a,b,c and d as:

$$\begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\ v_1 \\ v_2 \end{pmatrix} \tag{5.7}$$

where

$$A = \begin{pmatrix} 3a_1^2a_2^2-4a_1a_2^3+a_2^4 & -a_1^3a_2^2+2a_1^2a_2^3-a_1a_2^4 & a_1^4-4a_1^3a_2+3a_1^2a_2^2 & a_1^4a_2+2a_1^3a_2+3a_1^2a_2^2 \\ -6a_1^2a_2+6a_1a_2^2 & 2a_1^3a_2-3a_1^2a_2^2+a_2^4 & 6a_1^2a_2-6a_1a_2^2 & a_1^4-2a_1^2a_2^2-2a_1a_2^3 \\ 3a_1^2-3a_2^2 & -a_1^3+3a_1a_2^2-2a_2^3 & -3a_2^2+3a_2^2 & -2a_1^3+3a_2^2a_2-a_2^3 \\ -2a_1+2a_2 & a_1^2-2a_1a_2+a_2^2 & 2a_1-2a_2 & a_1^2-2a_1a_2+a_2^2 \end{pmatrix} \tag{5.8}$$

To test how well we do using this prescription we interpolate particles between two simulation outputs at $z_1 = 1.35$ and $z_3 = 0.92$ to a third output at $z_2 = 1.13$. The interpolated particles are then compared with their positions in the simulation itself. Figure 5.2 shows the positional accuracy in the Cartesian components of the particle positions. Most particles are accurately interpolated to within a few 100 h^{-1} kpc and the accuracy in practice will be higher as the time step used in this test is twice as large as that used in practice. This level of accuracy would be a problem for schemes where friends of friends codes are responsible for locating haloes as the typical orbital period of a halo will be much less than the typical snapshot separation of the simulation. In this case inaccurate interpolation will typically cause high velocity particles in the halo to leave the halo as it is hard to interpolate multiple circular orbits. In this case halos tend to "explode" between outputs leading to incorrect identification of structures. As we are not using particle positions to identify haloes this is no longer a problem and we can cope with the lower accuracy of interpolation. This in turn allows us to use a much coarser spacing of simulation outputs.

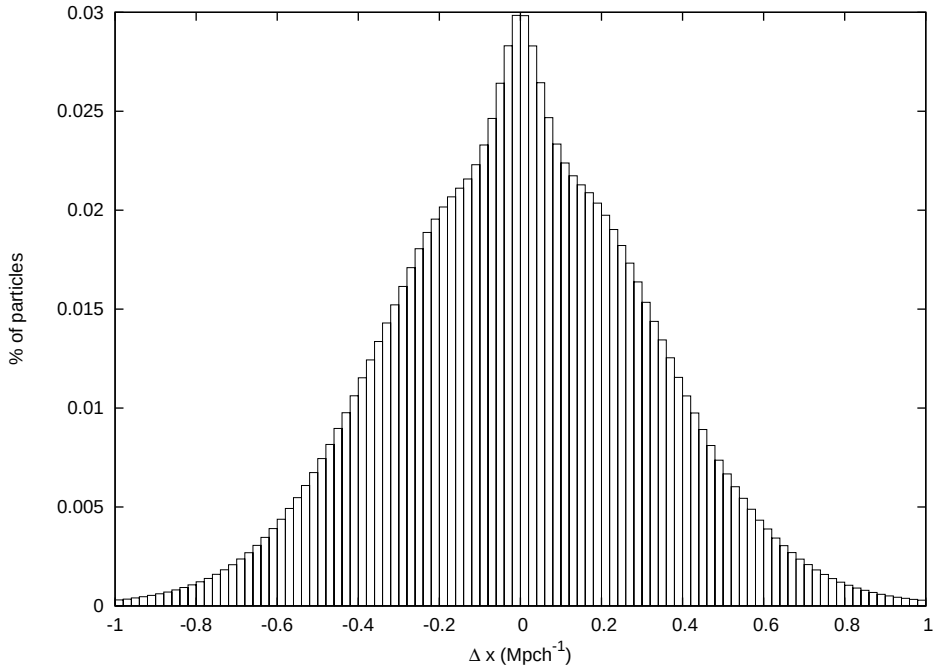

Figure 5.2: The accuracy of the interpolation algorithm described in Equation (5.7). The start and end redshifts where $z = 1.35$ and $z = 0.92$ interpolating to the intermediate redshift of $z = 1.13$. Most particles are accurately traced within a few 100 h^{-1} kpc. The actual accuracy in practice will be higher than this as the timestep of the simulations will actually be half that used in this test.

5.2.1 Populating cells

After applying the halo algorithm with evolution as discussed above we need to populate the haloes with galaxies. This is relatively straightforward as all we have to do is interpolate the conditional luminosity function to the redshift at which the halo sits relative to the observer and then apply the galaxy population algorithm of Chapter 3 as it stands.

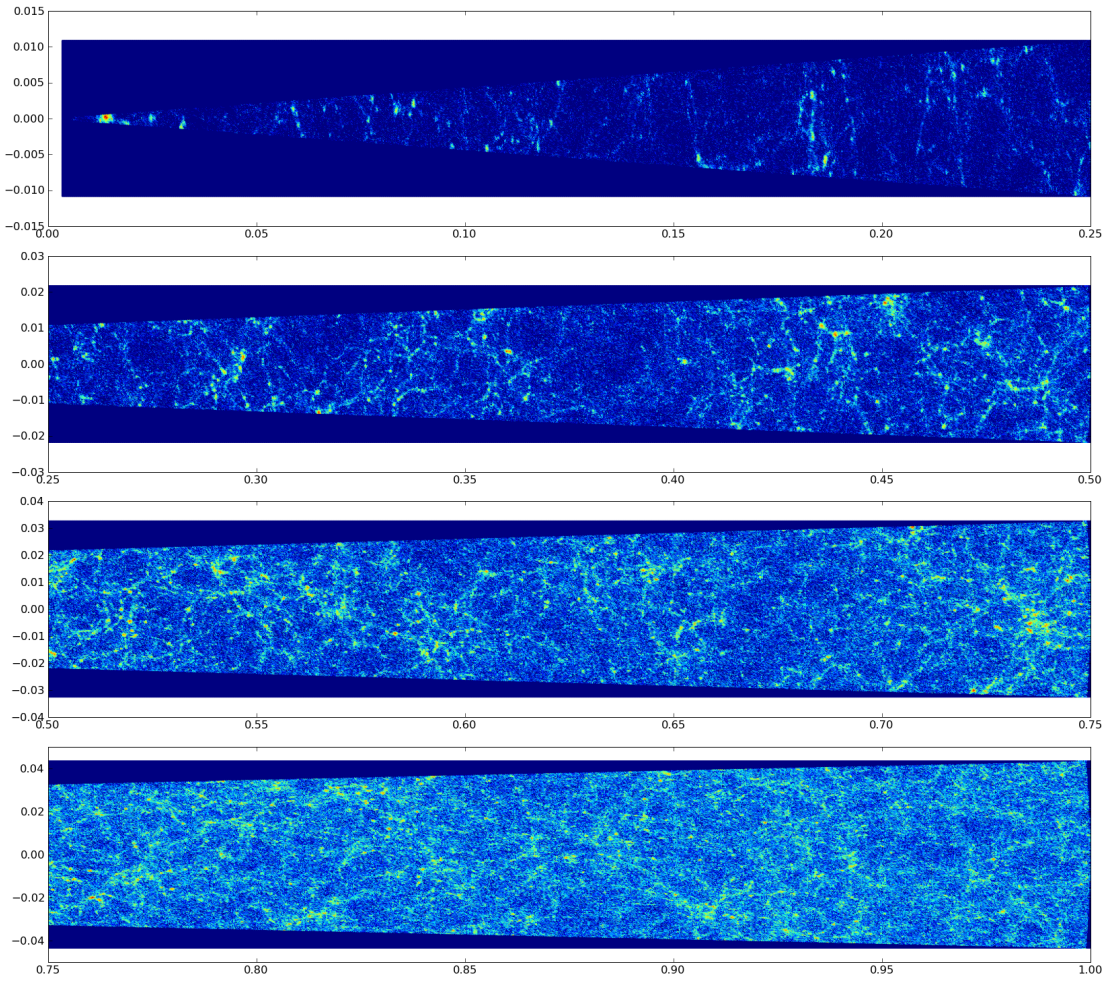

Figure 5.3: Logarithmic number density of galaxies in a lightcone of dimensions $5^\circ \times 5^\circ$ out to a redshift of $z = 1$. For each individual plot we rescale the y axis while the x axis shows the redshift. The lightcone is produced for an observer at the centre of a simulation box face. The orientation of the light cone is a random angle through the box. Any line of sight passes through at most only 3 replications of the simulation box.

Using the method we described above, we produce a lightcone with a similar geometry to the PAN-STARRS medium deep survey, a $5° \times 5°$ region of the sky out to a redshift of $z = 1$. Figure 5.3 shows the logarithmic number density of galaxies along the resulting lightcone. Each segment of the plot has had its y axis rescaled to fit the full lightcone in the plot. At low redshifts we see a number of small clusters, including a relatively large cluster next to the observer. As we increase redshift we observe a number of voids and the cosmic web begins to become more apparent.

5.3 Photometric redshifts and redshift space distortions

Once we have produced a catalogue of haloes on the lightcone we need to address the fact that when we measure the distance of a galaxy in a real survey we are not measuring the physical distance but rather its redshift. A further complication is that most large scale modern surveys will not determine a galaxy redshift directly but rather using a photometric redshift estimate. We need a prescription of how to take these effects into account within our light cone. Redshift space distortions are relatively easy to apply as each particle's redshift simply becomes a product of its cosmological redshift and the shift due to its peculiar velocity $\mathbf{v}$ projected along the line of site between the observer and the object $\mathbf{r}$:

$$z_{total} = (1 + z_{cos})(1 + z_{vel}) \tag{5.9}$$

where

$$1 + z_{vel} = \hat{\mathbf{r}} \cdot \mathbf{v}/c \tag{5.10}$$

There are two main methods for determining the redshifts of galaxies. A spectrum of the galaxy can be produced and then multiple features in the spectrum can be identified and the shift from their rest frame location measured. This method is very accurate but is time consuming and therefore limits the scope of surveys. An alternative approach is to use broadband filters in a number of bands to which a synthetic spectrum can then be fit to obtain an estimate of the redshift.

This introduces an error in the redshift which will in general depend on the particular algorithm that has been used to determine the redshift. It is much simpler however to bypass many of these complexities and simply model the probability of obtaining a photometric redshift $z_p(z)$ given the actual redshift z to be a Gaussian with a variance σ_z and a shift $z_b(z)$ in the mean.

$$p(z_p|z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{(z - z_p - z_b)^2}{2\sigma_z^2}\right] \tag{5.11}$$

The scatter introduced in the photometric redshift is due to inaccuracies within the method, but the shift is also important as the photometric redshift estimate is unbiased in z given z_p but not in z_p given z. In general it is hard for photometric redshift methods to achieve an accuracy greater than $\Delta z \simeq 0.05$, as the filters used in the photometry have a width of order 20% of their mean wavelength. While we will not explicitly use photometric redshifts in this work it is trivial to add them to our model by sampling the probability function in Equation 5.11.

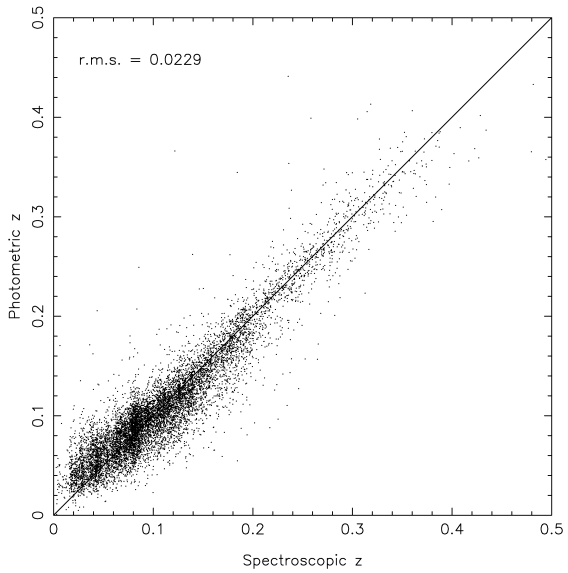

Figure 5.4: An example of the the error in photometric redshifts taken from Collister and Lahav (2004). In this work the photometric redshifts were obtained from an artificial neural network approach.

5.3.1 Redshift space distortions

The distortion of the correlation function along the line of sight by peculiar velocities is a problem when measuring the real clustering of the Universe, but it does contain information of its own. As structures form in the Universe, they have to acquire mass from their surrounding. This mass typically streams into the structure from the outside leading to high peculiar velocities. The rate of the growth of structure in the Universe can therefore be probed by measuring the shape of the redshift space distortions.

In linear theory, the growth of perturbations is related to the linear growth rate by:

$$f = \frac{d \ln \delta}{d \ln a}. \tag{5.12}$$

The growth rate depends strongly on the matter density $\Omega_m(a)$ at the epoch of interest (Heath, 1977; Linder, 2005). For many modified gravity models this dependence can be approximated by

$$f \approx \Omega_m^{\gamma} \tag{5.13}$$

where γ depends on the gravity model: $\gamma \approx 0.55$ for GR with a cosmological constant, $\gamma \approx 0.68$ for the DGP brane world model (Linder, 2005; Guzzo et al., 2008). In general γ is related to the effective equation of state of the model through $\gamma \approx 0.55 + 0.05[1 + w(z = 1)]$. If one can measure the growth of structure from redshift space distortions, then the distortions can become a good probe of alternative gravity models.

To do this we need a good theoretical model of the anisotropies created by the redshift space distortions. We write the correlation as a function of two variables $\xi(r_p, \pi)$, where r_p is the separation of a pair of galaxies perpendicular to the line of site while π is their separation parallel to the line of sight. Obviously the dependence of ξ on r_p remains the same, but there exist two main effects which can affect the line of sight correlations. On small scales, matter and galaxies located within dark matter haloes will have large line of sight virial velocities which will lead to galaxies traveling towards us appearing closer while those moving away appearing further. This effect known as the "fingers of god" has been found to be well modeled by a convolution of the correlation function with a Gaussian with

a typical 1D pairwise velocity dispersion of σ_{12},

$$\phi(v) = (\sigma_{12}\sqrt{2})^{-1} \exp \frac{\sqrt{2}|v|}{\sigma_{12}} \tag{5.14}$$

The second effect is an enhancement of the large scale correlations from the infall of matter into the cluster. In this case matter further than us from the halo centre will be falling towards the observer along the line of sight resulting in an underestimation of the matter's true position. Conversely, matter closer to us than the cluster will be falling away from the observer resulting in an overestimation of its true position. The effect then of infall is to make clusters appear more compact (Kaiser (1987)). To model this effect we follow Hamilton (1992) in expanding the correlation function in spherical harmonics

$$\xi(r_p, \pi) = \xi_0 P_0(\mu) + \xi_2 P_2(\mu) + \xi_4 P_4(\mu), \tag{5.15}$$

where μ is the cosine of the angle between the separation vector and the line of sight vector, $\mu = \mathbf{r} \cdot \mathbf{z}$, and the P_i are the Legendre polynomials and ξ_i are the weights:

$$\xi_0 = (1 + \frac{2}{3}f + \frac{1}{5}f^2\xi(r)) \tag{5.16}$$

$$\xi_2 = (\frac{4}{3}f + \frac{4}{7}f^2)[\xi(r) - \bar{\xi}(r)] \tag{5.17}$$

$$\xi_4 = \frac{8}{35}f^2[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)], \tag{5.18}$$

where

$$\bar{\xi} = 3r^{-3} \int_0^r \xi(s)s^2 ds, \tag{5.19}$$

and

$$\bar{\bar{\xi}} = 5r^{-5} \int_0^r \xi(s)s^4 ds. \tag{5.20}$$

For a power law correlation function $\xi \propto r^{-n}$ this gives:

$$\xi_0(r) = \left(1 + \frac{2}{3}f + \frac{1}{5}f^2\right)\xi(r) \tag{5.21}$$

$$\xi_2(r) = -\left(\frac{4}{3}f + \frac{4}{7}f^2\right)\frac{n}{3-n}\xi(r) \tag{5.22}$$

$$\xi_4(r) = \frac{8}{35}f^2\frac{n(2+n)}{(3-n)(5-n)}\xi(r) \tag{5.23}$$

To complete the model we convolve the result with the line of sight velocity distribution as given in Equation (5.14).

5.3.2 Fitting the redshift space distortions

In practice we want to measure the linear redshift distortion parameter

$$\beta = \frac{f}{b_l}, \tag{5.24}$$

where b_l is the linear bias value defined as the ratio between the root mean squared density contrasts for the galaxy and dark matter distributions. The other free parameter that exists is the pairwise velocity dispersion, which has to be fit from the data. The strategy we adopt is to fit the prescription for modelling the redshift space distortions described in the previous section to that measured in our simulations to determine the values of σ_{12} and β. Here to increase our number statistics and for ease of computation we calculate the correlation function for an individual simulation output, taking the z axis as our line of sight. At the end of this chapter we will contrast this wide angle approximation with a full result from the lightcone. In our modelling we have an additional two parameters, r_0, the normalisation, and n, the slope of the real space correlation function, which need to be determined. In real surveys (Guzzo et al., 2008) these parameters are determined by fitting by projecting $\xi(r_p, \pi)$ along the line of sight and inverting to obtain $\xi_r(y)$:

$$w_p = 2\int_0^\infty \xi(r_p, \pi)d\pi = 2\int_{r_p}^\infty \frac{y\xi_r(y)}{y^2 - r_p^2}dy \tag{5.25}$$

Here we will simply determine r_0 and n by fitting the spherically averaged, real space correlation function $\xi(r)$ in the range $r = 1\,h^{-1}\,\mathrm{Mpc}$ to $r = 10\,h^{-1}\,\mathrm{Mpc}$. We then proceed to fit the remaining two parameters β and σ_{12} using the 2D correlation function. We define our goodness of fit as:

$$\chi^2 = \sum_i \sum_j \left(\frac{y_{ij}^{model} - y^{sim}}{\sigma_{ij}} \right)^2 \tag{5.26}$$

and then calculate χ^2 over the ranges $0\,h^{-1}\,\mathrm{Mpc} < \pi < 20\,h^{-1}\,\mathrm{Mpc}$ and $0\,h^{-1}\,\mathrm{Mpc} < r_p < 20\,h^{-1}\,\mathrm{Mpc}$. We do this for 100 values in the range

$0.4 < \beta < 0.7$ and $300\mathrm{km\,s^{-1}} < \sigma_{12} < 700\mathrm{km\,s^{-1}}$ to obtain the best fitting values. Table 5.1 shows the results of the fitting of all 4 parameters for red galaxies, blue galaxies and all galaxies and Figures 5.5, 5.6 and 5.7 show the measured correlation function and the best fit model for all, blue and red galaxies, respectively.

Galaxy Type	r_0	n	β	σ_{12} (kms^{-1})
red	6.49 ± 0.8	2.56 ± 0.67	0.43 ± 0.23	543.32 ± 102.3
blue	3.95 ± 0.75	1.36 ± 0.52	0.42 ± 0.16	657.85 ± 94.5
Total	5.2 ± 1.2	2.0 ± 0.79	0.69 ± 0.34	405.34 ± 110.2

Table 5.1: Best fit parameters for redshift space distortions. The errors are calculates from 4 different runs of the algorithm

We compare these results with the results from the 2dFGRS (Madgwick et al., 2003) which can be found in figure 5.8.

Qualitatively we find similar results, the red passive galaxies have a much higher central dispersion than the blue galaxies from having a high proportion of satellite galaxies in massive haloes. The blue galaxies tend to live in smaller haloes and so have a much less pronounced finger of god effect. The 2dFGRS survey measured β to be 0.49 for active galaxies and 0.48 for passive galaxies. Our best fitting value of β for blue galaxies is 0.43 while for red galaxies prefer a value of 0.42. Our values for β are consistently low when compared with the 2sFGRS results while our values of σ_{12} are much higher. Part of this may be because we are calculating the result in a wide angle approximation, where we simply choose the z axis in a simulation to be the line of sight, and boost each particle in this direction. Figure 5.9 shows the equivalent redshift space correlation function for all galaxies calculated on the lightcone. While we have not fit the redshift space distortions to it, it has a greater resemblance to the results of Madgwick et al. (2003).

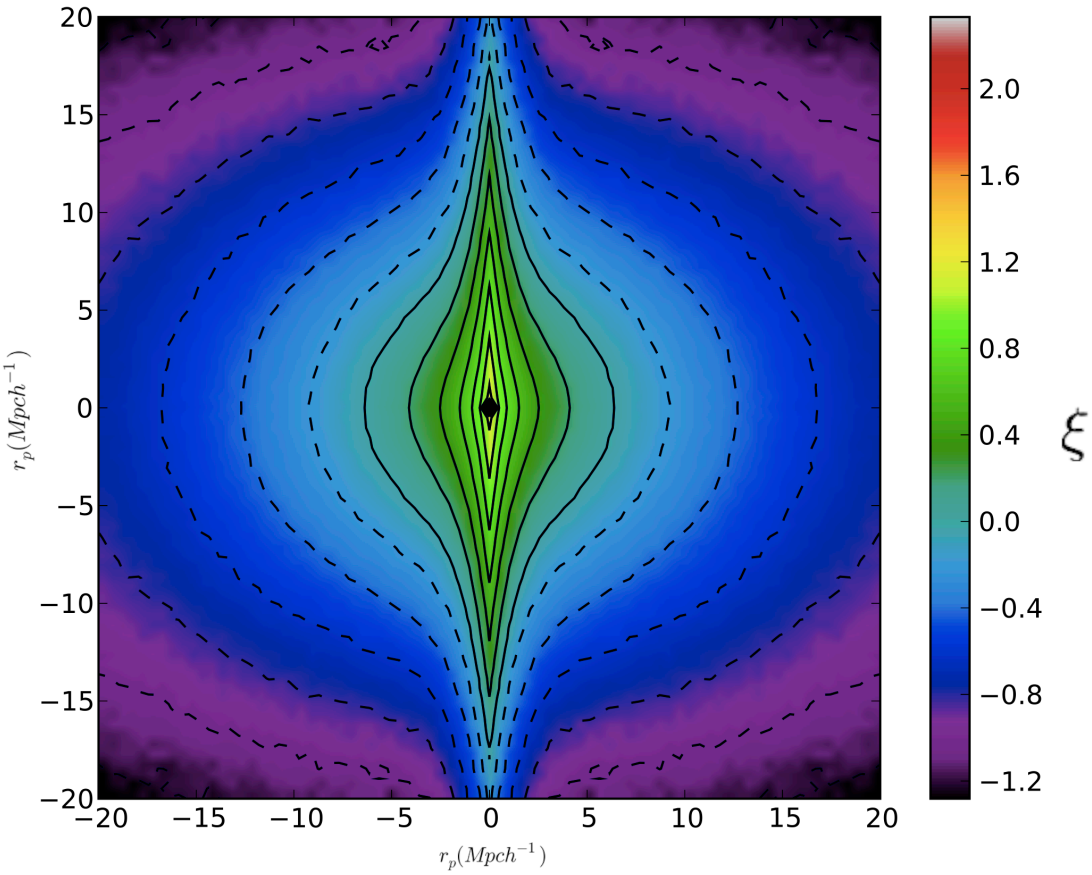

Figure 5.5: The 2D correlation function $\xi(r_p, \pi)$ of all galaxies for an individual simulation output. The contours show the best fit model to the data.

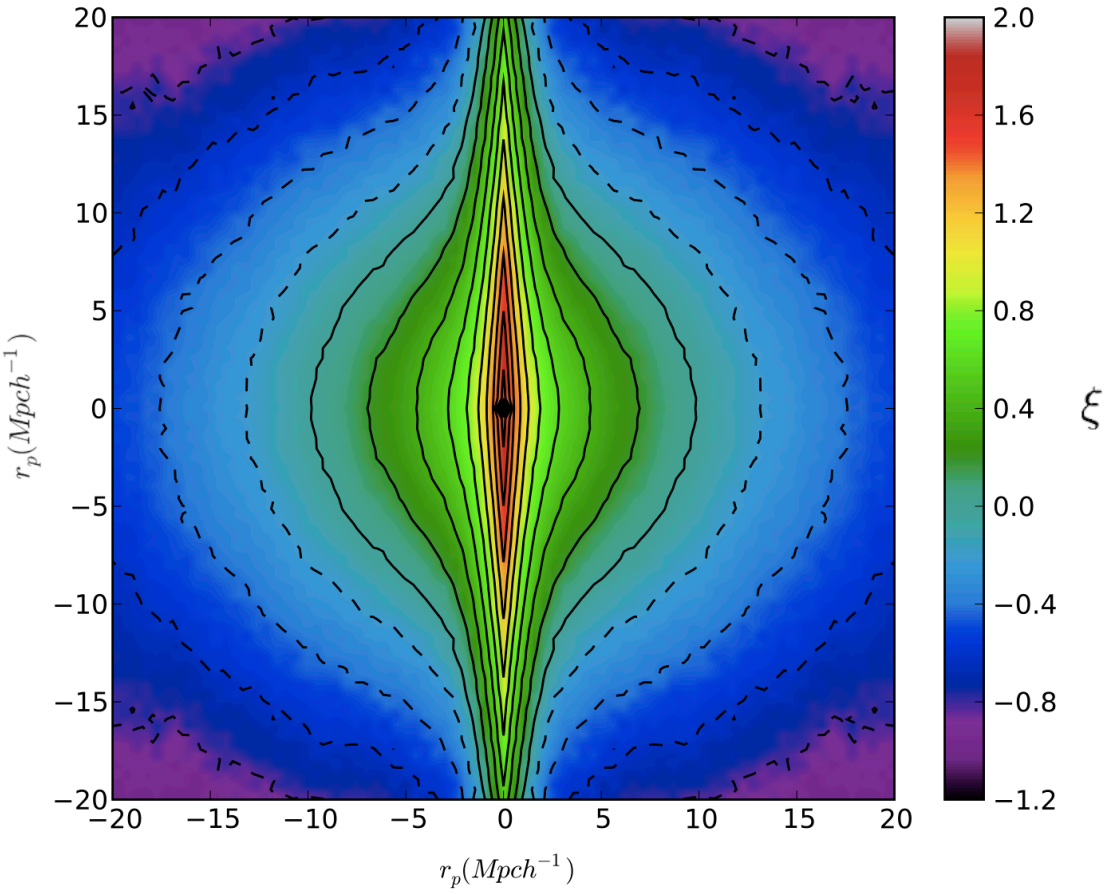

Figure 5.6: The 2D correlation function $\xi(r_p, \pi)$ of red galaxies for an individual simulation output. The contours show the best fit model to the data.

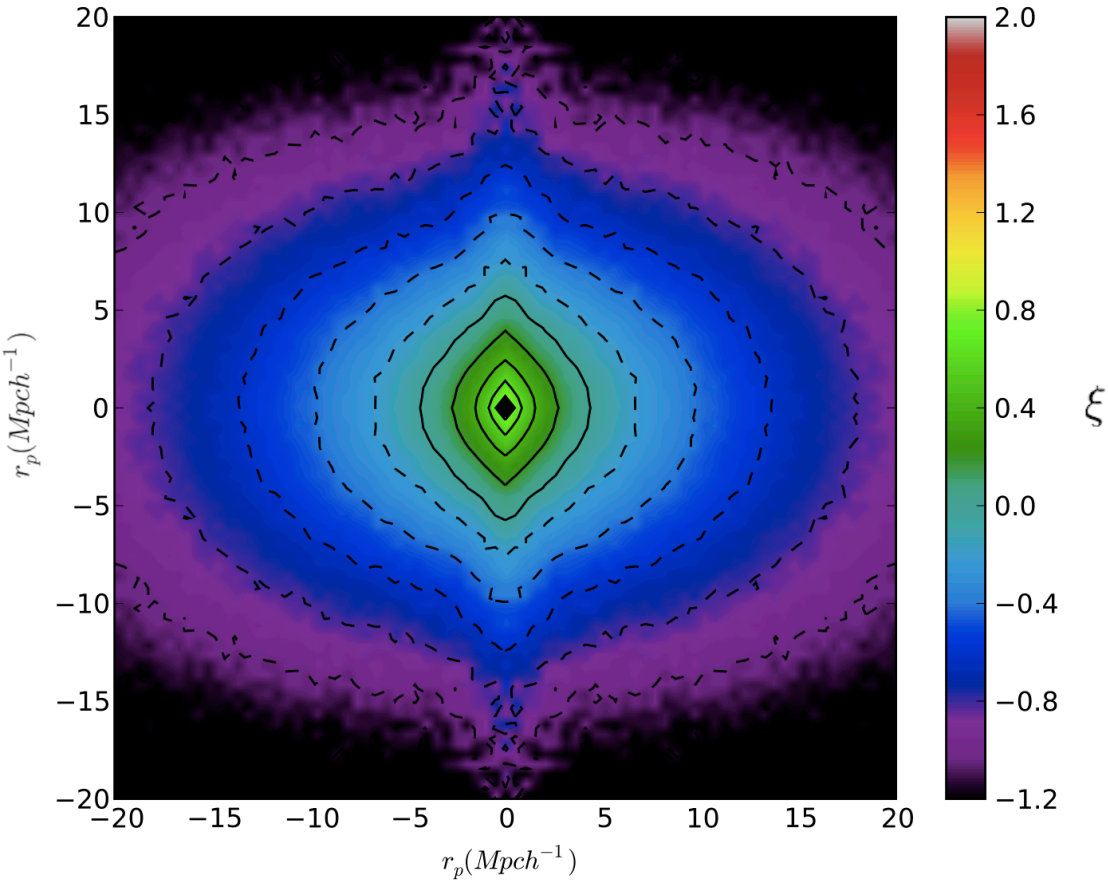

Figure 5.7: The 2D correlation function $\xi(r_p, \pi)$ of blue galaxies for an individual simulation output. The contours show the best fit model to the data.

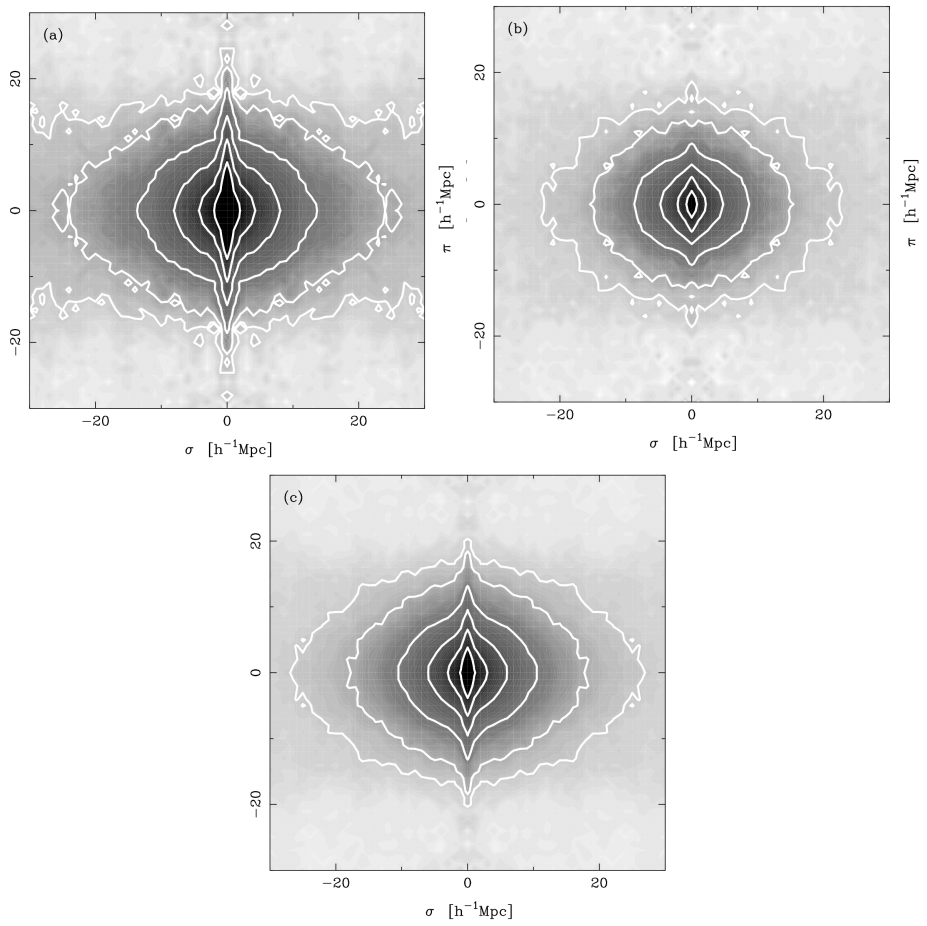

Figure 5.8: The 2D correlation function $\xi(r_p, \pi)$ of a) passive b) active c) all galaxies measured in the 2dfGRS survey (Madgwick et al., 2003)

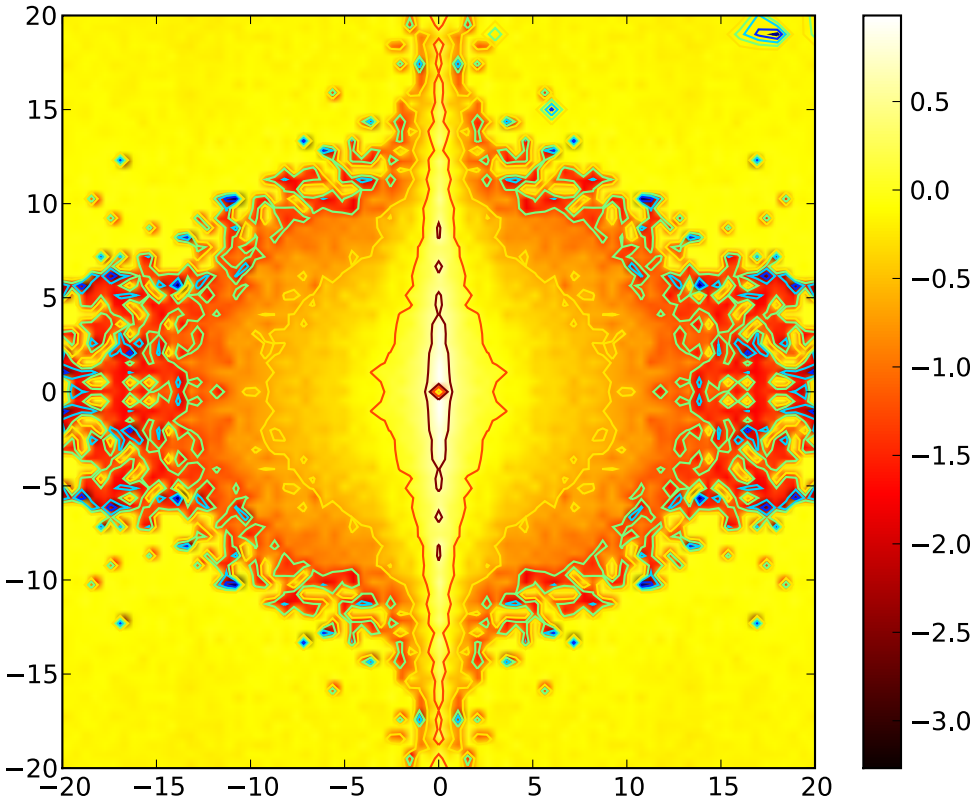

Figure 5.9: The 2D correlation function $\xi(r_p, \pi)$ of all galaxies calculated on the lightcone.

5.4 Topology

While the two point correlation function is one of the most natural ways to measure the growth of structure in the Universe, it is not the only statistic which can be used. The topology of the density has also been explored to a lesser extent. The idea is to smooth the density field on a given scale R and then calculate the genus, a measure of the connectivity of the resulting iso-density surface. For a given surface in 3 dimensions the topology can take 3 distinct forms, "meat ball" like which consists of a number of isolated high density regions, "Swiss cheese" like which have a number of isolated holes or a "sponge" like topology which consists of a combination of meatball and hole like configurations. The characteristic which separates the "sponge" topology form the others is that it is completely connected. Within both the regions contained in the isosurface and those excluded by it, it is possible to move to any other region of the same class. The interest of applying these ideas to cosmological density fields comes from the fact that a Gaussian field should have a "sponge" like topology, and so any departure from this is a measure of non-Gaussianity of the density field. For more details see Gott et al. (1986).

There are two ways in which the density field could be non-Gaussian, either a primordial non-Gaussianity signal was always present in the density field, imprinted along with the density perturbations from inflation, or through the non-linear growth of structure through gravity in the late Universe.

The topology of a field is quantified by its genus: a measure of the number of connected holes in a surface. For example, a sphere would have a genus of zero while a mug would have a genus of 1 etc. The genus of a Gaussian random field is (Gott et al., 1986):

$$g = N(1 - \nu^2)\exp(-\nu^2/2), \tag{5.27}$$

where ν is a measure of the mean density contour in units of the standard deviation away from the mean density and N, the normalisation factor, depends on the power spectrum of the field. Therefore the curve displayed in Figure 5.10 is sensitive to the power spectrum of the density field but it only effects its overall amplitude.

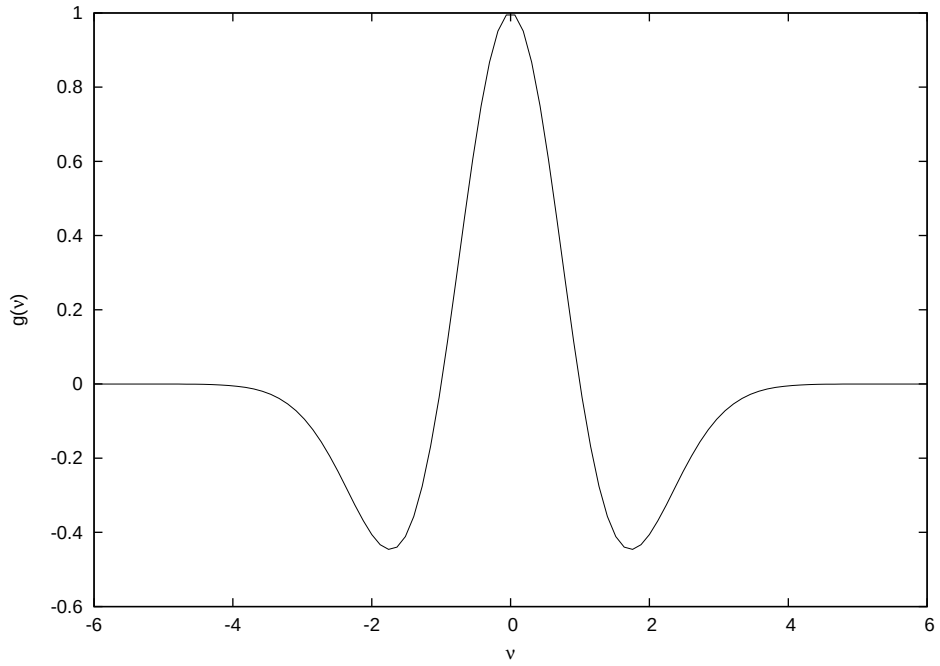

Figure 5.10: The genius function of a Gaussian random field.

5.4.1 Measuring the genus in simulations

To apply the above analysis to simulations we require a way of relating the global topological property, the genus, to a locally measurable geometric quantity. This is achieved through the Gauss-Bonnet theorem which relates the curvature of a surface to its genus. First we define the Gaussian curvature at a point x on the isodensity surface to be

$$K(\mathbf{x}) = \frac{1}{r_1(x)r_2(x)}, \tag{5.28}$$

where r_1 and r_2 are the radii of curvature for the two principal components on the surface. The Gauss-Bonnet theorem then tells us that the genus is proportional to the integral of $K(x)$ over the surface

$$4\pi(1-g) = \int K(\mathbf{x})dA. \tag{5.29}$$

For example, a sphere has a constant radius of curvature $1/R^2$ where R is the radius of the sphere and so the integral of curvature is simply 4π giving a genus of 0, as expected. To apply the Gauss-Bonnet theorem to a simulation volume we first have to define the iso-density surface for a given density. This is done by first binning up the simulation particles to produce a 3D gridded density

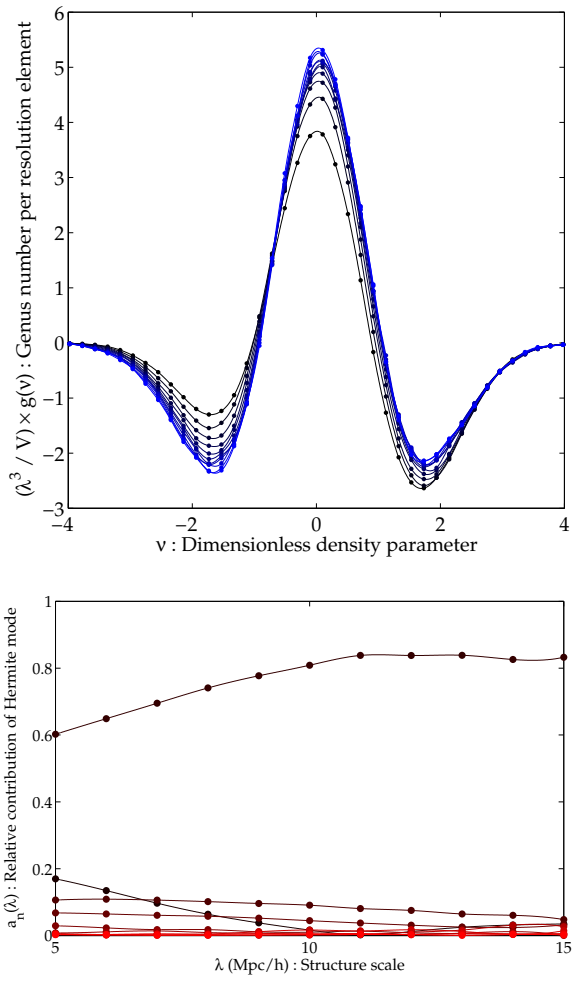

Figure 5.11: The genus of the halo population within the simulation (top). Each line represents the genus measured after smoothing on a given scale starting with $\lambda = 5\,h^{-1}$ Mpc (black curves)through to $\lambda = 15\,h^{-1}$ Mpc (blue curves). The bottom diagram shows the result of Hermite decomposition of the genus curve and shows the fractional contribution of the first ten modes to each smoothing scale.

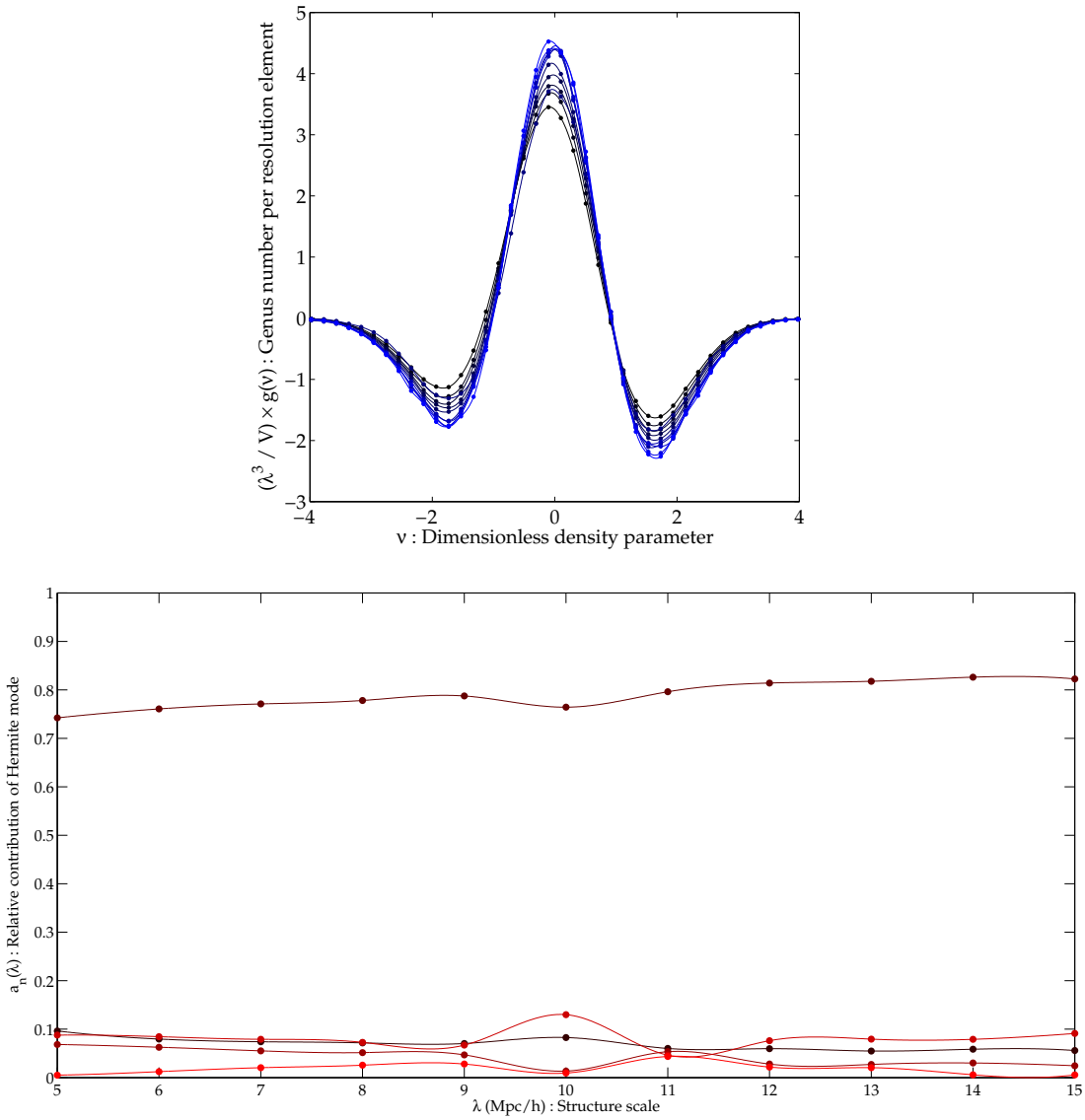

Figure 5.12: As for Figure 5.11 but for red galaxies.

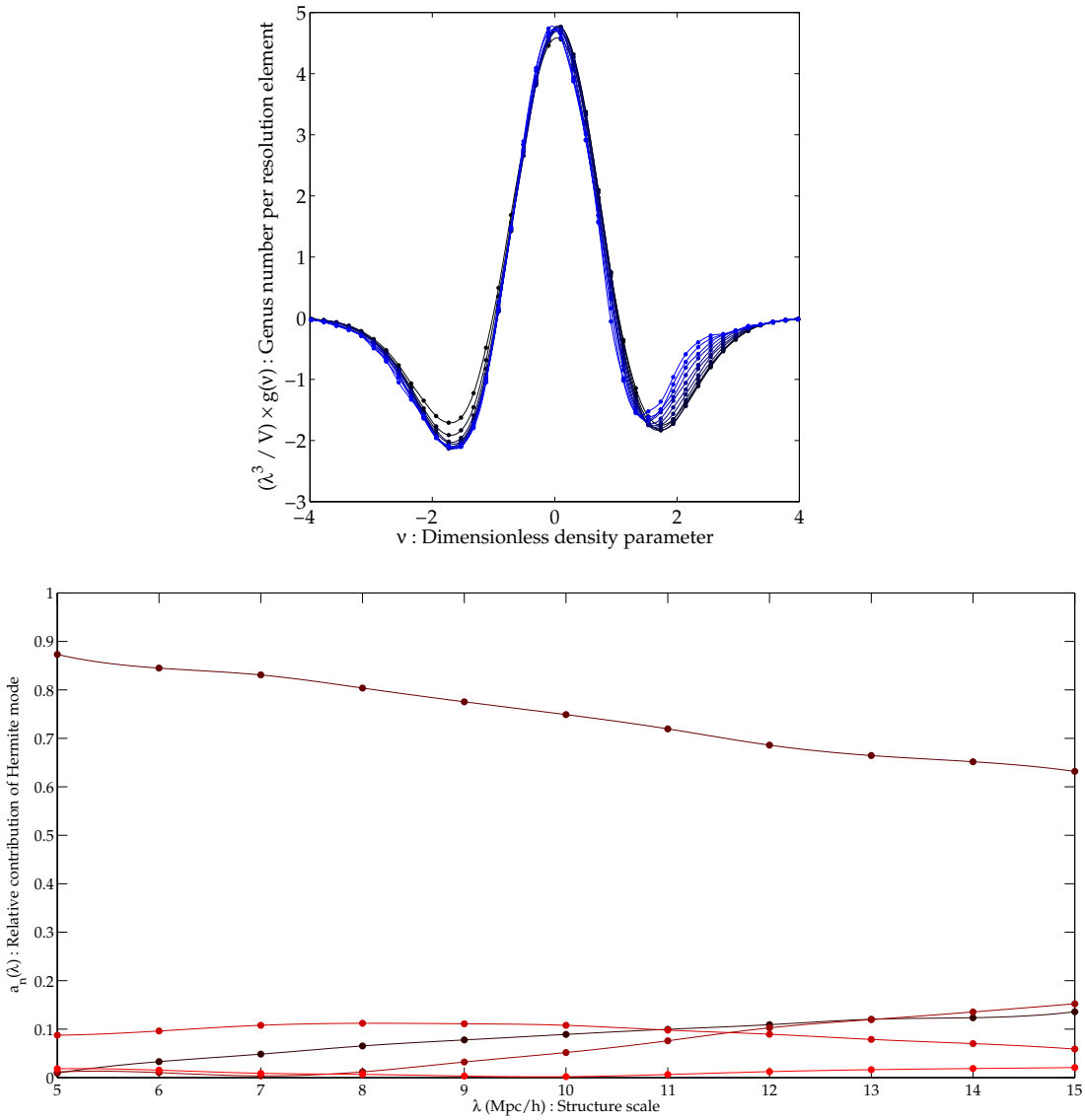

Figure 5.13: As for Figure 5.11 but for blue galaxies.

field, smoothing on a given scale λ using a Gaussian smoothing kernel and then classifying each cell as either greater than the density threshold ν (in which case it is inside the surface) or less than the density threshold ν (in which case it is outside the surface). The boundary between these surfaces is then approximated by a number of polygons which are in turn used to calculate the radius of curvature over the surface. For more details see Gott et al. (1986), Weinberg (1988) and Gott et al. (1987).

Once the genus has been measured, one method for characterising the departures from the Gaussian case is to decompose the curve in terms of Hermite functions via:

$$a_n(\lambda) = \int g_\lambda(\nu) e^{-v^2/4} H_n(\nu) d\nu, \tag{5.30}$$

where H_n is the n^{th} Hermite function. In a series of papers, (Matsubara, 1994; Matsubara and Suto, 1996) showed that the correlations between phases introduced by non-linear gravitational evolution lead to disruption of the symmetry about the mean density contour of the genus curve. Using cosmological perturbation theory they demonstrated that in a Gaussian field all connected moments higher than n=2 vanish, while in a non-Gaussian field the genus curve can be expressed as

$$g_{WNL} \approx -A \exp(-\nu^2/2)[H_2(\nu) + \sigma(P_3 H_3(\nu) + P_1 H_1(\nu))], \tag{5.31}$$

where $H_n(\nu)$ is the nth polynomial, P_n are their coefficients and σ is the rms fluctuations of the field.

We apply the genus measurement to each of the three populations in one of our redshift $z = 0$ simulation and decompose each into Hermite polynomials. This is done for a number of smoothing scales to establish the evolution of the genus with scale, starting from mildly non-linear at $15\,h^{-1}\,$Mpc through to the highly non-linear $5\,h^{-1}\,$Mpc scale. This analysis was carried out by Dr Berian James using data supplied by the author. Figure 5.11 shows the genus of the halo population and the fractional contribution of the nth Hermite polynomial to the genus. As we would expect, the halos exhibit a higher clumping than what would be expected from a Gaussian field. It is interesting to note that non-Gaussian signals are still present even at the weakly non-linear smoothing

scale. The galaxy results for red galaxies, Figure 5.12, and blue galaxies, Figure 5.13 showing very different characteristics. The red galaxies display an evolution with mean density similar to that of the dark matter haloes. In some respects this is unsurprising: the red galaxies will tend to be central galaxies associated with halos centres and so we would expect the results to be much the same. The blue galaxy evolution is very different, however, with the value of the genus curve for median density being constant with scale, indicating that the filament structures associated with the median density seem unaffected by non-linearity. The main deviation from the Gaussian field result comes at higher density scales. The decomposition into Hermite polynomials shows more significant differences in the different galaxy populations. Red galaxies have a similar decomposition to that of the dark matter halos but with a weaker decrease of the $n = 2$ mode. There appears to be a 'bump' in the results around the $10\,h^{-1}\,\mathrm{Mpc}$ scale which is likely to be an artefact of the decomposition. In comparison the blue galaxy decomposition shows an increase in the $n = 2$ mode at small scales. This is not consistent with the usual signature of non-linear gravitational evolution.

While these results are interesting they are preliminary in nature and it is unclear to what extent systematics play a role. We also only have run these results for a single simulation and so have no idea of the errors on the a_n values. This will be part of the aim of future work. Regardless of this it is striking that there exists such a difference between red and blue galaxies and this could indicate a strong dependence on environment. The preference of red galaxies to populate high mass halos in clusters and their increased likelihood of being a central galaxy, while blue galaxies tend to be satellites and field galaxies may contribute to the distinct difference in topology results. The dependence of galaxy colour on topology has been an unexplored area and this result suggests it is worthy of more attention.

5.5 Conclusions

In this Chapter we have extended the use of our algorithm to produce lightcones. We think our approach gives as a great advantage when it comes to producing light-cones as the galaxy and halo statistics along with the density field evolve smoothly along the history of the lightcone. This removes the need for many

of the methods required to ensure continuity of structure when constructing a lightcone from N-Body simulations and semi-analytics. While we are happy with the results, we are also keen to improve the scope of the lightcones we can produce. This is simply a matter of increasing the simulation size we are using as a seed and promises to produce full sky mocks with relative ease.

We have also had an initial examination of some of the other statistics used to characterise structure formation: topology and redshift space distortions. As with the BAO results in the last chapter we are more interested in the proof of concept here rather than a rigorous evaluation of these numbers and it is encouraging that our method is able to produce the required details.

The topological results is however interesting as it gives a distinctly different result for red and blue galaxies. This indicates they inhabit topologically different structures. These differences if interpreted correctly could give fresh insight in to the structure formation process.

Similarly the redshift space distortion results are, while a little off what has been measured from galaxy redshift surveys, encouraging. If we can refine the method to further match what is seen in the real Universe this will become a valuable tool. One exciting prospect which we will discuss more at the end of this thesis, is the idea that using this method we can produce redshift space distortions for other theories of gravity. By changing the underlying simulation which produces the density field and suitably altering the galaxy/halo statistics, we can quickly cover a number of modified gravities and produce predications on the distortion signature we would expect to find in each.

Chapter 6

Applications to Lensing

While the clustering of objects in the Universe allows us to trace the dark matter in an indirect way, gravitational lensing gives us a direct probe. As light travels to us from distant galaxies it has to pass through the cosmic web of large scale structure. The gravitational field created by the inhomogeneities in the Universe deflects light by distorting geodesics. In regions of large mass density like clusters the effect of the gravitational field is easily seen, distorting the background image of stars into large arcs. In less dense regions the distortion is more subtle but still present. As the level of deflection is proportional simply to the mass present within a region of space and not the nature of the mass, it can be used to map not only regular matter but dark matter. The method does not distinguish between classes of matter, but as can be seen in Figure 6.1 the mass on large scales is dominated by dark matter.

In regions of low density the effect is seen as subtle changes at around the 1% level to the ellipticity of galaxies, which is not detectable on a galaxy by galaxy basis but can be measured statistically by averaging the observed ellipticity of galaxies over an area of the sky. This regime, where the distortion is small, is known as weak lensing. Weak lensing is a promising method for not only mapping the dark matter distribution in both projection and in full 3D distribution (Simon et al., 2009), but also as a method of determining the nature of dark energy and

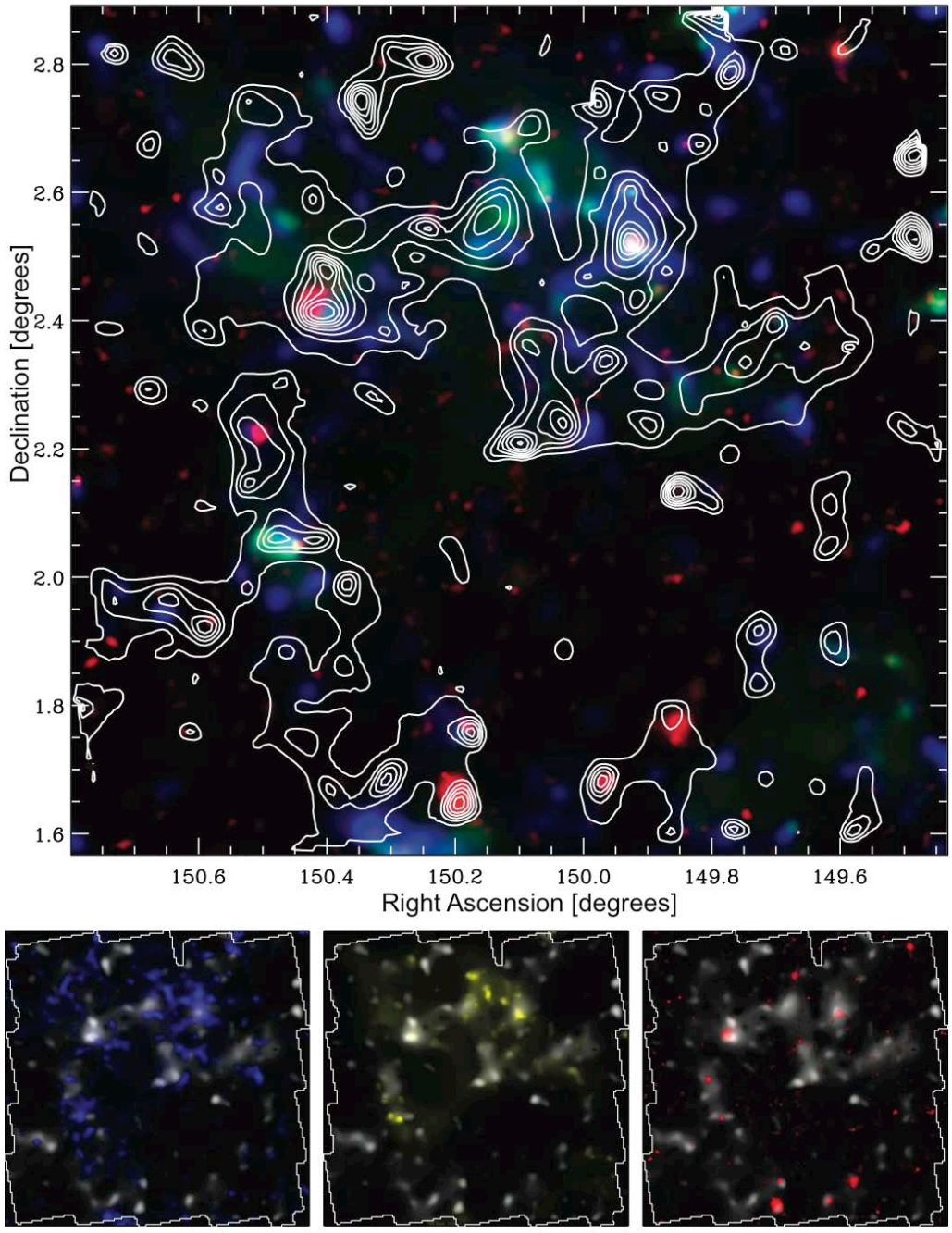

Figure 6.1: Dark matter map of the COSMOS field from Massey et al. (2007b). Top image shows the contours of the recovered mass distribution from lensing while the blue yellow and red indicate three tracers of baryonic matter: stellar mass, galaxy number count and hot x-ray gas, respectively.

can potentially provide signals of departures from General Relativity. However the method has a number of potential systematic errors which may adversely affect the results. In this chapter we will discuss the basic methods of gravitational lensing before discussing how the algorithm used to produce lightcones in the previous chapter can be extended to explore weak gravitational lensing.

6.1 Lensing basics

In general relativity light travels along geodesics, straight lines in the distorted space time. If we consider a thin lens, one in which the mass producing the effect is confined to a small region perpendicular to the line of sight, and introduce a lensing potential ψ, then the bend angle α is related to this potential through

$$\alpha = \nabla_\theta \psi(\theta_\mathbf{1}) \tag{6.1}$$

and from the Poisson equation the lensing potential is related to the integrated surface density $\Sigma = \int \rho dl$ and the critical surface density Σ_c by:

$$\nabla^2_\theta \psi = \frac{D_l D_{ls}}{D_s} \frac{8\pi G}{c^2} \Sigma = 2 \frac{\Sigma}{\Sigma_c}. \tag{6.2}$$

The bend angle α relates the position of an object on the source plane θ and the image plane β as

$$\beta = \theta - \alpha(\theta) \frac{D_{ls}}{D_s}. \tag{6.3}$$

In weak lensing we will not in general be able to use a thin lens approximation as the matter distribution lies continuously along the light of sight. In this case the convergence field can be constructed as a weighted sum of many individual thin lens approximation. We slice our simulation up into a series of j onion layers each of thickness dr_j and compute for a given pixel position i, the sum

$$\kappa(i) = \frac{3H_0^2 \Omega_m}{2c^2} \sum_j \delta(i,j) \frac{(r_s - r_j) r_j}{r_s a_j} dr_j \tag{6.4}$$

over simulation slices j which are each at a distance of r_j from the observer and thickness dr_j. The convergence is the central quantity for lensing from which all other quantities can be derived. The most important of these when it comes to weak lensing is the shear, which describes the deformation of circular background

images by the intervening matter. The shear is decomposed into two orthogonal directions giving γ_1 and γ_2 which are related to the lensing potential by

$$\kappa = (\psi_{11} + \psi_{22})/2 \tag{6.5}$$

$$\gamma_1 = (\psi_{11} - \psi_{12})/2 \tag{6.6}$$

$$\gamma_2 = \psi_{12}. \tag{6.7}$$

6.1.1 Convergence power spectrum

In much the same way that the density field is the key quantity for clustering and structure formation, convergence is the key quantity in lensing. The power spectrum of convergence is therefore one of the most important quantities in lensing. The theoretical form of the power spectrum can be obtained through the application of Limber's equation (Limber, 1953) to the matter power spectrum with a suitable weighting. We are essentially trying to observe a line integral through the density field weighted by a function $W(y)$,

$$\delta(\hat{\theta}) = \int_0^\infty \delta(\mathbf{y}) W(y) dy, \tag{6.8}$$

which translates, using the Fourier version of Limber's equation, to the dimensionless angular power spectrum (assuming a flat cosmological model)

$$\Delta^2_K = \frac{\pi}{K} \frac{9\Omega_m^2}{4} \left(\frac{c}{H_0}\right)^{-4} \int \Delta^2(K/y, y) G^2(y) y^3 dy, \tag{6.9}$$

where y is the comoving radius, y_s is the comoving radius of the source and $G^2 = a^{-1}(y_s - y)/y_s$. To compute this we have to provide a matter power spectrum $\Delta^2(y)$, we want to include some of the non-linear effects of structure growth, so we use the halofit method (Smith et al., 2003) to generate the matter power spectrum.

6.1.2 Halo surface densities

We can think of the lensing potential as being produced by the sum of individual contributions from many individual particles. When constructing the lensing potential from a simulation, this is exactly what we would do: bin the particles

onto a regular grid and then Fourier transform this grid to obtain the lensing potential. In general this requires a high resolution simulation to resolve small scale structures. If, however, we construct a surface density field from a catalogue as we produced in the previous chapter, we can apply the appropriate halo density profile around each halo centre and thus restore the small scale details to a relatively coarse simulation.

In principle any halo profile can be chosen to dress the halo centres. Here we list the surface density profiles of NFW and SIS halo profiles and in general we will stick to using the NFW profile. The surface density of a given profile is obtained simply by integrating the density profile along the line of sight:

$$\Sigma_h = \int \rho(r_\perp + r_s) dr_s, \tag{6.10}$$

where $r_\perp = D_d\sqrt{(\theta_1^2 + \theta_2^2)}$ is the distance from the halo centre perpendicular to the line of sight (with θ_1 and θ_2 being the measurement away from the halo centre in the two components of angular coordinates on the sky) and r_s is the separation from the halo centre parallel to the line of sight. Following Wright and Brainerd (2000) we define a dimensionless radial distance $x = r_\perp/r_s$ and integrate the NFW profile to obtain:

$$\Sigma_{NFW}(x) = \begin{cases} \frac{2r_s\delta_c\rho_c}{(x^2-1)}\left[1 - \frac{2}{\sqrt{1-x^2}}\tanh^{-1}\left(\sqrt{\frac{1-x}{1+x}}\right)\right] & x<1 \\ \frac{2r_s\delta_c\rho_c}{3} & x=1 \\ \frac{2r_s\delta_c\rho_c}{(x^2-1)}\left[1 - \frac{2}{\sqrt{x^2-1}}\tan^{-1}\left(\sqrt{\frac{x-1}{1+x}}\right)\right] & x>1 \end{cases} \tag{6.11}$$

which is displayed for a given halo in Figure 6.2.

Our task is then to combine this profile with haloes contained in the light cone catalogues we produced in Chapter 5. To produce a gridded surface density for a number of redshift slices. There are a number of different approaches we could take to accomplish this. Perhaps the most efficient approach would be to bin all haloes in a small mass range of $M_1 < M_h < M_2$ onto the surface density grid. The true density field would then be a convolution of the NFW profile appropriate for a mass of $M_{mean} = (M_1 + M_2)/2$ and the density field. The convolution can be performed in Fourier space, the process repeated for multiple mass ranges and the surface density summed to give the true surface mass density. This would be

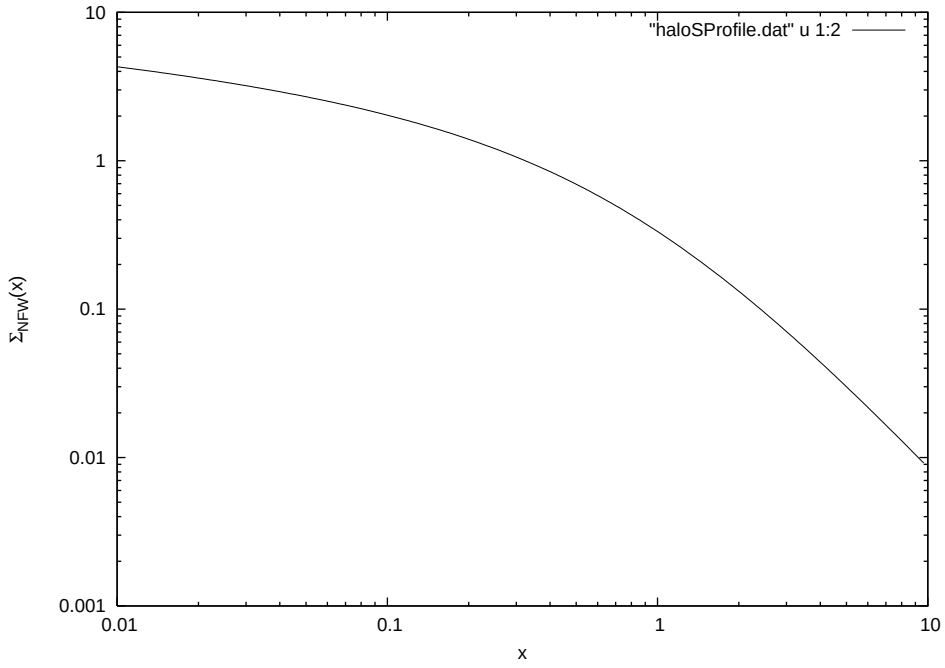

Figure 6.2: A scaled surface density of an NFW profile. x here is a dimensionless radial parameter $x = r_\perp / r_s$.

highly efficient given the existence of fast Fourier transform routines. However, smaller haloes and larger haloes at high redshift will typically only span one or a few grid points. In this case the above method becomes superfluous and the surface density field can be more easily obtained by simply binning haloes as if they are point particles of mass M_h. This will obviously not work when we have to deal with halos which span more than one grid point, so we need an extra strategy to deal with them. The approach we take is as follows:

For each halo in the lightcone:

- given the halo mass, concentration, redshift and location on the sky, project the halo to the surface density grid and determine the nearest grid point (θ_1, θ_2).
- Compute the number of cells on the surface density grid it subtends as $N_c = r_{200}/D_a/\triangle\theta$ along each dimension, where $\triangle\theta$ is the separation of grid points, D_a is the angular diameter distance.
- If $N_c <= 1$ simply assign the halo mass to the surrounding cells using a cloud in cell algorithm.
- if $N_c > 1$ Divide the cell into n subcells.
- Generate a NFW surface profile for the halo with appropriate mass, redshift and concentration parameters.
- At each subcell centre, sample the NFW surface profile and multiply by the cell area to obtain the contribution of the halo's mass associated with that subcell.
- Sum the contributions of each subcell and assign the resulting mass to the current cell.
- After all summing has been computed divide by the cell volume and the mean density of the Universe at the slice redshift to obtain an over-density field.

.

This ensures that the extended nature of large/close haloes are accounted for while more distant/smaller haloes can be quickly assimilated into the density field. A complication with both of the methods described above is that the spacing of the radial shells must be greater than the diameter of largest haloes. If this is not true there is a high probability that the projected mass of a halo will be assigned to an individual redshift slice rather than multiple ones. This could be overcome by replacing the 2D sampling of the halo surface density profile with a 3D sampling and then binning the resulting particles onto multiple slices. In practice however this should not be a concern as the typical redshift range over which evolution in the density field will occur, or over which the weak lensing properties will be of interest, will typically be much smaller than even the largest halo's diameter. The benefit of sampling a profile in this manner is that it allows us to have in principle an infinitely resolved density field. The limit of the resolution no longer depends on the particle density of the simulation, instead solely on the resolution of the mesh used to measure the density field. As we shall see this is a great advantage when we come to compute statistics of the convergence field. The lack of particle binning means that we do not experience a Poisson noise contribution to the power spectrum at small scales. This in turn lets us probe far beyond the resolution limit of the supporting simulation.

6.2 Source plane results

As an initial test of our method for producing lensing statistics we assume that the lensing sources are uniform and distributed regularly on a grid on the sky at a redshift z_s which is located at the end of the current lightcone . The lightcone is then sliced into a number of shells, each $\Delta z = 0.2$ in width and the surface density calculated. The surface density is then weighted to give the convergence which is summed over each slice to produce the convergence field.

Before we present the results obtained by the method described in this chapter, it is instructive to examine the results of a more traditional method of simulating lensing. Dr Alina Kiessling (Royal Observatory Edinburgh) has produced lensing results from dark matter simulations using a method which stacks a number of simulation boxes, applying to each a centroid shift and rotation as discussed

in Chapter 5. To model the evolution of the density field, the redshift of the simulation output used to construct the convergence field is dynamically chosen to be as close to the current "observed" redshift as possible. Figure 6.3 shows the convergence power spectrum results obtained using this approach applied to a dark matter simulation of side $1\,h^{-1}\,\mathrm{Gpc}$, containing 1024^3 particles. The simulation was run using the GADGET2 code.

The red line in Figure 6.3 represents the theoretical prediction of the power spectrum while the upper black line shows the result from the simulation. Despite the fact that the particle load used to produce these results is 8 times larger than that used to produce the simulations which form the basis for the algorithm, the results at small scales are still dominated by Poisson noise. While Poisson noise is hard to model in simulations in which we are accumulating power over a number of simulation slices, the blue line is an attempt to estimate the level on the noise by appealing to simple $1/\bar{n}$ arguments. Attempting to subtract this component from the power spectrum gives the lower line in Figure 6.3. It is apparent, then that to probe smaller scale features in the power spectrum we need to use either vastly more computationally expensive simulations, or develop approaches like the one described in this thesis. Aside from the issues on small scales, there is also a notable dip in the power spectrum at the peak between $\ell \sim 2000$ and $\ell \sim 10^4$. While this is getting close to the region in which Poisson noise dominates, it is a striking feature.

In an attempt to improve on these results, we apply the algorithm described in this thesis to produce a convergence field. We use a simulation with 8 times less particles as a starting point and generate the lightcone we presented in Chapter 5, Figure (5.3). We then generate surface densities at a number of redshifts, and use code written by Alina to generate the convergence maps, shear maps and convergence power spectrum. Figures 6.4 and 6.5 show the surface density maps and the cumulative surface density, in slices of width $\Delta r = 231.27\,h^{-1}\,\mathrm{Mpc}$. Unfortunately in the realisation used for this example, there exists a large cluster ($10^{14}h^{-1}M_{\odot}$) very close to the observer which dominates the density field at low redshift. This cluster is visible in the first density map in Figure 6.4 and in the lightcone plot (Figure 5.3). From the cumulative density field we then obtain convergence and shear maps for each redshift slice. Figure 6.6 shows

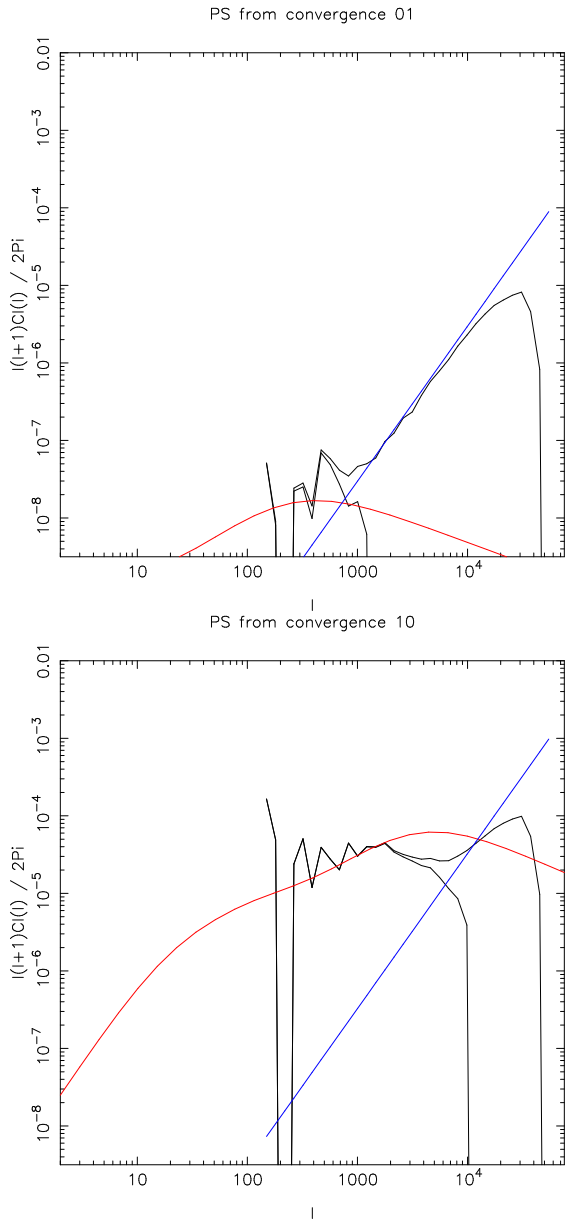

Figure 6.3: The convergence power spectrum for a source field at $z = 0.1$ (top) and $z = 1.0$ (bottom), computed by the binning of dark matter particles from a simulation with 1024^3 particles in a $1(\,h^{-1}\,\mathrm{Gpc})^3$ box. The red line is a theoretical prediction of the convergence power spectrum while the blue line is an estimate of the contribution from particle discreteness. The lower black line in each plot is the result of subtracting the Poisson noise estimate from the measured convergence. The power spectrum calculation was carried out by Dr Alina Kiessling.

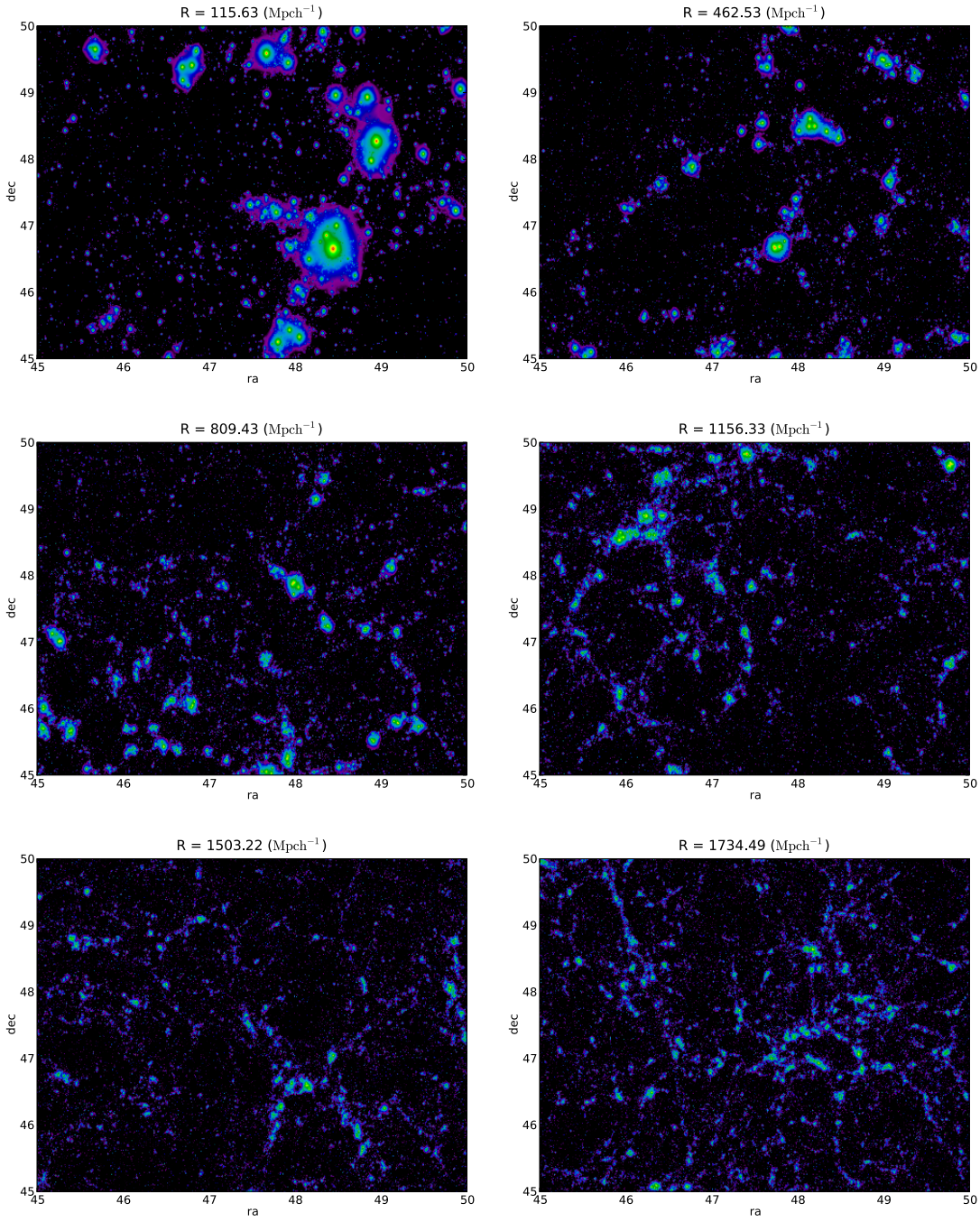

Figure 6.4: The contribution to the surface density field of slices in the radial direction. As we move to higher and higher redshifts the signal from individual haloes is weakened and the cosmic web becomes more visible.

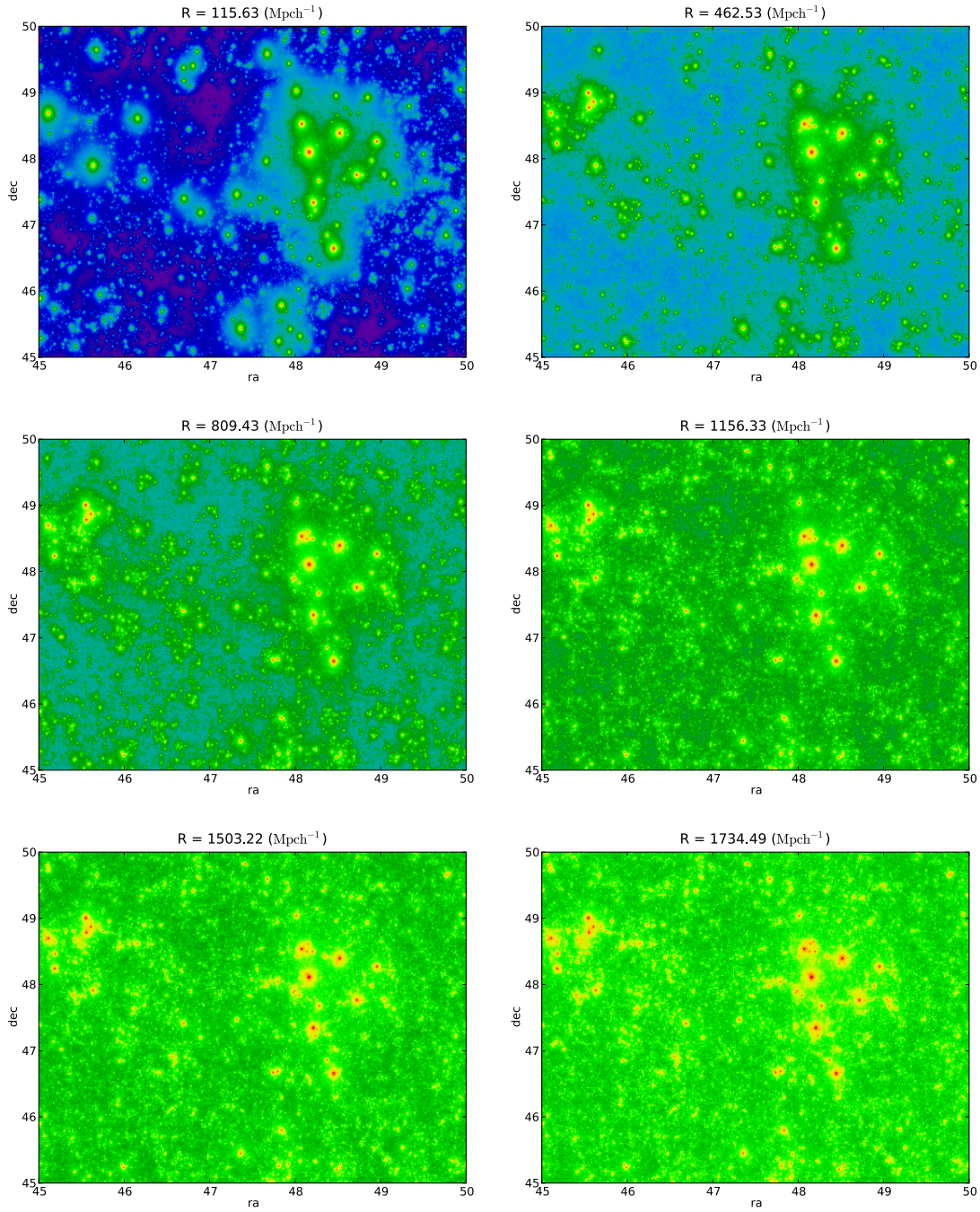

Figure 6.5: The evolution of the cumulative surface density field on a patch of sky $5^\circ \times 5^\circ$ across, moving out in comoving distance.

the convergence and shear from a slice centred at a comoving distance of $r = 1156\,h^{-1}\,\mathrm{Mpc}$. The influence of the large low redshift cluster can still be seen but its contribution to the convergence field is down-weighted as we place our sources at higher redshift by virtue of the distance factors, $D_l D_{ls}/D_s$, in Equation (6.2).

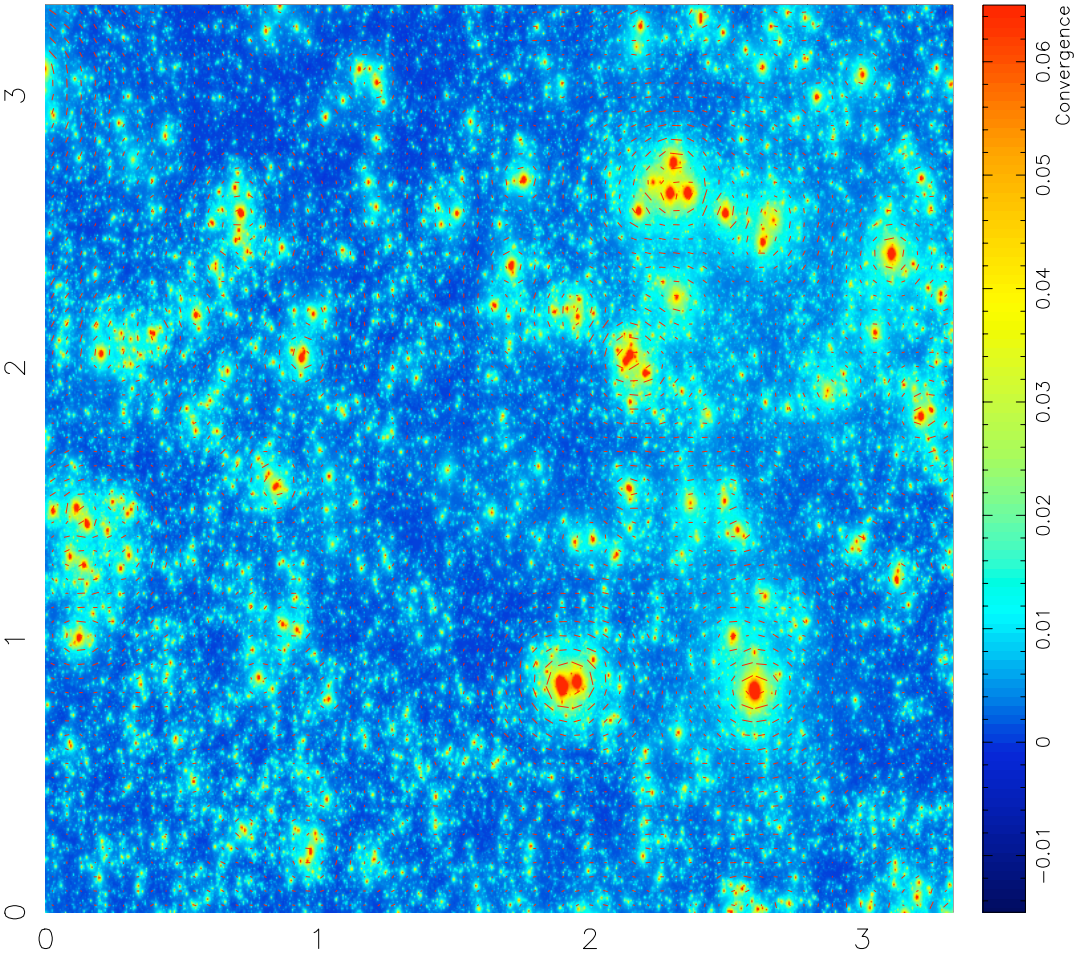

Figure 6.6: A convergence (colour scale) and shear (line elements) map calculated from the light cone. This is a slice through the light-cone centred on $z = 0.5$. The distorting effect of large clusters along with the smaller distortions induced by the large scale structure are clearly visible

.

The real test of the method however, comes from a comparison of the convergence power spectrums. The power spectrum is computed using the same code as the dark matter case. Figures 6.7 and 6.8 shows the convergence power spec-

trum at 4 redshifts. It shows a number of interesting features. Most notably at high ℓ the characteristic excess power from Poisson discreteness is absent. This illustrates one of the great advantages of our method, as we do not use discrete particles to construct the density field, but rather smooth profiles, the discreteness noise is not a factor. The spectrum appears to be initially high compared with theory, however we believe this is due to cosmic variance: the unfortunate presence of the low redshift cluster. Finally we observe the same worrying dip in power around the peak of the power spectrum. As we also observe hints of this feature in the more traditional dark matter calculation we believe that either again this is due to cosmic variance or some systematic in the convergence calculation which we have yet to determine.

6.2.1 Accounting for extra mass

The discussion above ignores an important fact which is that the mass in haloes is only a small fraction of the mass in the Universe. If we integrate the halo mass function from some threshold halo mass to ∞ we find that the fraction of the Universe in haloes will approach unity as the threshold approaches 0, but this convergence is slow. Even using the Monte Carlo approach defined in this thesis we will have to at some point include a minimum halo mass implying that when we construct the surface densities above we will in general be missing some fraction of the mass in the gravitational lens. Figure 6.9 shows the comparison between the convergence power spectrum for halo mass cuts of $10^{11}h^{-1}M_{\odot}$ and $10^{13}h^{-1}M_{\odot}$. The main effect of increasing the mass cut is to reduce the power at high ℓ. We have removed the contributions to the 1 halo term from small mass haloes. The scales in our power spectrum calculation are currently limited by the size of the 2D grid we are using to calculate. The cut off at $\ell \sim 40000$ is corresponds to this resolution limit, suggesting that even at a halo cut-off of $10^{11}h^{-1}M_{\odot}$ we are already getting close to a convergent result. However to be consistent and complete we require a method to account for this neglected mass.

While we have not fully implemented the following procedure it is one possible approach to restoring mass. When we populate our lightcone with haloes, we split the volume into cells of $15\,h^{-1}\,\mathrm{Mpc}$. If during this process we also take care to

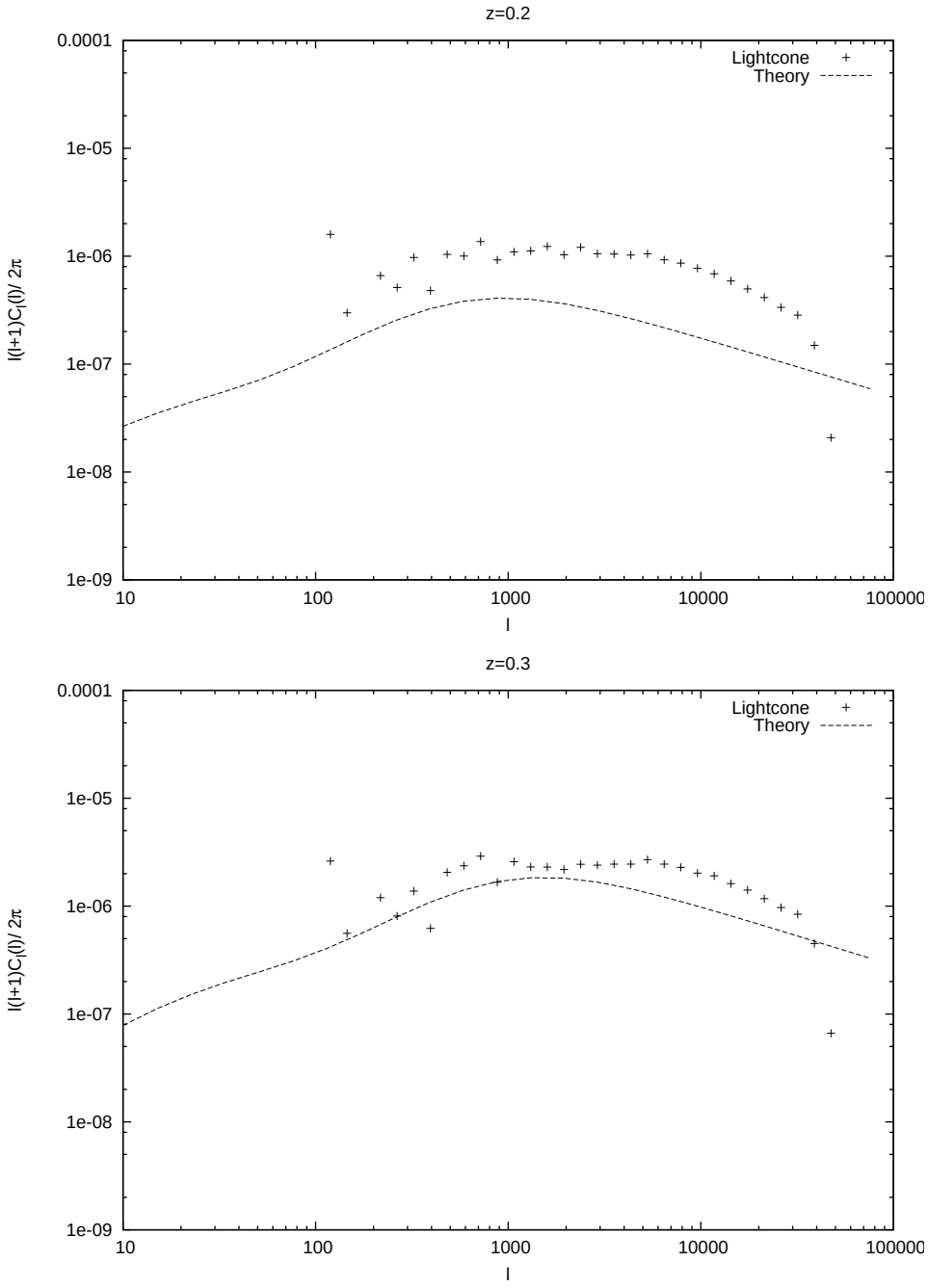

Figure 6.7: The cumulative convergence power spectrum as a function of distance from the observer. The sources are taken to be a regularly spaced grid of galaxies placed at the redshift of the furthest slice. The dashed line is the prediction from linear theory while the crosses are measured from simulations.

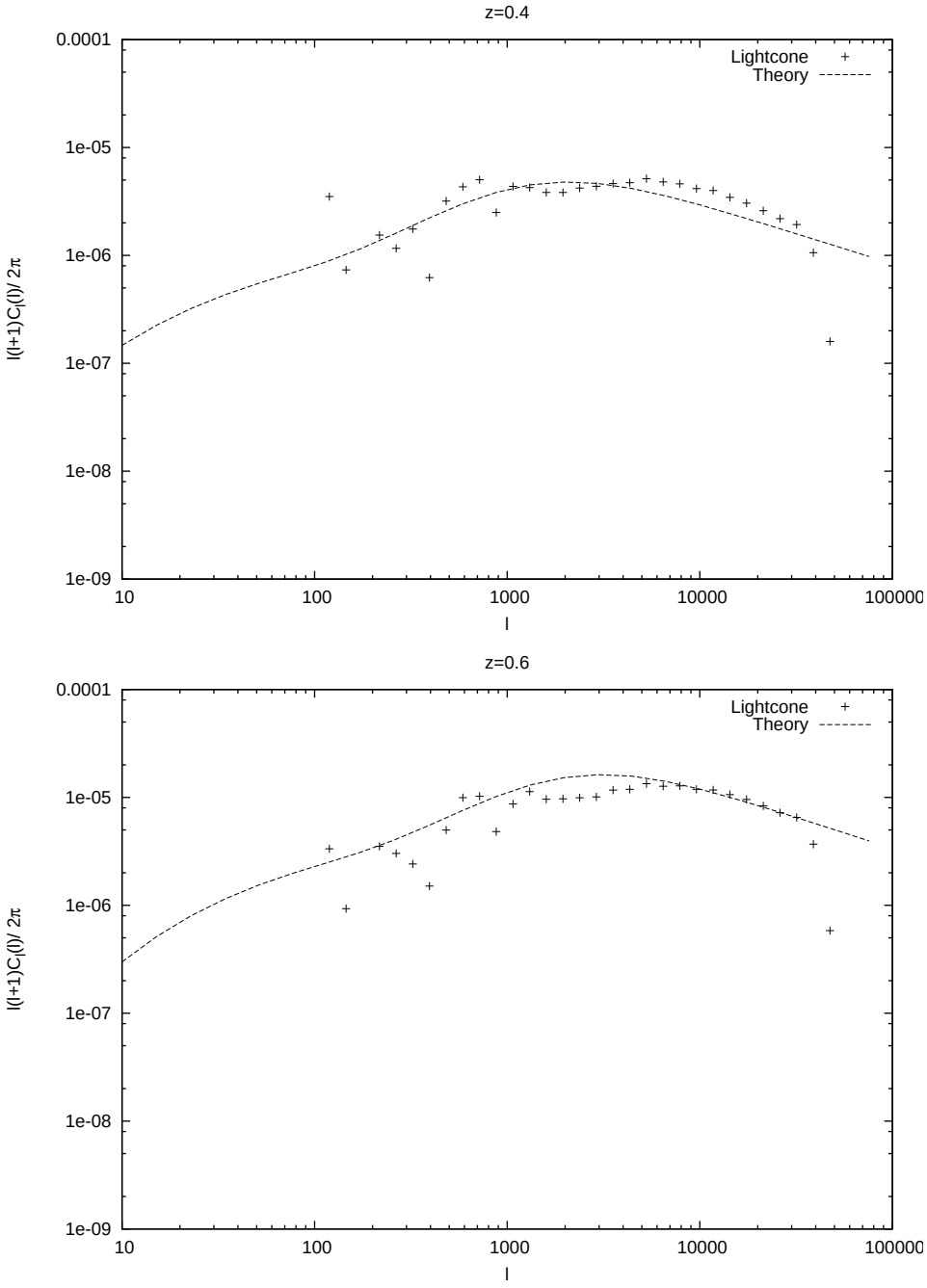

Figure 6.8: Continued from Figure 6.7.

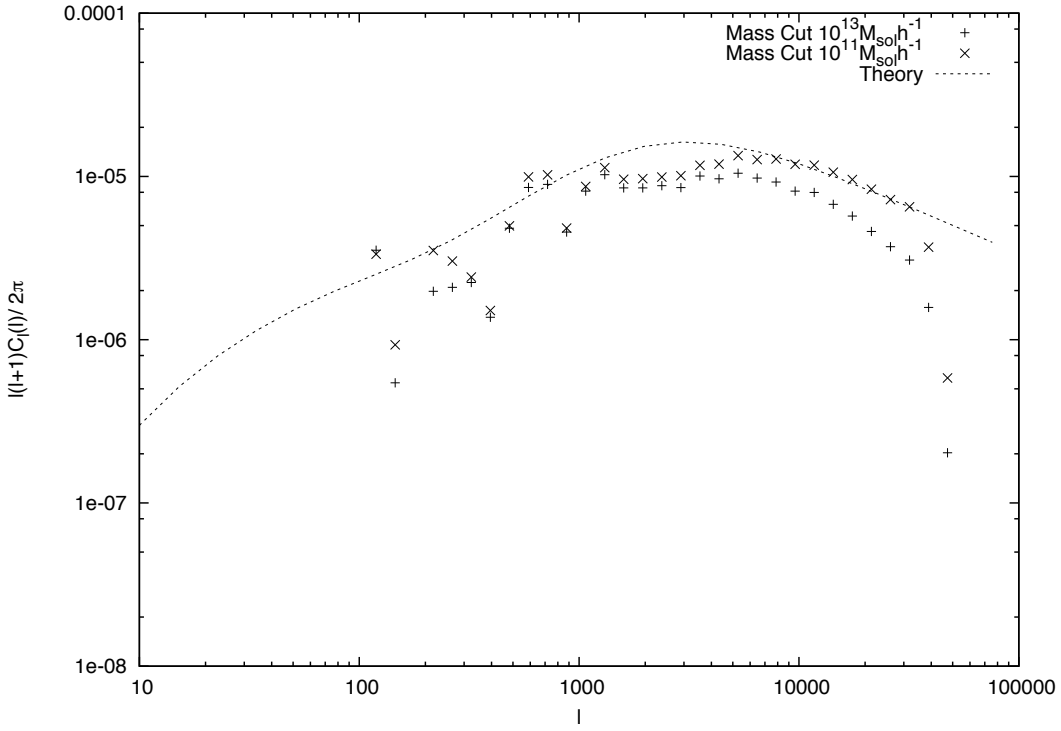

Figure 6.9: The cumulative convergence power spectrum up to $1734.49\,h^{-1}$ Mpc along the light cone ($z \sim 0.6$) for minimum mass cuts of $10^{11}h^{-1}M_{\odot}$ and $10^{13}h^{-1}M_{\odot}$. The biggest departure can be see at high ℓ as a decrease in the convergence. This is due to missing contributions of the smaller halos from the 1 halo term.

note the total mass partitioned into haloes, we will know on a cell by cell basis the exact mass deficit. At this point we know how much mass we need to account for but we need to know how to place it within the cell. The simplest method would just be to treat this mass as a constant background within the cell and distribute it randomly within the cell. The problem with this approach would be that some of this mass would end up in haloes, artificially producing an enhanced density and we would see sharp discontinuities at cell walls especially in regions bordering voids.

Instead we can treat the extra mass in much the same way as we placed the haloes in the original density field. After each halo has been placed in a cell the remaining mass can be divided into particles of a given mass M_p and placed by continuing to use the placement algorithm. This will obey the same collision rules as applied to haloes ensuring that the particles do not end up within the virial radius of any dark matter haloes and will in general trace the structure on a subcell level. When it comes to constructing the surface mass density we simply add in the extra mass particles along with the haloes.

6.2.2 Full sky survey

We have shown in this chapter that the application of our algorithm to lensing can not only replicate the results of more traditional particle based simulations, but can actually help probe the high ℓ domain. While this lightcone covers a small area of the sky there are a number of proposed surveys which will measure the lensing signal over large regions of the sky. The CFHTLS survey (Fu et al., 2008; Hoekstra et al., 2006) aims to cover an area of 410 square degrees while the PAN-STARRS survey aims to cover almost the entire visible sky. To produce an all sky mock, out to a redshift of $z = 1$, we would require a simulation volume which entirely encompasses a sphere with a radius of roughly $r = 3300\,h^{-1}\,\mathrm{Mpc}$. To accomplish this without an excessive amount of replication of boxes, will require simulation volumes of order $\ell = 2\,h^{-1}\,\mathrm{Gpc}$ or greater. At the time of writing the algorithm presented in this thesis can populate an individual simulation box of length $\ell = 1\,h^{-1}\,\mathrm{Gpc}$ with halos and galaxies in roughly 3 hours on a modern desktop computer. If we keep the cell size used to populate these boxes the same,

and simply increase the size of the box, the computational workload algorithm should scale linearly with the volume. A box of $\ell = 2\,h^{-1}\,\mathrm{Gpc}$ we would require a particle resolution of $n = 1024^3$, which is a very modest undertaking for even a small modern multiprocessor system.

There have already been a number of attempts to characterise the systematics of lensing on small scales, through projects such as the GREAT08 challenge (Bridle et al., 2009) and the Shear Testing Program (STEP) (Heymans et al., 2006; Massey et al., 2007a), both of which provide fake skies with a known matter density for groups to compare and contrast lensing codes. It is our hope that we can use this work to produce mock all sky surveys which could contribute to these programs.

6.3 Conclusions

We have adapted and expanded our method to produce shear maps from light-cones, and we have shown that we can recover the expected lensing signal well compared to theoretical predications and to traditional N-Body lensing simulations. Furthermore we have shown the potential of our method to provide a way of rapidly measuring the shear power spectrum without the worry of Poisson noise due to particle discreetness. The dressing of haloes with continuous density profiles allows us to use small simulations to probe much smaller scales than they would otherwise be able to.

With a little more work this approach can also provide us with a consistent catalogue of source galaxies. This will allow us to start probing questions about the systematics of lensing methods. Coupled with raytracing algorithms, this galaxy catalogue could also be used to produce fake optical observations which will hopefully prove to be a good data set to asses methods for measuring shear.

Chapter 7

Conclusions and future work

We have presented a novel method for producing mock catalogues from calibrated halo models. The accuracy of these models have been shown to be good when compared to a recreation of a known simulation volume, the Millennium Simulation. We have shown how the model can be extended to produce lightcones which incorporate evolution of the density field and galaxy properties. In turn we have shown how this density field can be adapted to produce lensing maps.

The mocks produced by this algorithm have been applied to a number of topics in modern cosmology; the baryon acoustic oscillation peak, redshift space distortions, topology and the production of convergence maps. For each application there is a scope to improve on the results we have obtained, by extending our algorithm and applying it to more areas. For the baryon acoustic oscillations we plan to extend the analysis to a larger number of simulations to get more accurate fits, and calculate the shift in the peak at a number of different simulation outputs to trace its evolution with time. With redshift space distortions we plan to refine our calculation of the correlation function to be completely on our lightcone, rather than in a wide angle approximation, and repeat our analysis of the distortion parameter β over a number of redshift slices for each galaxy population. This will allow limits to be placed on the ability of future surveys to identify departures from General Relativity. We aim to repeat the measurements we have

made of topology, extending the analysis to redshift space and for a number of other epochs to see if the distinction between red and blue galaxies is present in each, and to obtain a better understanding of how topology evolves with redshift. We also hope to test our prediction of markedly different topological signatures for red and blue galaxies on survey data from the 2dFGRS to determine if the effect is real.

7.1 Improving the model

The model has a number of areas in which it could be improved. By far the largest deviation from the MS results which the model displays is the excess clustering on scales $r < 2\,h^{-1}\,\mathrm{Mpc}$. This is obviously a feature of the halo placement model on a subcell scale. There are a number of ways this might be improved. A basic empirical fix could be to simply select two halos within a cell and increase their separation by a small random amount along the line joining their centres . Given two halos within the same subcell with position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$, with separation vector $\mathbf{r}_{12}$, the strategy would be to alter the positions to $\mathbf{r}_1 = \mathbf{r}_1 + x\hat{\mathbf{r}}_{12}$, where x is a small ($x << |\mathbf{r}_{12}|$) random number. The magnitude of the separation boost could then be tuned to match the correlation function on small scales. A more advanced approach would be to model the typical correlations between halos of given mass within the cell. A number of marked, conditional correlation functions could be measured from the MS allowing an estimate of the clustering between haloes of different masses in cells of differing overdensities, $\xi(r, M_1, M_2, \delta)$. The halo placement algorithm could then be adapted to place the largest halos in the cell first, as is done currently, and then for the placement of all subsequent haloes by randomly selecting a partner and sampling a separation for the conditional correlation function. This would define a radius, about an existing halo, at which we should place the new halo, but we would still need to select its orientation. This could be done by selecting the angles for which the sphere the new halo will sit on passes through the highest density. We aim to explore this and other ideas for better ways of placing the dark matter haloes.

The second improvement to the model would be to increase the number of galaxy properties which are calculated. In its present incarnation the algorithm

assigns each galaxy 8 properties, three Cartesian position coordinates, three Cartesian velocity components , an r-band magnitude and a colour. It would be highly beneficial to be able to select a full colour for a galaxy rather than simply an assignment to red or blue. This could potentially be done with multiple colours allowing the determination of a number of magnitude bands for the galaxy. Stellar mass would also be an important property to assign to our galaxies. This could either be done in an analogous way to the method currently used to assign luminosities to galaxies, or it could be obtained by considering a joint probability distribution between the halo mass, galaxy luminosity and colour of the galaxy, ie $P(M_\star|M_h, L, b-v)$. The only limits to the type of galaxy property we can implement is either, the availability of a theoretical model linking that galaxy property with the halo mass and having sufficient numbers of objects in a large simulations like the MS to calibrate the required statistics .

7.2 Future Applications

By far the most attractive properties of the algorithm developed in this thesis is its speed at producing mock catalogues. On a modern desktop computer we can produce a full $1\,h^{-1}\,\mathrm{Gpc}$ box, with galaxies and halos, in around 24 hours (including the time taken to run the initial simulation). In contrast to semi-analytic approaches, which can take weeks or even months on large multi-processor machines, this allows us to produce a large number of realisations in a modest time. By necessity, semi-analytic simulations usually adopt a given cosmology, typically the most recent constraints from observations. In reality the dependence of results on variations in cosmological parameters, and the degeneracies between these parameters, are as important as the mean values. An attractive use of the algorithm presented here would be to explore these degeneracies through large numbers of realisations with varying cosmological parameter sets. The dark matter component can be easily adjusted to a new set of cosmological parameters by adjusting the parameters of our simulation. Likewise the halo statistics have a well defined dependence on cosmology through the conditional mass function formalism. However galaxy statistics which form the input to our method are calibrated from the MS, which has a fixed cosmology. To adjust the galaxy statistics

to accommodate new cosmologies requires either a theoretically motivated model, or a way of empirically calibrating the statistics. One possible way to proceed, would be to assume that for small perturbations of the cosmological parameters, the variation in galaxy properties will be a smooth, progressive variation away from those of the MS. In this case the modelling which was applied to quantify the conditional luminosity function would still be valid. To quantify the parameters of the model we could run a number of small volume simulations, with varying cosmological parameters, to which semi-analytic codes could be quickly applied to produce galaxy catalogues. While these simulations would be too small to quantify the statistics of the full dynamic range of haloes, missing out the rarest fluctuations, it would allow for the identification of the trends in the halo occupation statistics. For a large simulation with a given cosmology, the galaxy statistics can then be interpolated to the parameters in question. This would allow us to test the dependence of the results in the areas we have explored in this thesis, redshift space distortions, BAOs, topology and lensing, on the uncertainties in the cosmological parameters.

The extension of the model to lensing not only recovers the results of more traditional particle based codes, but can also probe much smaller scales. The real power of our approach to lensing is that it produces a self-consistent galaxy and mass distribution. As we have discussed before in Chapter 6, this could be a valuable tool to allow lensing teams to calibrate and test methods for determining mass distributions from shear measurements. By placing either ellipsoids or full images of galaxies at the positions of galaxies within our lightcone, we can ray trace through the calculated convergence field to determine the distortion induced by weak lensing on the source. The resulting fake "skies" can then be used to test lensing codes by comparing the recovered mass distribution to the original simulation. In this case our model does not need to be perfect, as all that is required to test the analysis methods is that the lensing effect on the sources and the density field are consistent.

The analysis of lensing also needs to be concerned with a number of possible systematic effects. Most calculations and methods assume a number of simplifying conditions apply to the source galaxy distributions. The two most common assumptions are that the galaxy population itsself is not clustered and that the

ellipticities of galaxies are uncorrelated and random (Hirata and Seljak, 2004; Heavens et al., 2000). In practice this is not true. The background galaxy population is clustered to some level, and there are a number of mechanisms which can introduce correlations between the location and ellipticity of the a galaxy, such as the effect of large scale tidal fields on galaxy ellipticity. With a consistent dark matter / galaxy population these issues could be addressed. The main barrier to accomplishing this is that the model we have presented currently does not include any information about the morphology or orientation of a galaxy. An approximation which would allow us to include such a property would be to align each galaxy in the simulation with the tidal field. This could be done in practice in much the same way as we assign halos a velocity. Another possible method would be to introduce ellipticities by appealing to correlations between galaxy spin statistics and the density field (Lee and Pen, 2001).

A further future goal of this work is to make the source code used in our algorithm publicly available. We believe that this novel approach to modelling the large scale structure of the Universe will be very valuable to the next generation of surveys, and that many people, observers and theorists, will find the ability to populate large volumes quickly invaluable.

Bibliography

Angulo, R. E., Baugh, C. M., Frenk, C. S., and Lacey, C. G.: 2008, MNRAS **383**, 755

Barnes, J. and Hut, P.: 1986, Nature **324**, 446

Bell, E. F., Wolf, C., Meisenheimer, K., Rix, H.-W., Borch, A., Dye, S., Kleinheinrich, M., Wisotzki, L., and McIntosh, D. H.: 2004, ApJ **608**, 752

Benson, A. J., Frenk, C. S., Lacey, C. G., Baugh, C. M., and Cole, S.: 2002a, MNRAS **333**, 177

Benson, A. J., Lacey, C. G., Baugh, C. M., Cole, S., and Frenk, C. S.: 2002b, MNRAS **333**, 156

Blaizot, J., Wadadekar, Y., Guiderdoni, B., Colombi, S. T., Bertin, E., Bouchet, F. R., Devriendt, J. E. G., and Hatton, S.: 2005, MNRAS **360**, 159

Blake, C. and Glazebrook, K.: 2003, ApJ **594**, 665

Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N.: 1991, ApJ **379**, 440

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., and Lacey, C. G.: 2006, MNRAS **370**, 645

Bridle, S., Shawe-Taylor, J., Amara, A., Applegate, D., Balan, Berge, J. S. T., Bernstein, G., Dahle, H., Erben, T., Gill, M., Heavens, A., Heymans, C., High, F. W., Hoekstra, H., Jarvis, M., Kirk, D., Kitching, T., Kneib, J., Kuijken, K., Lagatutta, D., Mandelbaum, R., Massey, R., Mellier, Y., Moghaddam, B., Moudden, Y., Nakajima, R., Paulin-Henriksson, S., Pires, S., Rassat, A., Refregier, A., Rhodes, J., Schrabback, T., Semboloni, E., Shmakova, M., van

Waerbeke, L., Witherick, D., Voigt, L., and Wittman, D.: 2009, *Annals of Applied Statistics* **3**, 6

Bryan, G. L. and Norman, M. L.: 1998, ApJ **495**, 80

Bullock, J. S., Kolatt, T. S., Sigad, Y., Somerville, R. S., Kravtsov, A. V., Klypin, A. A., Primack, J. R., and Dekel, A.: 2001, MNRAS **321**, 559

Clowe, D., Bradač, M., Gonzalez, A. H., Markevitch, M., Randall, S. W., Jones, C., and Zaritsky, D.: 2006, ApJ **648**, L109

Cole, S. and Kaiser, N.: 1989, MNRAS **237**, 1127

Cole, S., Lacey, C. G., Baugh, C. M., and Frenk, C. S.: 2000, MNRAS **319**, 168

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., and Taylor, K.: 2005, MNRAS **362**, 505

Coles, P. and Jones, B.: 1991, MNRAS **248**, 1

Colless, M.: 1999, *Royal Society of London Philosophical Transactions Series A* **357**, 105

Collister, A. A. and Lahav, O.: 2004, PASP **116**, 345

Cooray, A.: 2006, MNRAS **365**, 842

Cooray, A. and Milosavljević, M.: 2005, ApJ **627**, L85

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., and Yoshida, N.: 2006, MNRAS **365**, 11

de Blok, W. J. G. and McGaugh, S. S.: 1997, MNRAS **290**, 533

De Lucia, G. and Blaizot, J.: 2007, MNRAS **375**, 2

Dekel, A. and Silk, J.: 1986, ApJ **303**, 39

Dickinson, C., Battye, R. A., Carreira, P., Cleary, K., Davies, R. D., Davis, R. J., Genova-Santos, R., Grainge, K., Gutiérrez, C. M., Hafez, Y. A., Hobson, M. P., Jones, M. E., Kneissl, R., Lancaster, K., Lasenby, A., Leahy, J. P., Maisinger, K., Ödman, C., Pooley, G., Rajguru, N., Rebolo, R., Rubiño-Martin, J. A., Saunders, R. D. E., Savage, R. S., Scaife, A., Scott, P. F., Slosar, A., Sosa Molina, P., Taylor, A. C., Titterington, D., Waldram, E., Watson, R. A., and Wilkinson, A.: 2004, MNRAS **353**, 732

Drory, N., Salvato, M., Gabasch, A., Bender, R., Hopp, U., Feulner, G., and Pannella, M.: 2005, ApJ **619**, L131

Dunkley, J., Komatsu, E., Nolta, M. R., Spergel, D. N., Larson, D., Hinshaw, G., Page, L., Bennett, C. L., Gold, B., Jarosik, N., Weiland, J. L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., and Wright, E. L.: 2009, ApJS **180**, 306

Efstathiou, G., Davis, M., White, S. D. M., and Frenk, C. S.: 1985, ApJS **57**, 241

Eisenstein, D. J. and Hu, W.: 1998, ApJ **496**, 605

Eisenstein, D. J., Seo, H.-J., and White, M.: 2007, ApJ **664**, 660

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J., Connolly, A., Csabai, I., Doi, M., Fukugita, M., Frieman, J. A., Glazebrook, K., Gunn, J. E., Hendry, J. S., Hennessy, G., Ivezić, Z., Kent, S., Knapp, G. R., Lin, H., Loh, Y.-S., Lupton, R. H., Margon, B., McKay, T. A., Meiksin, A., Munn, J. A., Pope, A., Richmond, M. W., Schlegel, D., Schneider, D. P., Shimasaku, K., Stoughton, C., Strauss, M. A., SubbaRao, M., Szalay, A. S., Szapudi, I., Tucker, D. L., Yanny, B., and York, D. G.: 2005, ApJ **633**, 560

Feldman, H. A., Kaiser, N., and Peacock, J. A.: 1994, ApJ **426**, 23

Fu, L., Semboloni, E., Hoekstra, H., Kilbinger, M., van Waerbeke, L., Tereno, I., Mellier, Y., Heymans, C., Coupon, J., Benabed, K., Benjamin, J., Bertin, E., Doré, O., Hudson, M. J., Ilbert, O., Maoli, R., Marmo, C., McCracken, H. J., and Ménard, B.: 2008, A&A **479**, 9

Gao, L. and White, S. D. M.: 2007, MNRAS **377**, L5

Glazebrook, K., Tober, J., Thomson, S., Bland-Hawthorn, J., and Abraham, R.: 2004, AJ **128**, 2652

Gott, III, J. R., Dickinson, M., and Melott, A. L.: 1986, ApJ **306**, 341

Gott, J. R. I., Weinberg, D. H., and Melott, A. L.: 1987, ApJ **319**, 1

Gray, A. G., Moore, A. W., Nichol, R. C., Connolly, A. J., Genovese, C., and Wasserman, L.: 2004, *Multi-Tree Methods for Statistics on Very Large Datasets in Astronomy*, Astronomical Society of the Pacific Conference Series, Astronomical Society of the Pacific, 390 Ashton Avenue - San Francisco, CA 94112

Guzzo, L., Pierleoni, M., Meneux, B., Branchini, E., Le Fèvre, O., Marinoni, C., Garilli, B., Blaizot, J., De Lucia, G., Pollo, A., McCracken, H. J., Bottini, D., Le Brun, V., Maccagni, D., Picat, J. P., Scaramella, R., Scodeggio, M., Tresse, L., Vettolani, G., Zanichelli, A., Adami, C., Arnouts, S., Bardelli, S., Bolzonella, M., Bongiorno, A., Cappi, A., Charlot, S., Ciliegi, P., Contini, T., Cucciati, O., de la Torre, S., Dolag, K., Foucaud, S., Franzetti, P., Gavignaud, I., Ilbert, O., Iovino, A., Lamareille, F., Marano, B., Mazure, A., Memeo, P., Merighi, R., Moscardini, L., Paltani, S., Pellò, R., Perez-Montero, E., Pozzetti, L., Radovich, M., Vergani, D., Zamorani, G., and Zucca, E.: 2008, Nature **451**, 541

Hamilton, A. J. S.: 1992, ApJ **385**, L5

Heath, D. J.: 1977, MNRAS **179**, 351

Heavens, A., Refregier, A., and Heymans, C.: 2000, MNRAS **319**, 649

Heymans, C., Van Waerbeke, L., Bacon, D., Berge, J., Bernstein, G., Bertin, E., Bridle, S., Brown, M. L., Clowe, D., Dahle, H., Erben, T., Gray, M., Hetterscheidt, M., Hoekstra, H., Hudelot, P., Jarvis, M., Kuijken, K., Margoniner, V., Massey, R., Mellier, Y., Nakajima, R., Refregier, A., Rhodes, J., Schrabback, T., and Wittman, D.: 2006, MNRAS **368**, 1323

Hinshaw, G., Spergel, D. N., Verde, L., Hill, R. S., Meyer, S. S., Barnes, C., Bennett, C. L., Halpern, M., Jarosik, N., Kogut, A., Komatsu, E., Limon, M., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L.: 2003, ApJS **148**, 135

Hirata, C. M. and Seljak, U.: 2004, Phys. Rev. D **70(6)**, 063526

Hoekstra, H., Mellier, Y., van Waerbeke, L., Semboloni, E., Fu, L., Hudson, M. J., Parker, L. C., Tereno, I., and Benabed, K.: 2006, ApJ **647**, 116

Kaiser, N.: 1987, MNRAS **227**, 1

Kim, J. and Park, C.: 2006, ApJ **639**, 600

Kitzbichler, M. G. and White, S. D. M.: 2007, MNRAS **376**, 2

Kolb, E. W., Salopek, D. S., and Turner, M. S.: 1990, *NASA STI/Recon Technical Report N* **90**, 25915

Kuo, C. L., Ade, P. A. R., Bock, J. J., Cantalupo, C., Daub, M. D., Goldstein, J., Holzapfel, W. L., Lange, A. E., Lueker, M., Newcomb, M., Peterson, J. B., Ruhl, J., Runyan, M. C., and Torbet, E.: 2004, ApJ **600**, 32

Lacey, C. and Cole, S.: 1993, MNRAS **262**, 627

Larson, R. B.: 1974, MNRAS **169**, 229

Lee, J. and Pen, U.-L.: 2001, ApJ **555**, 106

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

Linder, E. V.: 2005, Phys. Rev. D **72(4)**, 043529

Madgwick, D. S., Hawkins, E., Lahav, O., Maddox, S., Norberg, P., Peacock, J. A., Baldry, I. K., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lewis, I., Lumsden, S., Peterson, B. A., Sutherland, W., and Taylor, K.: 2003, MNRAS **344**, 847

Massey, R., Heymans, C., Bergé, J., Bernstein, G., Bridle, S., Clowe, D., Dahle, H., Ellis, R., Erben, T., Hetterscheidt, M., High, F. W., Hirata, C., Hoekstra, H., Hudelot, P., Jarvis, M., Johnston, D., Kuijken, K., Margoniner, V., Mandelbaum, R., Mellier, Y., Nakajima, R., Paulin-Henriksson, S., Peeples, M., Roat, C., Refregier, A., Rhodes, J., Schrabback, T., Schirmer, M., Seljak, U., Semboloni, E., and van Waerbeke, L.: 2007a, MNRAS **376**, 13

Massey, R., Rhodes, J., Ellis, R., Scoville, N., Leauthaud, A., Finoguenov, A., Capak, P., Bacon, D., Aussel, H., Kneib, J.-P., Koekemoer, A., McCracken, H., Mobasher, B., Pires, S., Refregier, A., Sasaki, S., Starck, J.-L., Taniguchi, Y., Taylor, A., and Taylor, J.: 2007b, Nature **445**, 286

Matsubara, T.: 1994, ApJ **434**, L43

Matsubara, T. and Suto, Y.: 1996, ApJ **460**, 51

Mo, H. J. and White, S. D. M.: 1996, MNRAS **282**, 347

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., and Tozzi, P.: 1999a, ApJ **524**, L19

Moore, B., Quinn, T., Governato, F., Stadel, J., and Lake, G.: 1999b, MNRAS **310**, 1147

Nagamine, K., Cen, R., Hernquist, L., Ostriker, J. P., and Springel, V.: 2004, ApJ **610**, 45

Nagamine, K., Cen, R., Hernquist, L., Ostriker, J. P., and Springel, V.: 2005, ApJ **627**, 608

Navarro, J. F., Frenk, C. S., and White, S. D. M.: 1997, ApJ **490**, 493

Neto, A. F., Gao, L., Bett, P., Cole, S., Navarro, J. F., Frenk, C. S., White, S. D. M., Springel, V., and Jenkins, A.: 2007, MNRAS **381**, 1450

Park, S. K. and Miller, K. W.: 1988, *Commun. ACM* **31**, 1192

Parkinson, D., Blake, C., Kunz, M., Bassett, B. A., Nichol, R. C., and Glazebrook, K.: 2007, MNRAS **377**, 185

Percival, W. J., Reid, B. A., Eisenstein, D. J., Bahcall, N. A., Budavari, T., Frieman, J. A., Fukugita, M., Gunn, J. E., Ivezić, Ž., Knapp, G. R., Kron, R. G., Loveday, J., Lupton, R. H., McKay, T. A., Meiksin, A., Nichol, R. C., Pope, A. C., Schlegel, D. J., Schneider, D. P., Spergel, D. N., Stoughton, C., Strauss, M. A., Szalay, A. S., Tegmark, M., Vogeley, M. S., Weinberg, D. H., York, D. G., and Zehavi, I.: 2010, MNRAS **401**, 2148

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and The Supernova Cosmology Project: 1999, ApJ **517**, 565

Press, W. H. and Schechter, P.: 1974, ApJ **187**, 425

Readhead, A. C. S., Mason, B. S., Contaldi, C. R., Pearson, T. J., Bond, J. R., Myers, S. T., Padin, S., Sievers, J. L., Cartwright, J. K., Shepherd, M. C., Pogosyan, D., Prunet, S., Altamirano, P., Bustos, R., Bronfman, L., Casassus, S., Holzapfel, W. L., May, J., Pen, U., Torres, S., and Udomprasert, P. S.: 2004, ApJ **609**, 498

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J.: 1998, AJ **116**, 1009

Robertson, H. P.: 1935, ApJ **82**, 284

Rubiño-Martín, J. A., Betancort-Rijo, J., and Patiri, S. G.: 2008, MNRAS **386**, 2181

Sánchez, A. G., Baugh, C. M., Percival, W. J., Peacock, J. A., Padilla, N. D., Cole, S., Frenk, C. S., and Norberg, P.: 2006, MNRAS **366**, 189

Scoccimarro, R. and Sheth, R. K.: 2002, MNRAS **329**, 629

Seo, H. and Eisenstein, D. J.: 2005, ApJ **633**, 575

Sheth, R. K. and Diaferio, A.: 2001, MNRAS **322**, 901

Sheth, R. K., Mo, H. J., and Tormen, G.: 2001, MNRAS **323**, 1

Sheth, R. K. and Tormen, G.: 1999, MNRAS **308**, 119

Simon, P., Taylor, A. N., and Hartlap, J.: 2009, MNRAS **399**, 48

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., and Couchman, H. M. P.: 2003, MNRAS **341**, 1311

Smith, R. E., Scoccimarro, R., and Sheth, R. K.: 2008, Phys. Rev. D **77(4)**, 3525

Somerville, R. S.: 2002, ApJ **572**, L23

Springel, V.: 2005, MNRAS **364**, 1105

Springel, V., White, S. D. M., Tormen, G., and Kauffmann, G.: 2001, MNRAS **328**, 726

Sutherland, R. S. and Dopita, M. A.: 1993, ApJS **88**, 253

Weinberg, D. H.: 1988, PASP **100**, 1373

White, S. D. M. and Rees, M. J.: 1978, MNRAS **183**, 341

Wright, C. O. and Brainerd, T. G.: 2000, ApJ **534**, 34

Zel'Dovich, Y. B.: 1970, A&A **5**, 84