THE ENVIRONMENTAL DEPENDENCE OF DARK MATTER AND GALAXY CLUSTERING

by

Ummi Abbas

M.S. in Physics, Utkal University, 2000

M.S. in Physics, University of Pittsburgh, 2002

Submitted to the Graduate Faculty of

the Department of Physics and Astronomy in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

DEPARTMENT OF PHYSICS AND ASTRONOMY

This dissertation was presented

by

Ummi Abbas

It was defended on

April 13, 2006

and approved by

Ravi K. Sheth, Assistant Professor, Physics & Astronomy

David A. Turnshek, Professor, Physics & Astronomy

Donna Naples, Associate Professor, Physics & Astronomy

Andrew J. Connolly, Associate Professor, Physics & Astronomy

Rupert A. C. Croft, Associate Professor, Physics & Astronomy, CMU

Dissertation Advisors: Ravi K. Sheth, Assistant Professor, Physics & Astronomy,

David A. Turnshek, Professor, Physics & Astronomy

THE ENVIRONMENTAL DEPENDENCE OF DARK MATTER AND GALAXY CLUSTERING

Ummi Abbas, PhD

University of Pittsburgh, 2006

We have addressed important astrophysical issues of the large scale structure of the universe with the help of analytical and numerical tools. In particular, we have concentrated on obtaining a better understanding of the environmental dependence of clustering of dark matter and galaxies. We will discuss the development of the analytical framework, the generation of mock catalogs of data, and present results from numerical studies of the comparison of the theory with the simulated data samples and subsequently with actual observational data.

The N-body simulations from the Virgo consortium and SDSS data were used to study the clustering properties of dense and less dense subsamples with the help of the correlation function. The dense sample shows a correlation function that is about a factor of ten times or more higher than the less dense sample at large scales. This discernible difference holds true for mock galaxies, dark matter, and data from the Sloan Digital Sky Survey. By utilizing the halo model and adapting the formalism to different environments, we present the correspondence between theoretical models and simulated samples and real data.

The very good agreement between the model and results from simulations and data provides strong support for the hierarchical clustering models. This suggests that unless care is taken to study galaxies at fixed halo mass, observed correlations between gastrophysical effects and environment are dominated by the fact that these effects also correlate with the halo mass, and halo mass correlates with environment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Life is like a road. As one moves along, there are many cross-sections, stop signs, red lights and more importantly green lights. Through certain sections of this road one drives along with other people, sometimes more than just one. My drive into the career of academics was spearheaded by my father. He always thought that if one has the brains then one should put them to good use. And my mother was always there to stand by the choices I made. She is the most amazing and beautiful woman I know, I have always felt that I have taken after her in more ways than one. My brother, Samar, always taught me that nothing is impossible and the importance of having a back-up plan. I owe everything to my family for standing by me at all times. The decision of taking up Physics as my field of research was due to the reason that in college I enjoyed not having to memorize unconnected facts. Theorems and problems could be solved in a straightforward and simple manner. Not only did I find this stimulating, but enjoyable as well. It is true that at times I felt that I was in overdrive, and had to stop and reconsider the path I was taking. But, in the end, the will to think prevailed. The decision to come to the USA to pursue my PhD was extremely difficult and I owe a lot to my parents for standing by me. If for once they had said no, I would never have come to be here.

My theoretical inclination was mostly a byproduct of the academic environment I was brought up in, very little work in labs, mostly deriving and solving equations. As such when I first came to PITT I was stumped as to what topic and with whom I wanted to do research with. My fascination for Astrophysics led me to attend Astrolunch seminars and meetings. The first summer I tried my hand at Neutrino astrophysics with the HEP group under the supervision of Donna Naples and Vittorio Paolone. It was an interesting first experience and introduction into the world of research. Vittorio taught me how to approach problems and

solve them. The theory side of me itched and was relieved by the arrival of a theoretical astrophysicist, Ravi Sheth, who eventually became my thesis advisor. Donna encouraged me to take up research with Ravi, and I thank her and Vittorio for their encouragement and support.

Ravi Sheth has taught me many things during my transition from being a student to being a researcher. It was a painstaking and slow transition and he has always been patient and kept me stimulated with encouraging remarks and thought provoking discussions. He would always remind me that I should think for myself and not take arguments for granted. I thank him for all the help he has given me and for giving me emotional support. Through him I have to come to slowly understand and enjoy the complicated world of research.

The road until now would not have been smooth without the wonderful people who helped me through it all. Each in their own particular way have enriched my life and made the experience as complete as can be. At the beginning the girls were there to give me advice and guide me through the rigorous coursework, Michele, Rukmini, Voica, Ching-Wa and Yvette. In later years, the people I shared my office with, Vahit, Lorenzo, Niraj, Sam, Kip, Phil, would give me emotional support directly and often indirectly as we all were going through similar experiences. Insightful discussions and help from Ramin and Cameron helped me through many difficult scientific problems. Outside my office space and department, Barun, Anand and Mahesh, have all commanded my respect, and helped me through difficult times. I thank all these great friends and many others I have come in touch with. The office staff, Leyla, Michele, Heidi, Judy, Laurie, Laura, Mary and Jim have made my time at the department as smooth and problem-free as can be. Leyla has always impressed me with her efficiency, and work ethic, and above all her caring attitude.

The members of my thesis committee have all helped me through trying times. Dave Turnshek has given me a lot of encouragement. Under his direction and guidance the road to my PhD has been very smooth. Andy and Rupert have both helped me too, thanks to them all.

Finally, thanks to Alessandro Sozzetti for just being there, always. The past years have allowed me to see things through his eyes and learn many an important lesson. The rusty car I had been in has now become a slick and shiny sports car. Obstacles seem to be mere

pebbles...

I am also dedicating this thesis to my old and dear friend, Pinky. Her sweet nature will always have a special corner in my heart. Even though a short life, the memories are an everlasting treasure.

We thank the Virgo consortium for making their simulations available to the public, and the Pittsburgh Computational Astrostatistics group (PiCA) for the NPT code which was used to measure the correlation functions in the simulations in the first part of the work of this thesis. We thank Ramin Skibba for many helpful discussions, Cameron McBride and Jeff Gardner for the NTropy code which was also used to measure the correlation functions and projected statistics in the simulations and data, Andrew Connolly and Ryan Scranton for providing the SDSS data samples used in the thesis.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, the Korean Scientist Group, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

Part of the work in this thesis was supported by the National Science Foundation under Grants No. 0307747 and 0520647, Mary Warga Fellowship, and several Teaching Assistantships.

To see the world in a grain of sand,

Heaven in a wild flower,

Hold infinity in the palm of your hand,

And eternity in an hour.

-William Blake

TO MY FAMILY.

1.0 INTRODUCTION

Present observations with powerful telescopes and recent space borne missions have shown that we are living in a universe that is comprised of 73% dark energy and 23% non-baryonic dark matter, while only 4% is in the form of ordinary baryonic matter. The dark energy and non-baryonic dark matter have not been reproduced in a lab; they are indirectly inferred from observations of stars and galaxies. In Appendix A the components of the Universe are briefly discussed. This picture has led to the Lambda-Cold Dark Matter (ΛCDM) cosmological model, based on the general theory of relativity and the cosmological principle. In this model the universe is continuously evolving and it is currently undergoing a phase of accelerated expansion. The ΛCDM model is corroborated by various observations such as the Hubble flow, primordial nucleosynthesis, the cosmic microwave background and distance measures from type IA supernovae. This is further elaborated on in Appendix B. In addition, observations show that the distribution of matter at large scales is in the form of well-defined structures. This large scale structure provides further support in favor of the current cosmological model.

1.1 LARGE SCALE STRUCTURE

The Universe appears to be hierarchically structured. Ordering the structures seen on smallest to largest scales, we have stars, galaxies, clusters and superclusters. Until the 1980's superclusters were assumed to be the largest structures in the universe with an almost uniform distribution. These are massive structures of about 10^{16} solar masses consisting of several tens of thousands of galaxies grouped into tens of clusters. The Coma cluster is an example

of one of the largest superclusters with a size of roughly 3 million light years. The picture changed in 1989 when Margaret Geller and John Huchra discovered the 'great wall' in the redshift survey they were conducting (Geller & Huchra 1989). The 'great wall' is a gigantic structure 500 million light years long, 200 million wide, and 15 million light years thick.

Redshift surveys provide a 3-dimensional view of the Universe and show the presence of large scale structures called voids, sheets and filaments. The galaxies are found to be concentrated along sheets and filaments, with superclusters generally at the intersection of many filaments. The voids are mostly empty and are the dominant feature occupying about 90% of the Universe. Figure 1.1 shows a slice from a redshift survey, the Sloan Digital Sky Survey (SDSS, York et al. 2000), depicting the various features. Due to the virial motions of galaxies within clusters, the effect known as the 'finger-of-god' is caused by long lines of galaxies pointing towards the observer. Whereas at large scales, gravitation causes clusters to appear slightly flattened. These two effects are described in more detail in Chapter 3.

The dark matter distribution is seen to follow a very similar web-like structure. Detailed studies of dark matter have been carried out with the help of large scale N-body simulations. Currently, with the help of supercomputers, simulations of billions of particles (galaxies) are easily followed through cosmic time. This is accomplished by starting with an initial distribution of density fluctuations, and then following the redshift evolution of this distribution (e.g. the Millenium simulation; Springel et al. 2005). Through these studies we can hope to better understand the distribution of galaxies.

Figure 1.2 shows different stages of structure formation in an N-body simulation box at different redshifts. At high redshift the distribution of particles is quite uniform. With the passing of time, as redshift decreases, structure becomes increasingly more clumped and non-linear. The voids in turn become more and more empty. The appearance of the large scale structure resembles a spider web, hence the popular term 'cosmic web' (Bond et al. 1996).

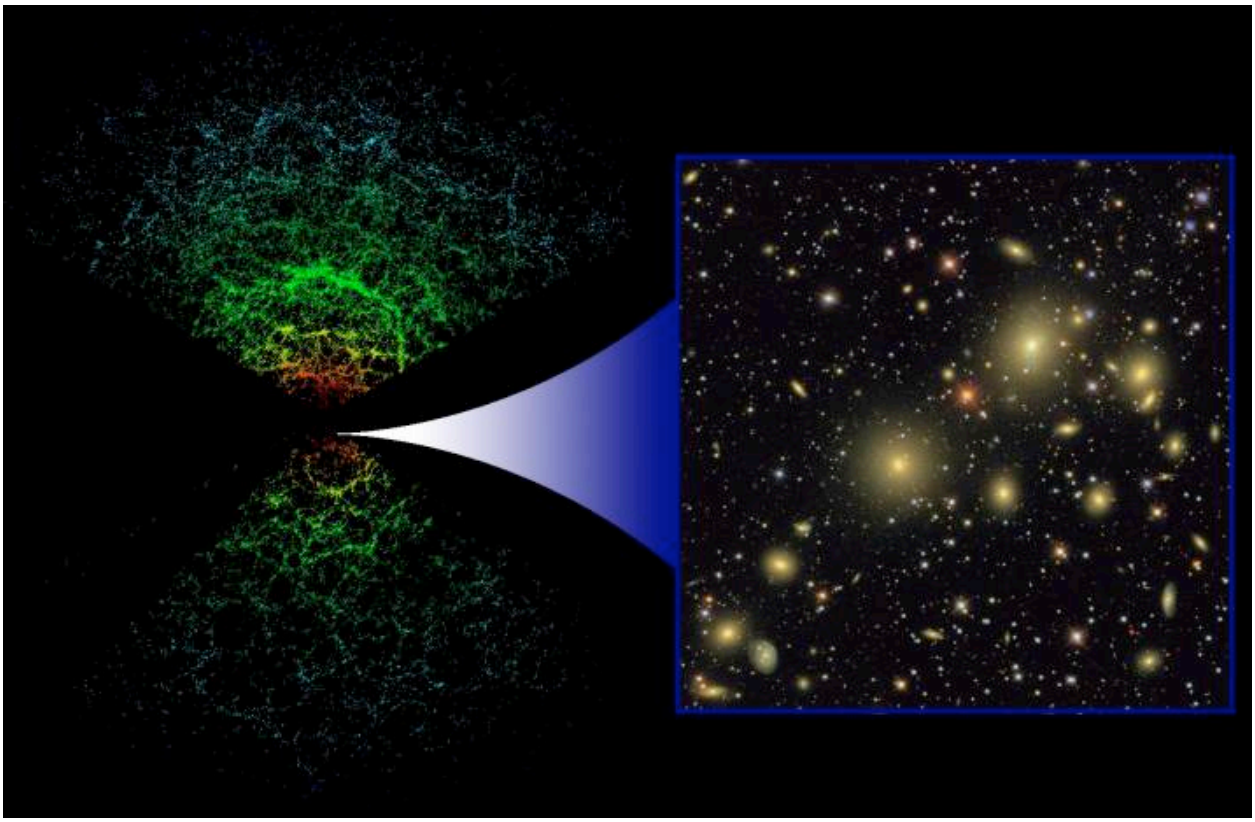

Figure 1.1: A redshift slice from the SDSS survey depicting approximately 67,000 galaxies that lie near the plane of Earth's equator. The 3D map at the left is 2 billion light years deep. Each point represents a galaxy with the color depicting its luminosity. The inset at the right shows the 2D image of galaxies used to create the 3D map. Image credit: SDSS.

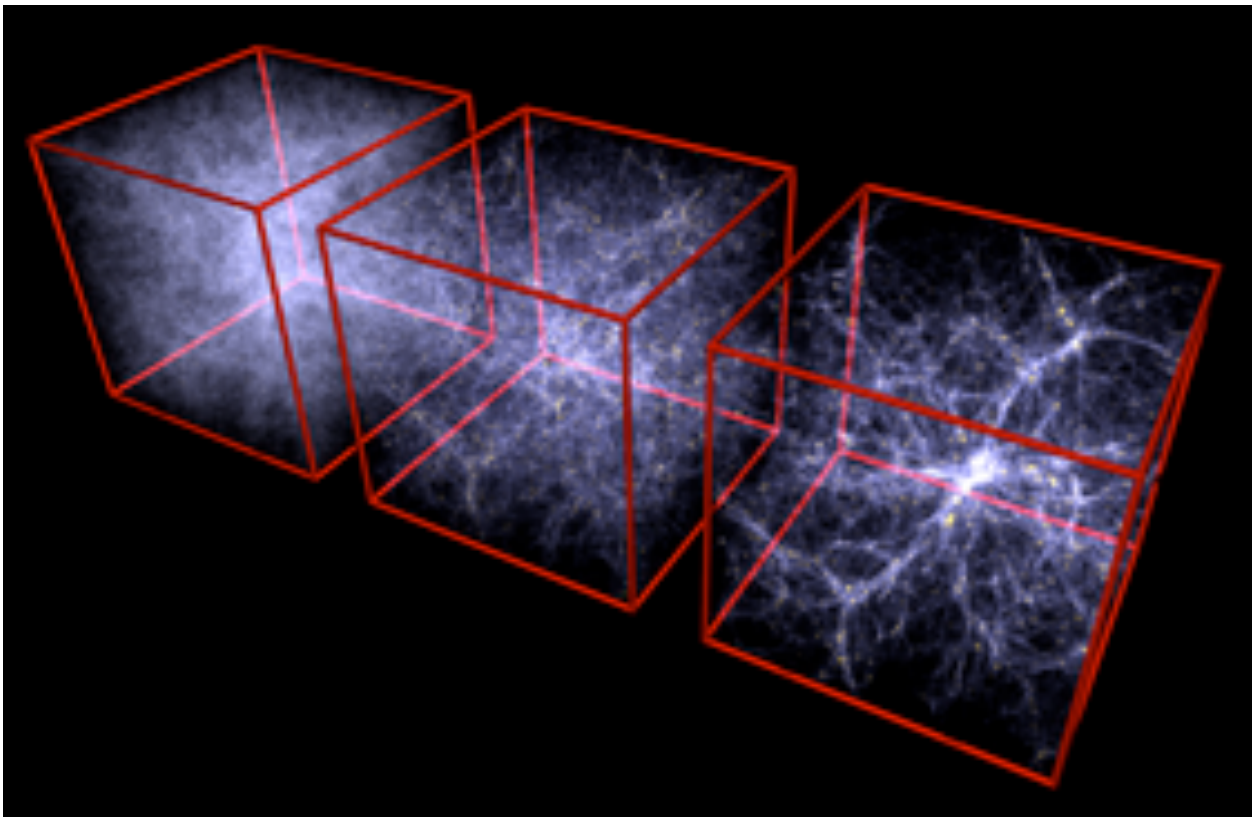

Figure 1.2: The development of nonlinear structure in an N-body simulation from z = 6 (leftmost cube) to z = 0 (right cube) through z = 2 (middle cube). Simulation box is 100 Mpc/h on a side.

Figure courtesy: Volker Springel.

http : //www.mpa − garching.mpg.de/galform/data_vis/index.shtml

1.2 THE CORRELATION FUNCTION

The spatial distribution of galaxies can be completely described by an n-point correlation function. Recent measurements of the present day density correlations and the CMB fluctuations are consistent with the gravitational instability theory: the structures observed today have evolved from infinitesimal, adiabatic, Gaussian, scale invariant density fluctuations generated by inflation. The statistical nature of such a Gaussian density field is completely specified by the two-point correlation function. The Gaussian nature of the density field is preserved for linear evolution, however this is not the case for non-linear evolution, as gravity can generate skewness and higher order moments.

Following the definition from Peebles (1980), given a random galaxy in a location, the correlation function describes the probability that another galaxy will be found within a given distance. The differential probability of finding two objects in volumes V_1 and V_2 separated by a distance r given that the number density is n can be written as

$$dP = n^2[1 + \xi(r)]dV_1 dV_2 \tag{1.1}$$

The n-point correlation function for a statistical ensemble is given by:

$$\xi(r) =< \delta(x)\delta(x + r)... > \tag{1.2}$$

where δ represents the density perturbation field and is written as:

$$\delta(x) = \frac{\rho(x) - < \rho >}{< \rho >} \tag{1.3}$$

For a Poisson distribution, $\xi = 0$, if the positions are correlated $\xi > 0$, and for anti-correlation, $-1 < \xi < 0$. Groth & Peebles (1977) using the Lick catalogue and later Baugh (1996) using APM catalogue, found that the galaxy correlation function at small scales can be approximated by a power law, $\xi(r) = (r/r_0)^{-\gamma}$ with the best fit values of $r_0 \approx 5h^{-1}Mpc$ and $\gamma \approx 1.8$. However, the correlation function of cold dark matter differs from the galaxy correlation function by being more complex with a non-power law behavior for the case of different cosmologies (Klypin, Primack, & Holtzman 1996; Peacock 1997; Jenkins et al. 1998).

Given a sample of data and random sets each containing n_d and n_r number of points respectively, the two-point correlation function can be obtained by using the simple estimator (Peebles 1980) given by;

$$\xi(r) = \frac{DD(r)/(n_d(n_d-1))}{RR(r)/(n_r(n_r-1))} - 1 \tag{1.4}$$

where DD(r) is the number of data-data pairs, and RR(r) is the number of random-random pairs all with separation r. This estimator is limited in its use in dealing with the interactions of the sample data with the boundaries of the sampling space. The random set behaves differently near the boundaries than the data set, which can be taken into account through the number of data-random pairs in the estimator. Instead, the current popular choice is the Landy-Szalay estimator (Landy & Szalay 1993);

$$\xi(r) = \frac{DD(r)/(n_d(n_d-1)) - 2DR(r)/(n_r n_d) + RR(r)/(n_r(n_r-1))}{RR(r)/(n_r(n_r-1))} \tag{1.5}$$

where once again DD(r) is the number of data-data pairs, DR(r) is the number of data-random pairs, and RR(r) is the number of random-random pairs all with separation r. This estimator has the advantage of having a nearly Poisson variance and is accurate even for small number counts.

Another popular choice used to describe the amount of clustering on a given length scale is the power spectrum, which is the fourier transform of the correlation function. The measured power spectrum from different galaxy surveys is shown in Figure 1.3.

Measurements of clustering give very useful constraints on cosmological parameters and models of galaxy formation. With the availability of large galaxy catalogs even more precise constraints can be obtained.

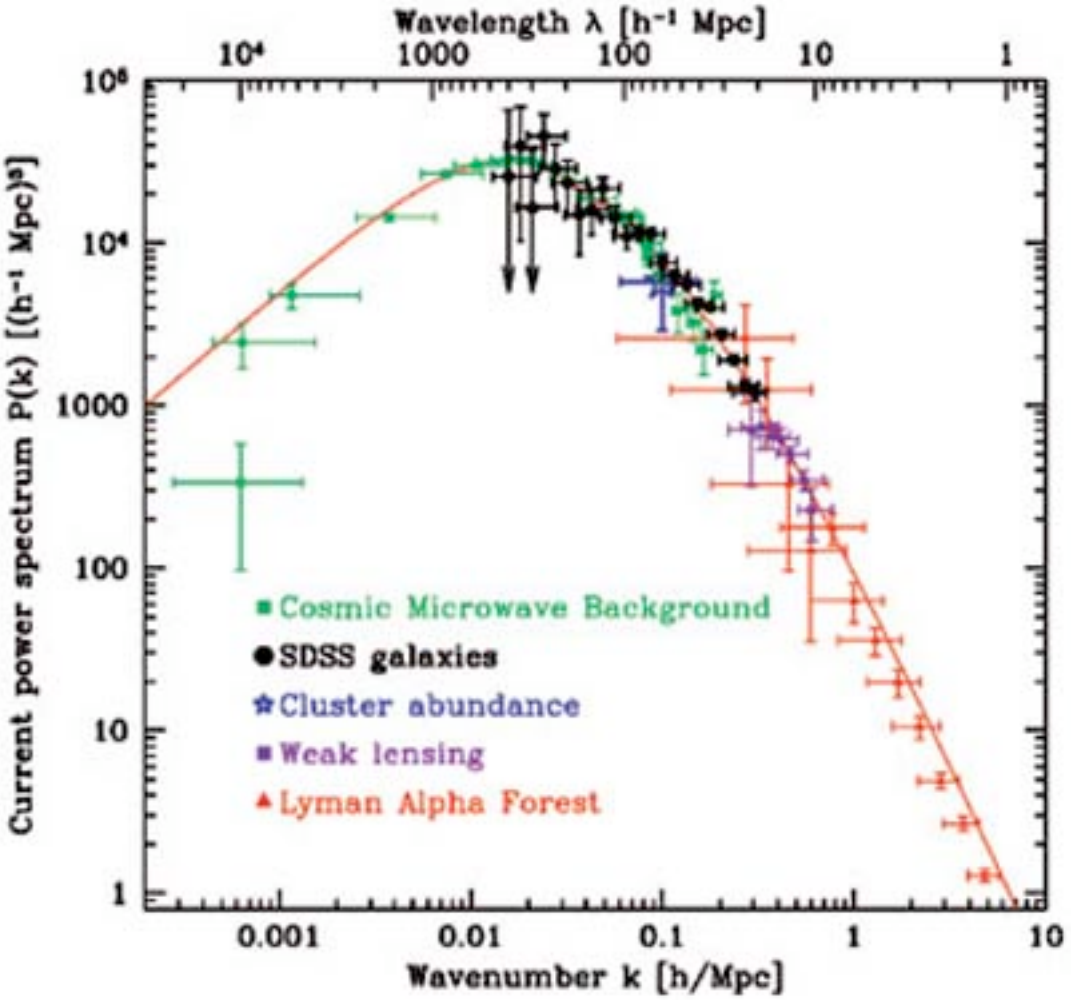

Figure 1.3: The measured power spectrum from different surveys (Tegmark et al. 2004). The negative slope for $\lambda < 400$ Mpc/h is caused by wavelengths entering the horizon before matter-radiation equality time, which translates into a suppression of the structure formation at those scales.

1.3 WHAT DO WE KNOW ABOUT ENVIRONMENTAL DEPENDENCE?

At very large scales, of the order of a few hundred Mpc, the distribution of dark matter and galaxies can be expected to be almost uniform and isotropic. As one decreases the scale, different samples having sufficiently small volumes (say a cube of 10 Mpc on side) can be expected to contain different spatial distributions of matter. The natural questions that arise are: what causes the differences seen between various samples? How does the large scale environment affect such a behavior? Is there an easy way to quantify this?

Previously, we had shown the distribution of dark matter in an N-body simulation box and the distribution of galaxies in redshift surveys. The striking resemblance leads one to believe that it can be possible to relate the two by taking into account an effect known as biasing. Thanks to the wealth of data, which is becoming richer by the day, statistical samples allow us to probe such effects to unprecedented accuracy.

A number of physical mechanisms are expected to play a role in determining the properties of a galaxy: e.g., ram pressure stripping (the removal of matter from galaxies moving through a cluster), harassment (high-speed galaxy-galaxy encounters), and strangulation (truncation of star-formation in cluster galaxies due to tidal heating of gas) (Gunn & Gott 1972; Farouki & Shapiro 1980; Moore et al. 1996; Balogh & Morris 2000). Many of these operate in dense environments. So the existence of a morphology–density relation—the fraction of galaxies which have elliptical rather than spiral morphologies is higher in denser regions (Dressler 1980)—is not unexpected. For similar reasons, recent measurements of lower star-formation rates in denser regions, which appears to persist even at fixed morphology (Balogh et al. 2002; Gomez et al. 2003) and stellar mass (Kauffmann et al. 2004), are not unexpected. Massive galaxies are seen to be more strongly clustered than less massive galaxies and in turn preferentially occupy denser regions. More luminous galaxies and red galaxies too show a tendency to be more clustered than fainter and bluer ones (Figure 1.4). However, determining which, if any, of the physical mechanisms mentioned above is the dominant one is more difficult.

This leads to the conclusion that different samples of galaxies are biased relative to one another. As galaxies represent the observable part of the universe, they should somehow be

related to the underlying dark matter distribution. Hence the visible matter on the whole should be biased relative to the dark matter. Understanding the nature of biasing is crucial in relating the known to the unobservable distribution. The complex shape of the matter correlation function can be quantified in terms of a scale-dependent and varying bias.

During the late 1970s and early 1980s it became clear that clusters of galaxies were more clustered than galaxies. Kaiser (1984) introduced the 'high-peaks' model to describe this effect in terms of the primordial density field. The density field can be decomposed into a complete set of density fluctuations of varying length. The high-peaks model claims that a rare high density fluctuation corresponds to a massive object with a higher-than-average correlation function. This model naturally explained the excess of massive objects in overdense regions and was successful at describing the strong clustering of elliptical galaxies as well as that of the Abell clusters.

Recently, hierarchical galaxy formation models have been rather successful at reproducing the morphology–density relation (e.g. Benson et al. 2001). In these models, a correlation between galaxy-type and environment arises even if none of the physical mechanisms mentioned above are present. The correlation is a consequence of the following assumptions. Gravity has transformed small fluctuations in the early Universe into the structures we see today. This transformation was hierarchical, in the sense that small virialized objects formed first, and then merged with one another to form more massive virialized objects at a later time. The virialized objects present at any given time, called dark matter haloes, are approximately 200 times denser than the background universe at the time (Gunn & Gott 1972). Galaxies form from gas which cools within virialized dark matter haloes (White & Rees 1978). The properties of a galaxy are determined entirely by the mass and formation history of the dark matter halo within which it formed (e.g. White & Frenk 1991; Kauffmann et al. 1997; Somerville & Primack 1999; Cole et al. 2000). Halo masses and formation histories are directly related to the structure of the initial density fluctuation field from which they formed (Press & Schechter 1974; Lacey & Cole 1993; Sheth et al. 2001). In hierarchical models, there is a correlation between fluctuations on different scales, and this induces correlations between halo mass and/or formation and the larger scale environment of a halo (Mo & White 1996; Lemson & Kauffmann 1999; Sheth & Tormen 2002). This, in turn,

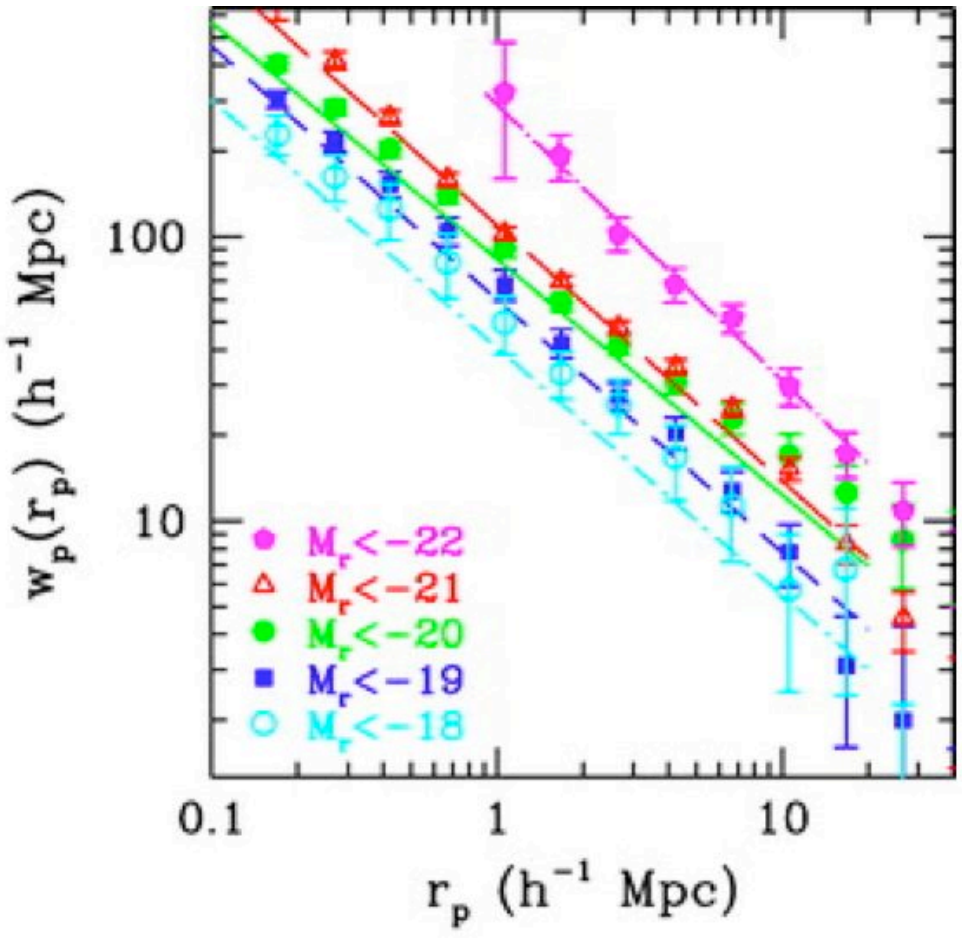

Figure 1.4: The luminosity dependence of galaxy clustering seen in the SDSS. Image has been taken for illustrative purposes from Zehavi et al. (2005). The different samples correspond to galaxies brighter than the indicated absolute magnitudes and show that the projected correlation function of bright galaxies is higher.

induces a correlation between galaxy-type and environment, which is entirely *statistical* in nature. So it is interesting to ask if this statistical correlation is sufficient to explain most of the observed correlation between galaxy-type and environment. The main goal of this thesis is to address the aforementioned issue.

1.4 THE HALO MODEL

The analytical framework in this thesis is based on the highly successful and increasingly popular halo model. This model allows us to quantify the correlation function and power spectrum of dark matter and galaxies in a simple and consistent fashion. It easily bypasses the route of following simulations or semi-analytic calculations that require descriptions of complex physics, such as feedback from supernovae, harassment etc. All the complications of galaxy formation are encoded via the halo occupation number, which is the number of galaxies found above some luminosity threshold in a virialized halo of a given mass. In a nut-shell, the halo model describes the non-linear structures as virialized halos of different mass, placing them in space according to the large scale density field.

The use of an analytical framework to describe the distribution of galaxies dates back to the 1950's. Neyman & Scott (1952) were the first to attempt an analytical description of the clustering of galaxies as a realization of the random distribution of galaxy clusters. Further developments took place in the 70's and 80's with papers by Peebles (1974), McClelland & Silk (1978), and, Scargle (1981). Due to the then lack of statistical samples of data, the work was limited in its scope of prediction. During the 90's modern versions of the halo model were formulated and further refined by Scherrer & Bertschinger (1991), Sheth & Jain (1997) and Sheth & Lemson (1999). Recent papers dealing with the halo model include Seljak (2000), Ma & Fry (2000), Peacock & Smith (2000), Scoccimarro et al. (2000), Berlind & Weinberg (2002), Sheth & Jain (2003), Kravtsov et al. (2004), Sheth (2005), Abbas & Sheth (2005), Smith et al. (2006) etc. This is just a small subset of the papers that have come out in recent years and is an indication of the burst of activity in this field. A very nice review of the halo model is in Cooray & Sheth (2002).

The distribution of dark matter can be described in terms of a distribution of dark haloes

that have a mass function and density profile following that of the dark matter. As such, dark matter simulations of structure formation suggest the halo model. Within the analytical framework of the halo model the statistics can be described at two scales, non-linear small scales and large linear scales. In this model, all mass is bound up in dark matter haloes which have a range of masses. Hence, the background density is

$$\bar{\rho} = \int dm \, \frac{dn(m)}{dm} \, m. \tag{1.6}$$

where $dn(m)/dm$ denotes the number density of haloes of mass m.

The correlation function is the Fourier transform of the power spectrum $P(k)$:

$$\begin{aligned} \xi(r) &= \int \frac{d^3k}{(2\pi)^3} \, P(k) e^{i\,\vec{k}.\vec{r}} \\ &= \frac{1}{(2\pi)^3} \int_0^\infty \frac{dk}{k} k^3 P(k) \int_{-1}^{1} d(\cos\theta) e^{ikr\cos\theta} \int_0^{2\pi} d\phi \\ &= \int \frac{dk}{k} \, \frac{k^3 P(k)}{2\pi^2} \, \frac{\sin kr}{kr}. \end{aligned} \tag{1.7}$$

In the halo model, $P(k)$ is written as the sum of two terms: one that arises from particles within the same halo and dominates on small scales (the 1-halo term), and the other from particles in different haloes which dominates on larger scales (the 2-halo term). At small scales the density profile of the haloes becomes important, whereas the spatial distribution of haloes does not affect the particles within the halo. On the other hand, at large scales the distribution of haloes is more crucial than the density profile. This implies that the spatial statistics can be studied in two steps and is the underlying basis for the halo model. As such,

$$P(k) = P_{1h}(k) + P_{2h}(k), \tag{1.8}$$

where

$$\begin{aligned} P_{1h}(k) &= \int dm \, \frac{dn(m)}{dm} \, \frac{m^2}{\bar{\rho}^2} \, |u(k|m)|^2, \\ P_{2h}(k) &= \left[\int dm \, \frac{dn(m)}{dm} \, \frac{m}{\bar{\rho}} \, b(m) \, u(k|m) \right]^2 P_{\rm Lin}(k). \end{aligned} \tag{1.9}$$

Here $u(k|m)$ is the Fourier transform of the halo density profile, $b(m)$ is the bias factor which describes the strength of halo clustering, and $P_{Lin}(k)$ is the power spectrum of the

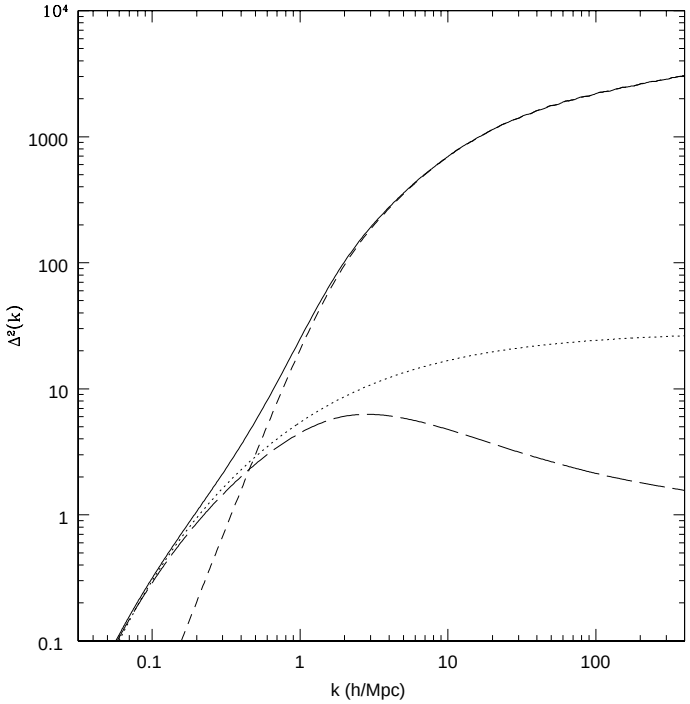

Figure 1.5: The different contributions to the power spectrum as predicted by the halo model. The dashed, long dashed and dotted curves show the 1 halo, 2 halo and linear terms respectively. The solid curve shows the full power spectrum. Here the y-axis denotes the dimensionless quantity given by the power per logarithmic interval in wavenumber.

mass in linear theory. This is depicted in Figure 1.5. The dependence of the bias factor on halo mass is shown in figure 1.6, which highlights the trend of high mass haloes with strong biasing and the anti-biasing of low mass haloes.

When explicit calculations are made, we assume that the density profiles of haloes have the form described by Navarro et al. (1997) shown in figure 1.7, and that halo abundances and clustering are described by the parameterization of Sheth & Tormen (1999):

$$\begin{aligned}
\frac{m}{\bar{\rho}}\frac{dn(m)}{dm}\,\mathrm{d}m &= f(m)\,\mathrm{d}m = f(\nu)\,\mathrm{d}\nu \\
&= \frac{\mathrm{d}\nu^2}{\nu^2}\sqrt{\frac{a\nu^2}{2\pi}}\,\exp\left(-\frac{a\nu^2}{2}\right) A\,\left[1+(a\nu^2)^{-p}\right] \\
b(m) &\approx 1 - \frac{d\ln dn(m)/dm}{d\delta_c} \\
&= 1 + \frac{a\nu^2-1}{\delta_{\rm sc}} + \frac{2p/\delta_{\rm sc}}{1+(a\nu^2)^p} \\
\nu &= \frac{\delta_{\rm sc}}{\sigma(m)} \quad \text{and} \\
\sigma^2(m) &= \int_0^\infty \frac{\mathrm{d}k}{k}\,\frac{k^3 P_{\rm Lin}(k)}{2\pi^2}\,W^2(kR_0),
\end{aligned} \tag{1.10}$$

where $W(x) = (3/x^3)[\sin(x) - x\cos(x)]$ and $R_0 = (3M/4\pi\bar{\rho})^{1/3}$. That is to say, $\sigma(m)$ is the rms value of the initial fluctuation field when it is smoothed with a tophat filter of comoving size R_0, extrapolated using linear theory to the present time. Here $\delta_{\rm sc}$ is the critical density required for spherical collapse, extrapolated to the present time using linear theory (it is 1.686 for an Einstein de-Sitter model), and $a \approx 0.71$, $p = 0.3$ and $A = (1+\Gamma(1/2-p)/\sqrt{\pi}/2^p)^{-1} \approx 0.322$. If $a = 1$, $p = 0$ and $A = 1/2$, then $\mathrm{d}n/\mathrm{d}m$ is the same as the universal mass function first written down by Press & Schechter (1974). The mass function is shown in Figure 1.8.

We now discuss how the model above can be extended to describe galaxy clustering. The mean number density of galaxies $\bar{n}_{gal}$ is given by replacing the weighting by m in Equation (1.6) for the mean density by $\langle N_{gal}|m\rangle$, the mean number of galaxies in an m-halo. Similarly, the weighting by $m/\bar{\rho}$ in $P_{2h}(k)$ is replaced with a weighting by $\langle N_{gal}|m\rangle/\bar{n}_{gal}$. And the weighting by $(m/\bar{\rho})^2$ in the 1-halo term becomes $[2\langle N_s|m\rangle\,u(k|m) + \langle N_s|m\rangle^2\,|u(k|m)|^2]/n_{gal}^2$. This weighting assumes there is always one galaxy at the centre, and that the number of satellite galaxies in an m-halo follows a Poisson distribution with mean $\langle N_s|m\rangle$ and is motivated by the halo occupation distribution model (HOD) of Kravtsov et al. (2004).

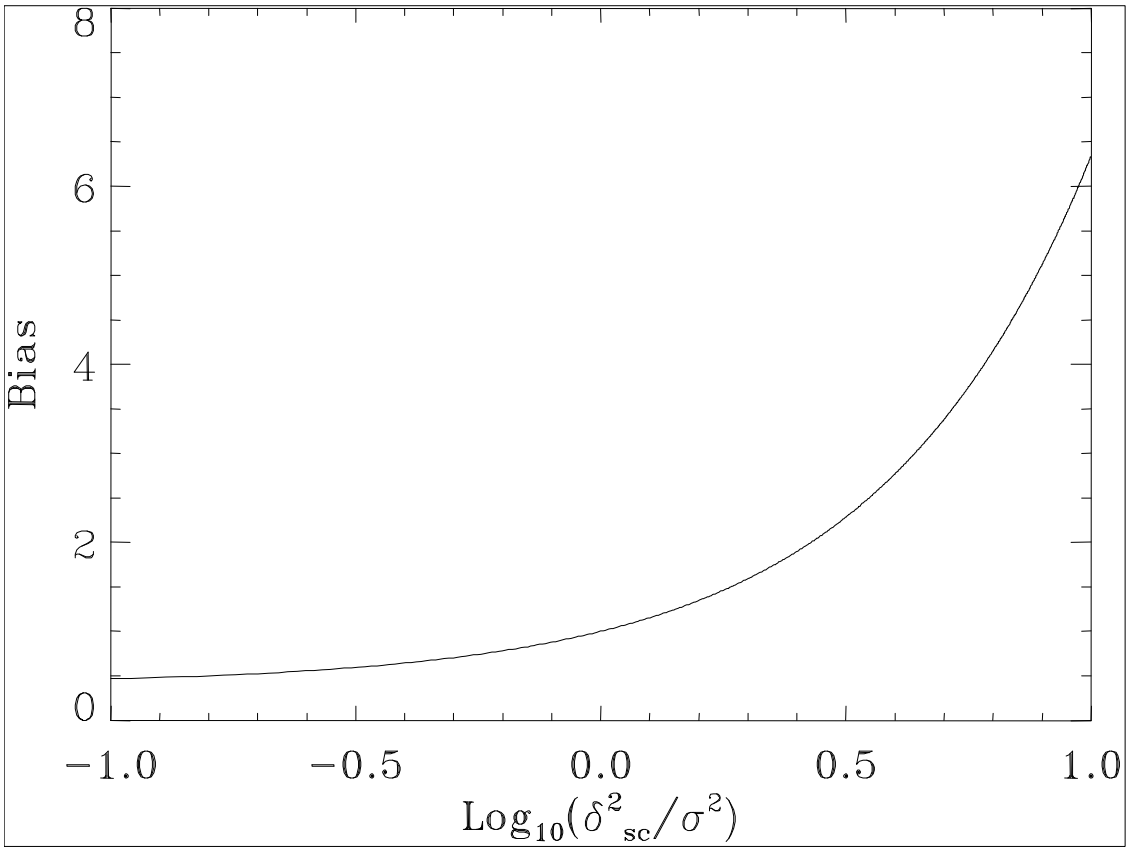

Figure 1.6: Graphical representation of the mass dependence of the bias factor where the x-axis scales as the mass of the halo (Equation 1.10). Low mass haloes have a bias less than 1, and are defined as anti-biased. High mass haloes have a positive bias greater than 1, and in turn are defined as biased.

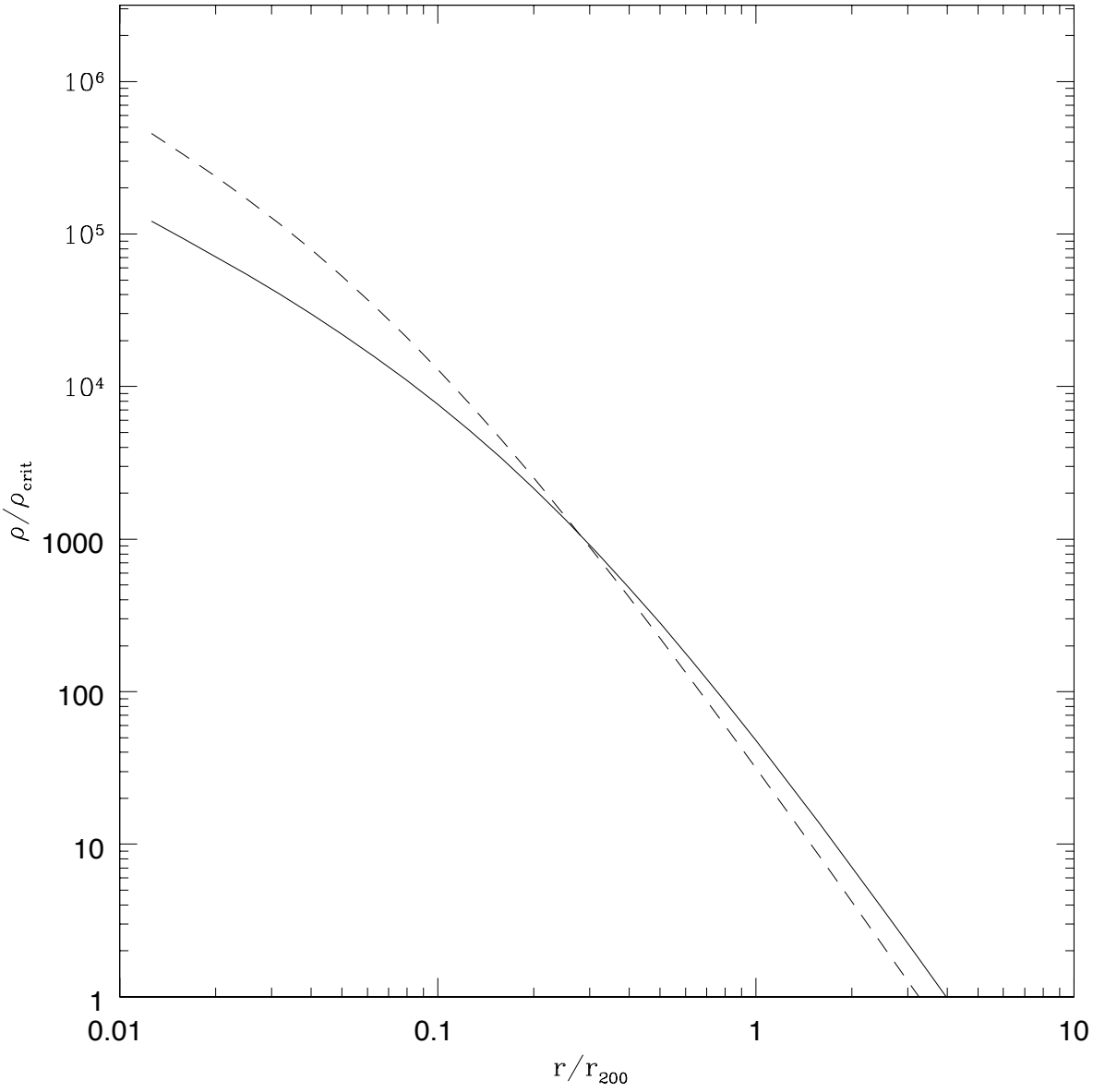

Figure 1.7: The NFW density profile of dark matter haloes seen in dark matter simulations. The fits to the density profiles have been scaled to the virial radius (r_{200}) and to the critical density of the universe at $z = 0$ (ρ_{crit}). The dashed and solid lines represent the low- and high-mass systems respectively.

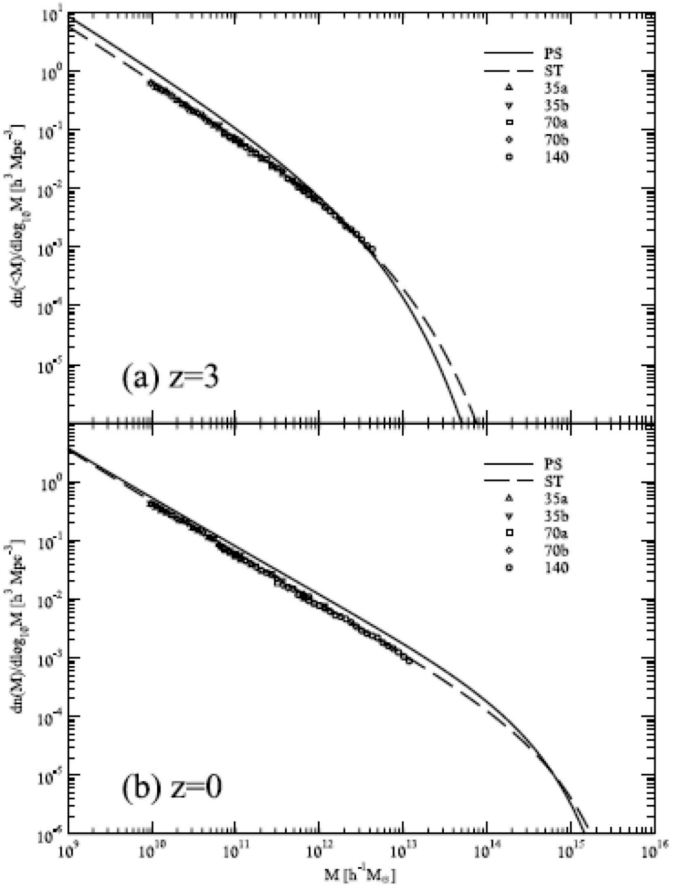

Figure 1.8: The mass function of dark matter haloes for different redshifts. The symbols represent measurements from N-body simulations, which are well fitted by the Sheth-Tormen mass function (dashed line), whereas the Press-Schechter mass function (solid line) is slightly different. Figure has been taken from Yahagi et al. (2004).

The standard halo model assumes that the density profiles and $\langle N_{gal}|m\rangle$ are independent of environment. This thesis will discuss the consequences of the assumption that *all* environmental effects come *only* from the dependence of $dn(m)/dm$ on environment. The present analysis is intended to complement traditional measurements of the environmental dependence of clustering which tend to focus on correlations between the distribution of an observable (e.g., the luminosity function, or the star formation rate, etc.) and the environment (see e.g. Yang et al. 2003; Mo et al. 2004; Kauffmann et al. 2004; Balogh et al. 2004; Blanton et al. 2004, for some recent analyses). This is particularly interesting given recent work which suggests that, at fixed mass, haloes in dense regions form at higher redshifts (Sheth & Tormen 2004), and that this effect is a decreasing function of halo mass (Gao, Springel & White 2005).

1.5 CHAPTER SUMMARIES

Chapter 2 studies the dependence of clustering of dark matter and galaxies in simulations and mock catalogs. The analytical framework is developed for the first time based on the halo model. Comparisons between the analytical model and simulations are illustrated with the help of plots showing the correlation function in real space. This has been taken from a paper with myself as the leading author, published in MNRAS (Abbas & Sheth 2005) and is the backbone of the thesis. I have generated the mock galaxy samples, undertaken numerical analyses of the simulations, and contributed to the definition and implementation of the analytical model.

Chapter 3 shows the dependence of galaxy clustering on environment for mock and real data samples. The analytical model (described in the previous chapter) now incorporates redshift space effects to successfully describe quantitatively what is seen in observational data. This has been taken from a paper with myself as the leading author, published in MNRAS (Abbas & Sheth 2006). I have undertaken numerical analyses of the SDSS data samples, and corresponding mock samples, and contributed to the definition and implementation of the analytical model, taking into account redshift space distortions and projection effects.

Chapter 4 describes a test to find out if galaxy clustering can provide constraints on σ_8, the amplitude of fluctuations. We find that the model provides useful constraints.

Chapter 5 wraps up the thesis with a summary and discusses future prospects.

2.0 THE ENVIRONMENTAL DEPENDENCE OF CLUSTERING IN HIERARCHICAL MODELS

2.1 INTRODUCTION

The main goal of this chapter is to show how the analysis of the clustering of galaxies can be used to quantify the strength of the statistical correlation between galaxy properties and environment through the dependence of the halo mass function on environment. The idea is to show the strength of the statistical effect alone: any discrepancy with observations can then be ascribed to other physical processes (Sheth, Abbas & Skibba 2004). Although our argument is general, we focus in particular on the two-point correlation function, $\xi(r)$, and show how it is expected to depend on environment if the only environmental effects on galaxy properties are those which arise from the statistics of the initial fluctuation field.

This chapter is arranged as follows. Section 2.2 uses numerical simulations to illustrate how clustering is expected to depend on environment if the entire environmental dependence arises from the correlation between haloes and their environments. Section 2.2.1 shows the effect for dark matter. A simple toy model which captures most of the relevant features of this effect is presented in Section 2.2.2. Section 2.2.3 uses mock galaxy samples to show that the effect should be easily measured in surveys such as the SDSS. Section 2.3 shows how to describe the effect using the language of the halo model of large scale structure (given in Section 1.4). The final section summarizes our results and discusses some implications.

2.2 ENVIRONMENTAL DEPENDENCE OF ξ

Throughout, we show results for a flat ΛCDM model for which $(\Omega_0, h, \sigma_8) = (0.3, 0.7, 0.9)$ at $z = 0$. Here Ω_0 is the density in units of critical density today, the present Hubble constant is $H_0 = 100h$ km s^{-1} Mpc^{-1}, and σ_8 describes the rms fluctuations of the initial field, evolved to the present time using linear theory, when smoothed with a tophat filter of radius $8h^{-1}$ Mpc. The GIF and VLS numerical simulations we use to illustrate our arguments were made available to the public by the Virgo consortium. Both were run with the same ΛCDM cosmology, but with slightly different parameterizations of the initial fluctuation spectrum. The GIF simulation had 256^3 particles in a cubic box with sides $L = 141h^{-1}$Mpc. The VLS simulation had 512^3 particles in a cubic box with sides $L = 479h^{-1}$Mpc.

2.2.1 Dark matter simulations

To illustrate how clustering depends on environment, we used the halo and particle distributions in the ΛCDM GIF simulations (Kauffmann et al. 1999). The simulations contain approximately 90,000 haloes which each have at least ten particles, where the virialized halos were found using a friends-of-friends group finding program. We defined the environment of a halo by using the mass M_R within a sphere of radius R centred on the halo. We set $R = 5h^{-1}$ or $8h^{-1}$Mpc, but any value which is substantially larger than the typical virial radius (a few hundred kpc), but smaller than the scale on which the Universe is essentially homogeneous, will do. We then ranked all the haloes in decreasing order of M_R. The particles belonging to the top one-third of the haloes were labeled as belonging to densest environments, and the particles in the bottom one-third of the halo sample were labeled as belonging to the least dense environments. The dot plots for these two samples are shown in Figure 2.1. Finally, we computed the correlation function of particles belonging to the haloes in the densest and least dense regions using the Landy-Szalay estimator (Equation 1.5). The results are shown in Figure 2.2.

There are obvious differences between the two correlation functions. The correlation function for the particles in dense regions extends to larger scales; if fit to a power-law, it

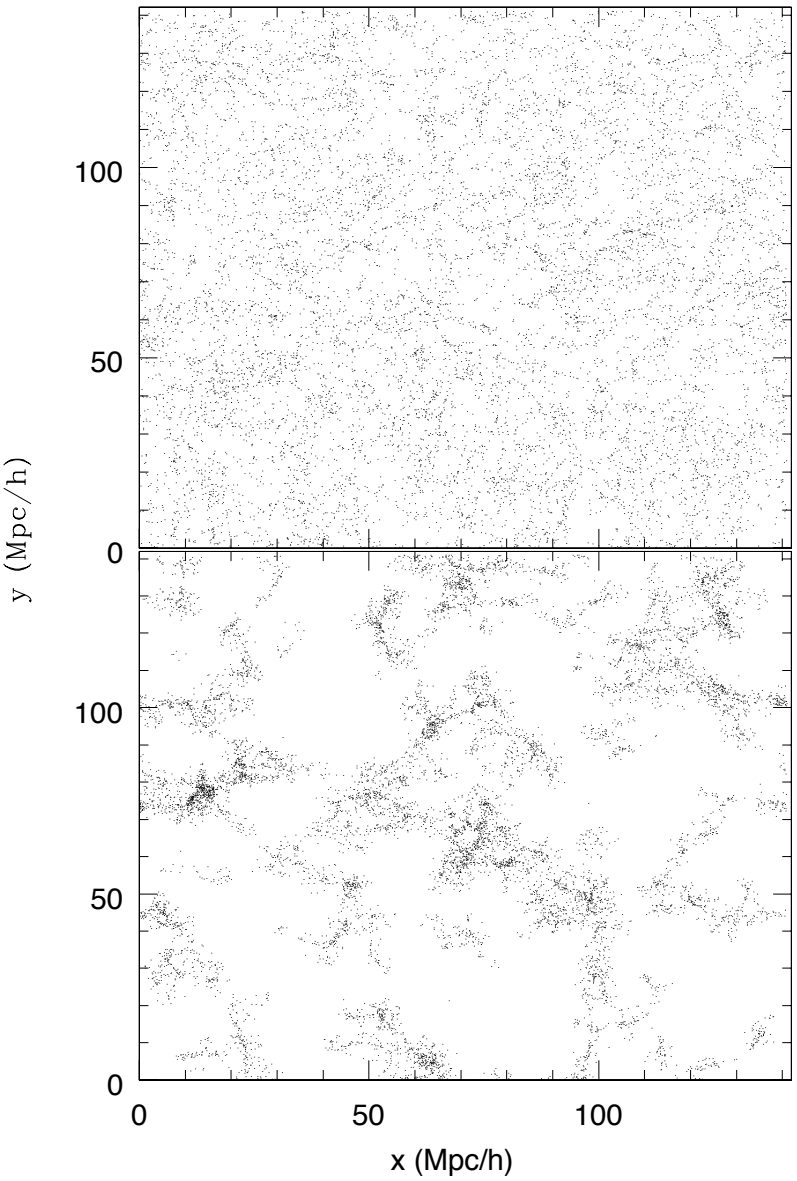

Figure 2.1: Dot plots of the dark matter particles in the GIF simulation, where the x vs y coordinate for the particles are plotted with $z < 50$. The top panel shows the one-third underdense sample, and lower panel shows the dense sample with the same number of points. The dark matter particles in the underdense sample fill the void regions which are present in the dense sample.

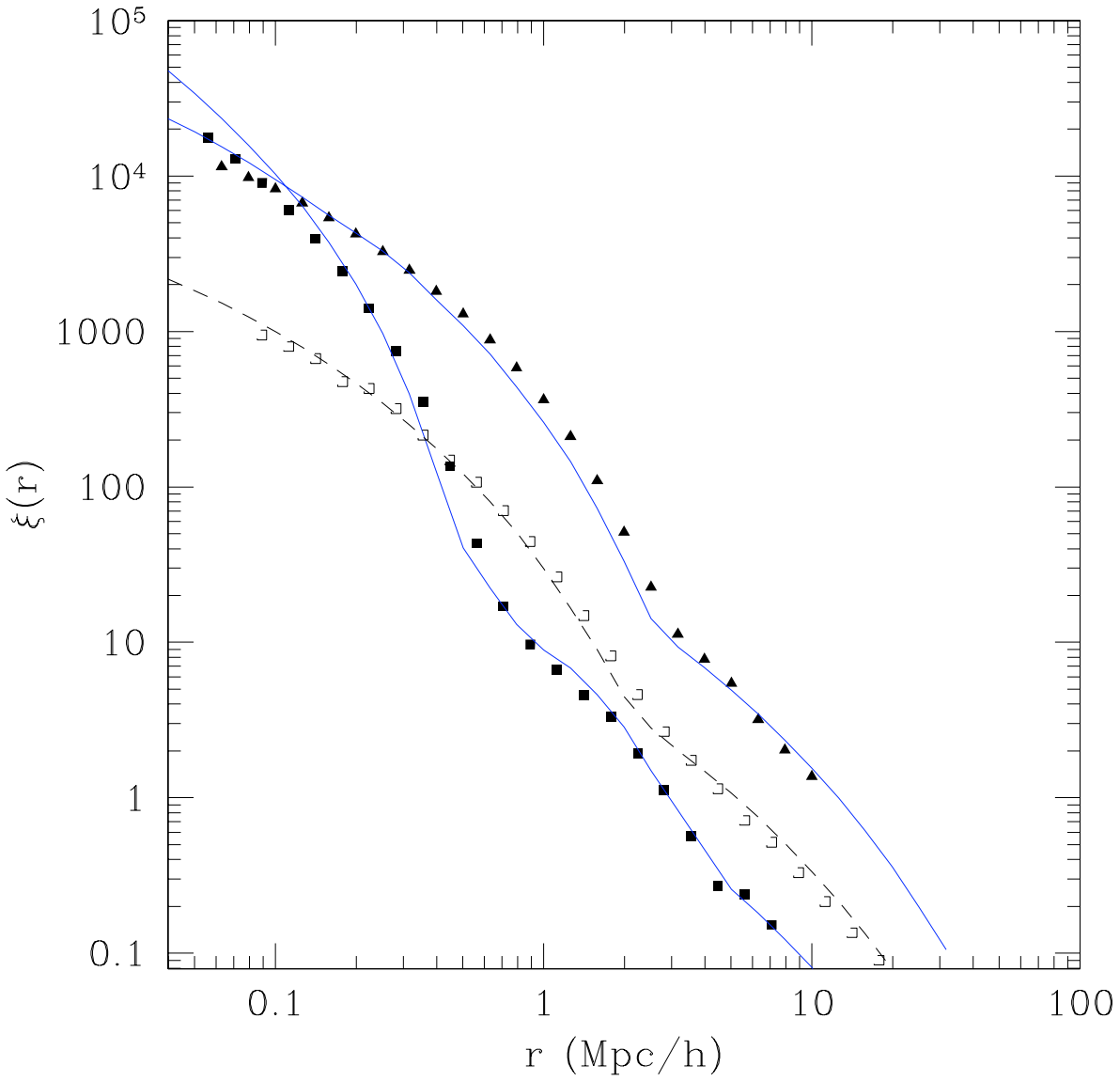

Figure 2.2: Environmental dependence of the dark matter correlation function in the ΛCDM GIF simulation. Triangles show $\xi(r)$ of particles in haloes which were defined as having the densest environments (defined by counting the mass within a sphere of radius of $5h^{-1}$Mpc centred on each halo), and squares are from particles in underdense regions. Smooth curves show the analytic model for this environmental effect that is developed in Section 2.3. The dashed curve and corresponding symbols represent the correlation function of the full sample.

would have a shallower slope. The next section describes a simple model for these differences. The smooth curves in the figure show the result of a more complete analytic model that is developed in Section 2.3.

2.2.2 A toy model

Let $dn(m,\delta_c)/dm$ denote the number density of dark matter haloes with mass m at a time when the linear theory overdensity required for spherical collapse is δ_c, and let $N(m,\delta_c|M,V)$ denote the average number of m haloes in regions of volume V which contain mass M. Define $M/V \equiv \bar{\rho}(1+\delta)$, where $\bar{\rho}$ denotes the average density of the background. Dense regions have $\delta > 0$. Mo & White (1996) showed that a generic feature of hierarchical models is that $N(m,\delta_c|M,V) \neq (1+\delta)\,V\,dn/dm$: i.e., dark halo abundances in dense and underdense regions do not differ by a simple factor which accounts for the difference in density. Rather,

$$N(m,\delta_c|M,V) \approx \Big[1+b(m)\delta\Big]V\frac{dn(m,\delta_c)}{dm}, \tag{2.1}$$

where

$$b(m,\delta_c) \approx 1 - \frac{d\ln dn(m,\delta_c)/dm}{d\delta_c} \tag{2.2}$$

is a function which typically increases monotonically with m (e.g. Sheth & Tormen 1999) and is shown in Figure 1.6. As a result, one expects the ratio of the number of massive to low mass haloes to be larger in dense regions than in less dense regions: the mass function in overdense regions should be 'top-heavy'. Measurements in simulations indicate that this is indeed the case (e.g. Lemson & Kauffmann 1999; Sheth & Tormen 2002): the average halo mass increases with δ as shown in Figure 2.3.

In the GIF simulations, the average mass of the $\sim 30,000$ haloes which reside in the densest regions is approximately $m = 2.5\times10^{12}h^{-1}M_\odot$, whereas the average mass of the $\sim 30,000$ haloes which reside in the least dense regions is $m = 5.2\times10^{11}h^{-1}M_\odot$. (Hence, the two overdensities differed by a factor of approximately 5.)

The fact that dense regions have a top-heavy mass function has an important consequence for the environmental dependence of the correlation function which we now describe. Let $\xi(r|\delta)$ denote the shape of the correlation function in regions with overdensity δ. Although

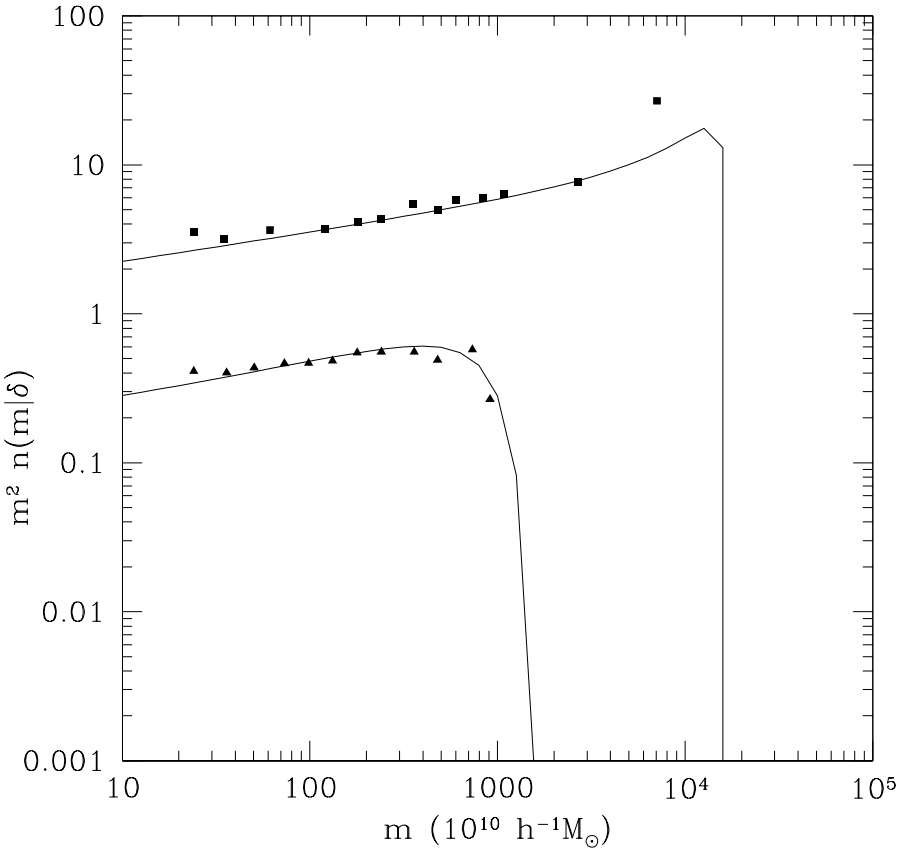

Figure 2.3: Mass function for the underdense and dense samples shown in Figure 2.2. The curve showing the analytic model and symbols showing the corresponding measurements have been offset by a factor of 10 for the top (dense sample) curve, whereas the underdense sample has not been shifted. It can be seen that massive haloes occupy the densest sample.

such regions contain haloes with a range of masses, suppose we require that all haloes have the same mass m_δ, chosen to match, say, the mean halo mass in the regions. If these haloes are not clustered, then $\xi(r|\delta)$ is simply a consequence of the shape of the density profiles $\rho(r|\delta)$ around haloes (Peebles 1974):

$$\xi(r|\delta) = n_{clus}(m_\delta) \int ds \, \frac{\rho(\mathrm{s}|m_\delta,\delta)}{\bar{\rho}_\delta} \frac{\rho(\mathrm{s}+\mathrm{r}|m_\delta,\delta)}{\bar{\rho}_\delta}, \tag{2.3}$$

where $n_{clus}(m_\delta)$ is the average number density of haloes surrounded by regions with overdensity δ, $\bar{\rho}_\delta \equiv m_\delta n_{clus}$, and $m_\delta \equiv \int ds\, \rho(\mathrm{s}|m_\delta,\delta)$. In fact, the haloes are clustered, but Sheth & Jain (1997) show why ignoring halo clustering should be a good approximation on small scales. Hence, to estimate the small scale correlations as a function of environment, we require an estimate of the shapes of halo density profiles.

When fit to spherical models, the density profiles of haloes in simulations are well fit by the functional form $\rho(r|m)/\bar{\rho} = \Delta_c(r_{vir}/cr)/(1+cr/r_{vir})^2$ where $r_{vir} \equiv (3m/4\pi\Delta_{vir}\bar{\rho})^{1/3}$ and the concentration $c \equiv 9\,(10^{13}h^{-1}M_\odot/m)^{0.1}$ (Navarro et al. 1997). If we assume that the density profiles of m haloes are the same in all environments (we will modify this assumption shortly), then $\xi(r|\delta)$ is given by inserting this expression for the density profile into the convolution integral above. The result is

$$\xi(r|\delta) = \frac{\Delta_{vir}}{\bar{\rho}_\delta/\bar{\rho}}\,\lambda(r|m_\delta,\delta), \qquad \text{provided} \quad r \le 2r_{vir}, \tag{2.4}$$

where λ is a messy function of c and cr/r_{vir} (given in Sheth et al. 2001). Since Δ_{vir} is independent of m, in such a model, the environmental dependence of $\xi(r|\delta)$ comes entirely from the fact that dense regions host the more massive haloes, and halo density profiles depend on halo mass. The factor of $\bar{\rho}_\delta/\bar{\rho}$ in the denominator derives from the fact that haloes which have a fixed overdensity relative to the global background density $\bar{\rho}$ have a different overdensity relative to the local background $\bar{\rho}_\delta$.

Figure 2.4 shows the result of this simple analytic estimate of $\xi(r|\delta)$ for the two sets of GIF simulation particles: those which reside in the 30,000 haloes with the largest surrounding overdensities (as described previously), and those which reside in the 30,000 haloes with the smallest surrounding overdensities The curves are qualitatively similar to the measurements shown in Figure 2.2, at least out to scales on which the measurements show an inflection:

ξ falls to zero on smaller scales in the less dense regions. The inflections at $\sim 0.8h^{-1}$Mpc and $\sim 3h^{-1}$Mpc in the simulations (Figure 2.2) denote the scales which are approximately twice the virial radii of the typical haloes in the two regions. Beyond this scale, halo-halo correlations become important; we build a model for this in Section 2.3.

The curves which extend to larger scales are those for particles in the denser regions. This is easily understood, because dense regions are those for which m_δ is larger, and hence r_{vir} is larger. The reason why $\xi(r|\delta)$ on small scales is larger for the less dense regions is more subtle.

The solid curves show results in which we have set $c = 10$ and ignored the mass dependence of c, and dashed curves include the mass dependence but assume that there is no additional dependence on environment. Clearly, the mass dependence of c is not a dominant effect even on scales smaller than r_{vir}/c. Thus, the difference in amplitudes on small scales derives from the factor of $\bar{\rho}_\delta/\bar{\rho}$ in the expression above, and not from the mass dependence of c. In the present example, the number of haloes in the two (dense and underdense) regions is the same, but m_δ in dense regions is larger, so $\bar{\rho}_\delta/\bar{\rho}$ is larger for the dense regions. Since the shape of ξ is approximately constant on small scales, it has a lower amplitude in denser regions.

The relative unimportance of c has the following interesting consequence. Suppose that halo density profiles do depend on environment (numerical simulations are only just reaching the resolution required to address this question). A simple way of parameterizing this dependence is to allow c to depend both on m and δ. If c depends only weakly on δ, then the effect on ξ will only be noticeable on scales smaller than $r_{vir}/\bar{c}$, where $\bar{c}$ denotes the mean c (averaged over environments).

On the smaller scales, where halo-halo correlations are not important, the differences between the two curves in Figure 2.2 are qualitatively like those of the simple toy model described above, indicating that our use of a mean density-dependent halo mass m_δ, rather than a distribution of masses, does capture the essential features of the density dependence of $\xi(r|\delta)$. On larger scales, where halo correlations are important, Figure 2.2 shows that $\xi(r|\delta)$ is stronger in dense regions. This is not unexpected in the context of the linear peaks-bias model of Kaiser (1984), if, on average, the densest regions at the present time

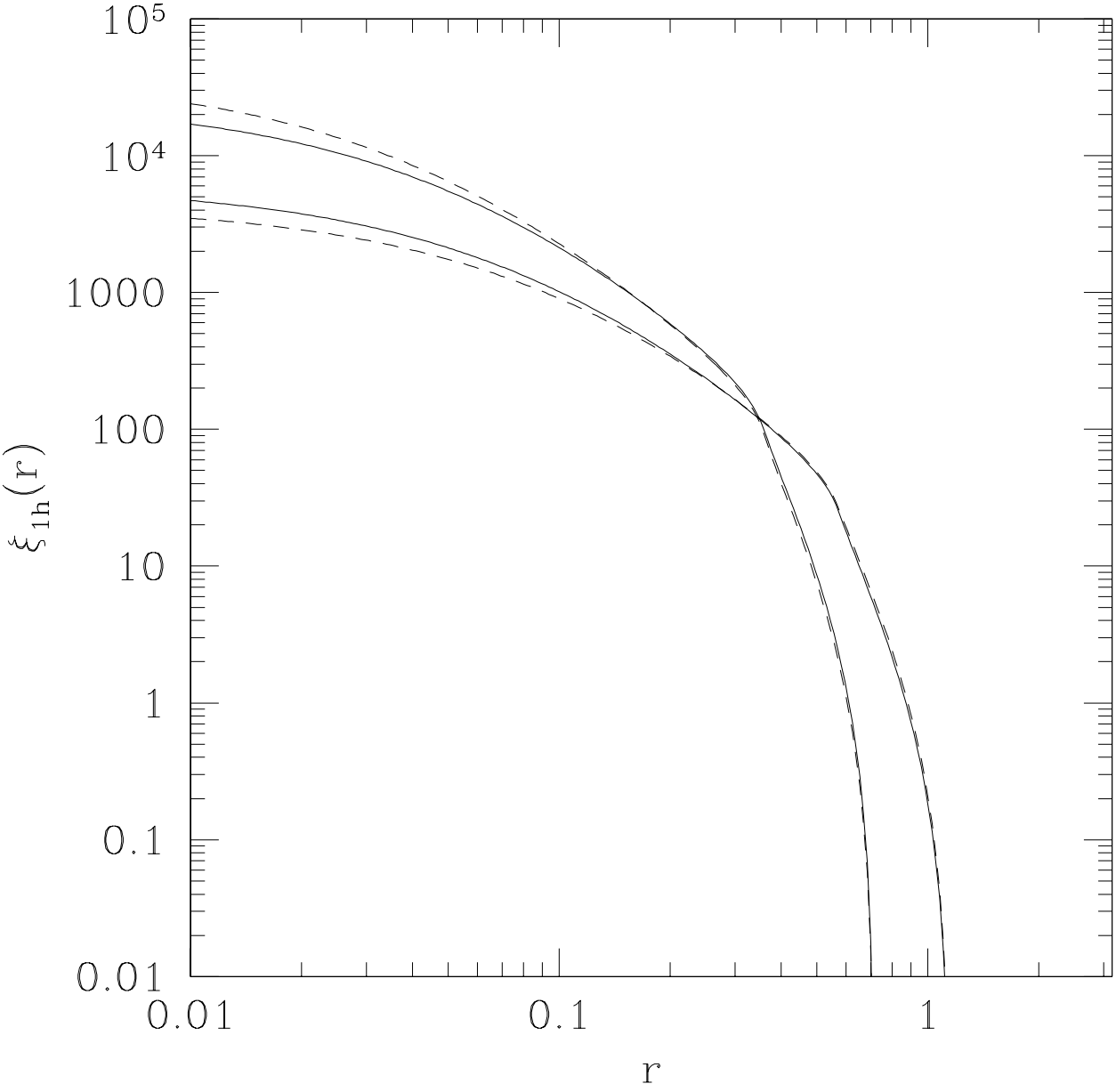

Figure 2.4: The correlation function of particles in haloes which are surrounded by dense regions (curves which extend to larger r), and by less dense regions (curves which fall to zero at smaller r), in our toy model. The number of haloes in the two environments was assumed to be the same, but the halo mass in dense regions was assumed to be larger by a factor of 5. Solid curves have $c = 10$ in both cases, whereas dashed curves include a realistic prescription for the weak mass dependence of c.

formed from the densest regions in the initial fluctuation field. This is because, in the initial Gaussian random field, the densest regions were more strongly clustered than regions of average density. Therefore, our simple model, in which the environmental dependence of ξ is entirely due to the environmental dependence of the halo mass function, suggests the following generic features for $\xi(r|\delta)$: on scales larger than the virial radius of a typical halo, the amplitude of ξ should increase as δ increases; on smaller scales, the amplitude of ξ in underdense regions should be larger; hence, the slope of ξ in dense regions should be shallower than in less dense regions.

2.2.3 Mock galaxy samples

To illustrate that the features described above really are generic, and to make a closer connection to observations, we assigned mock 'galaxies' to haloes in the simulations as follows. Sufficiently low mass haloes contain no galaxies. Haloes more massive than some m_L contain at least one galaxy. The first galaxy in a halo is called the 'central' galaxy. The number of other 'satellite' galaxies is drawn from a Poisson distribution with mean $N_s(m)$ where

$$N_s(m) = \left(\frac{m}{m_1}\right)^{\alpha} \text{ if } m \geq m_L. \tag{2.5}$$

This procedure is motivated by Kravtsov et al. (2004). We distribute the satellite galaxies in a halo around the halo centre so that the radial profile follows that of the dark matter (i.e., the galaxies are assumed to follow an NFW profile). We set $m_L = 10^{11.27} h^{-1} M_{\odot}$, $m_1 = 23 m_L$, and $\alpha = 0.92$; Zehavi et al. (2005) show that this choice is appropriate for SDSS galaxies more luminous than $M_r < -18$. We then compute N_5, the number of galaxies in a $5h^{-1}$Mpc sphere around each galaxy, and rank the galaxies in order of decreasing N_5. The top one-third are labeled as being galaxies in dense regions, and the bottom third as being in underdense regions.

Figure 2.5 presents the correlation functions of the mock galaxies classified as being in dense (top) and underdense (bottom) regions. Once again we see the generic tendency for ξ computed using objects in denser regions to be shallower than when objects in less dense environments are used.

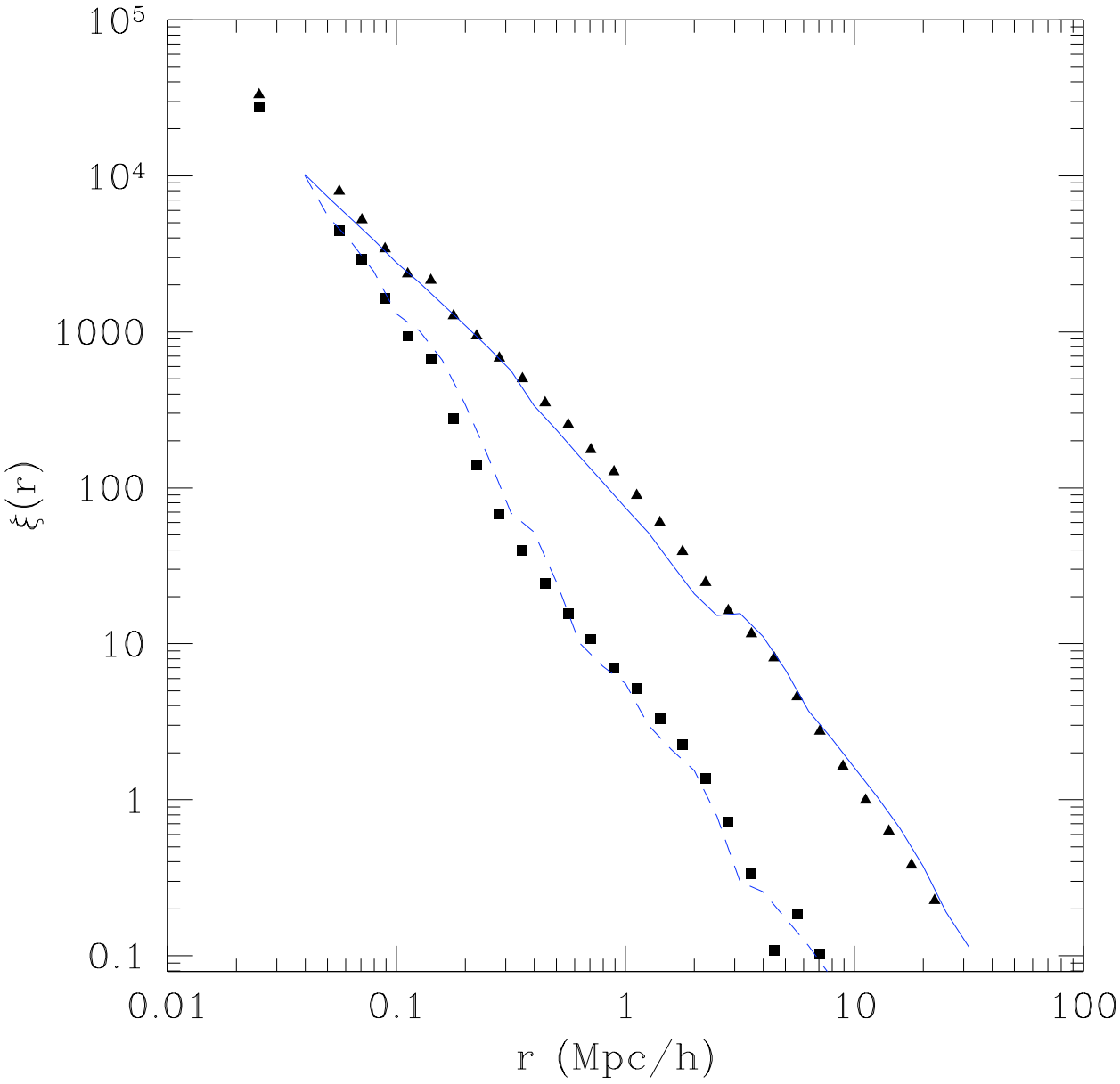

Figure 2.5: Similar to Figure 2.2, but now for model galaxies distributed according to the model described in the text around Equation 2.5, and density defined by counting galaxies in spheres centred on the galaxies themselves, rather than the mass in spheres centred on haloes. Curves show the analytic model developed in Section 2.3.

The important point, which we note explicitly here, is the following: By assuming that Equation (2.5) is the same function of m for all environments, and by assuming that the radial profile of the galaxies depends only on halo mass and not on environment, we have constructed a galaxy catalog in which all environmental effects are *entirely* a consequence of the correlation between halo mass and environment. Therefore, the locii traced out by the two sets of symbols shown in Figure 2.5 represent the predicted environmental dependence of ξ if there are no environmental effects other than the statistical one determined by the initial fluctuation field.

Over the $1 - 10h^{-1}$Mpc range which the SDSS data currently probe most reliably, ξ for the dense and underdense samples differs by an order of magnitude. This difference is easily measurable in data sets which are currently available. Comparison of this predicted difference with measurements in the SDSS will provide a sharp test of the assumption that environmental effects are dominated by the statistical correlation between the halo mass function and environment rather than by other 'gastro'physics, and is studied in the next chapter.

2.3 THE ANALYTIC MODEL

This Section provides an analytic model which incorporates the assumption that environmental effects are determined entirely by the statistical correlation between halo mass and environment. Note that to perform the test with data, this analytic model is unnecessary—measurements of clustering in the mock catalogs described previously are sufficient.

The analysis which follows uses the framework of the halo model of large scale structure. It extends that of the toy model described in Section 2.2.2 in two ways: it allows for a range of halo masses, and it allows for correlations between haloes.

2.3.1 Including the environmental effect

The expressions for the halo model are the result of averaging over environments. Including the environmental dependence of the halo distribution explicitly is not entirely straightfor-

ward, because, as we describe below, it requires knowledge of the probability that a region of volume V has overdensity δ. As we describe below, we use the excursion set model described in Sheth (1998) to do this.

In this model, spherical evolution is described by a 'moving barrier':

$$\delta_0(\Delta) = \frac{\delta_{\rm sc}}{1.686}\left[1.686 - \frac{1.35}{\Delta} - \frac{1.124}{\Delta^{1/2}} + \frac{0.788}{\Delta^{0.587}}\right] \tag{2.6}$$

where $\Delta \equiv M/\bar{\rho}V$. This barrier is said to be 'moving' because, for general V, it is a function of M. The excursion set model attributes special significance to the first crossing distribution $f(M,V)dM$ of this barrier by Brownian motion random walks: it is a measure of the mass fraction in regions of size V which contain mass M. In this approach, a halo can be thought of as a patch which has collapsed to vanishingly small volume. In the limit of $V \to 0$, $\delta_0 \to \delta_{\rm sc}$ is the same constant for all M. In this limit the barrier is said to have constant 'height', and the first crossing distribution $f(M,\delta_{\rm sc})dM$ represents the mass fraction in haloes of mass M. Thus, $f(M,\delta_{\rm sc})$ equals $(M/\bar{\rho})$ times $dn(M,\delta_{\rm sc})/dM$, the halo mass function.

Notice that, for general V, $\delta_0(\Delta) \le \delta_{\rm sc}$ for all M. Hence, $f(m,\delta_{\rm sc}|M,V)$, the first crossing distribution of $\delta_{\rm sc}$ by walks which first crossed the moving barrier associated with non-zero V at M, denotes the mass fraction of cells of size V which contain mass M which is in haloes which have mass m for some $m \le M$. This fraction equals (m/M) times $N(m,\delta_{\rm sc}|M,V)$, the environmental dependent mass function discussed previously (cf. Equation 2.1).

In Section 2.2.1, we define the environment of a halo by considering the mass in a patch of volume V surrounding it. To classify haloes by their environment, we must be able to estimate the number density of haloes of mass m which are surrounded by regions of volume V which contain mass M. If $dn(m)/dm$ denotes the number density of m haloes, and $f(M,V|m)$ denotes the fraction of m haloes which contain mass M in the surrounding volume V, then the number density of such haloes is $f(M,V|m)\,dn(m)/dm$. In the excursion set model, this equals

$$\begin{aligned}\frac{dn(m)}{dm}\,f(M,V|m)dM &= \frac{\bar{\rho}}{m}\,f(m)f(M,V|m)\,dM \\ &= \frac{\bar{\rho}}{m}\,f(M,V)f(m|M,V)\,dM \\ &= n(M,V)\,N(m|M,V)\,dM,\end{aligned} \tag{2.7}$$

where $f(M,V)dM$ is given by computing the first-crossing distribution of the moving barrier associated with spherical collapse described above. (In practice, we use the analytic approximation to the first crossing distribution of such moving barrier problems given by Sheth & Tormen 2002).

The mass density contributed by haloes that are embedded in regions of mass $M_{min} \leq M \leq M_{max}$ is;

$$\begin{aligned}\bar{\rho}_\delta &= \int_{M_{min}}^{M_{max}} dM \int_0^{M_{max}} dm \, \frac{dn(m)}{dm} \, f(M,V|m) \, m \\ &= \int_{M_{min}}^{M_{max}} dM \, n(M,V) \int_0^{M} dm \, N(m|M,V) \, m \\ &= \int_{M_{min}}^{M_{max}} dM \, n(M,V) \, M. \end{aligned} \tag{2.8}$$

In the standard model, the density profile of a halo depends on its mass, but not on the surrounding environment. In this case, the one-halo term is

$$\begin{aligned}P_{1h}(k|\delta) = &\int_{M_{min}}^{M_{max}} dM \, n(M,V) \\ &\times \int_0^{M} dm \, N(m|M,V) \left(\frac{m}{\bar{\rho}_\delta}\right)^2 |u(k|m)|^2. \end{aligned} \tag{2.9}$$

This reduces to the standard 1-halo term in the limits $M_{min} \to 0$ and $M_{max} \to \infty$.

The two-halo term is more complex as it now has two types of contributions: pairs which are in the same patch (2-halo–1-patch), and pairs in different patches (2-halo–2-patch). The 2-halo–1-patch term can only be important on intermediate scales (i.e., those which are larger than the diameter of a typical halo but smaller than the diameter of a patch). It is more complex to model this term accurately, as we describe shortly. The 2-halo–2-patch term, on the other hand, is simpler. It should be well approximated by

$$\begin{aligned}\frac{P_{2h-2p}(k|\delta)}{P_{\rm Lin}(k|R_p)} = &\left[\int_{M_{min}}^{M_{max}} dM \, n(M,V) \, B(M,V) \right. \\ &\left. \times \int_0^{M} dm \, N(m|M,V) \frac{m}{\bar{\rho}_\delta} \, u(k|m)\right]^2, \end{aligned} \tag{2.10}$$

where $P_{\rm Lin}(k|R_p)$ denotes the power spectrum associated with setting the linear theory correlation function to -1 on scales smaller than the diameter of a patch $2R_p$. This truncation

has little effect on small $kR_p \ll 1$, where $P_{\rm Lin}(k|R_p) \approx P_{\rm Lin}(k)$. The factor $B(M,V)$ describes bias associated with the clustering of the patches, and depends on the abundance of such patches in exactly the same way that $b(m)$ is related to $dn(m)/dm$ (c.f. Equation 2.2). Note that this expression assumes that the 2-halo–2-patch term is given simply by weighting the patch-patch correlation by the halo abundance within a patch.

To a first approximation, patches do not overlap with one another, so the 2-halo–2-patch term should drop on scales smaller than the diameter of a patch. It is on these scales that P_{2h-1p} should begin to dominate. A first approximation for the net effect of P_{2h-1p}, then, is to not enforce this small-scale decrease of P_{2h-2p}, and to simply use the expression above for P_{2h-2p} but $P_{\rm Lin}(k)$ instead of $P_{\rm Lin}(k|R_p)$, for all k. The appendix shows that this expression reduces to the standard two-halo term when $M_{min} \to 0$ and $M_{max} \to \infty$, so it is at least a reasonable approximation. However, we will see below that the correlation function of objects in underdense regions sometimes shows a clear signature of the fact that patches do not overlap. Thus, a more sophisticated approximation is required to describe clustering in underdense regions.

To estimate the 2-halo–1-patch term, it is convenient to think of the patches as haloes, and of the haloes as patch-substructure. Sheth & Jain (2003) have developed the halo model description of clustering when haloes have substructure. They allow for the possibility that a halo may be made up of a smooth component plus a population of subclumps. Our present case corresponds to assigning all the mass to subclumps, and none to the smooth component, so that only the final two of the four terms in their equation (31) contribute. Our expression for $P_{1h}(k|\delta)$ is effectively the same as their final term, so their third term is our $P_{2h-1p}(k|\delta)$. Namely,

$$\begin{aligned} P_{2h-1p}(k|\delta) = & \int_{M_{min}}^{M_{max}} dM\, n(M,V) \\ & \times \int_0^M dm_1 \left(\frac{m_1}{\bar{\rho}_\delta}\right) |u(k|m_1)| \\ & \times \int_0^{M-m} dm_2 \left(\frac{m_2}{\bar{\rho}_\delta}\right) |u(k|m_2)| \\ & \times \; N(m_1,m_2|M,V)\; U(k|m_1,m_2,M)^2. \end{aligned} \tag{2.11}$$

Here U denotes the normalized Fourier transform of the spatial distribution of m_1 and m_2 haloes within a patch. A simple first estimate would use the correlation function of the haloes to model this profile, but to truncate this at the patch radius, since both haloes are required to lie within the patch:

$$U^2 \approx b(m_1)b(m_2)\,\frac{P_{\rm Lin}(k) - P_{\rm Lin}(k|R_p)}{M/\bar{\rho}} \tag{2.12}$$

where $P_{\rm Lin}(k) - P_{Lin}(k|R_p)$ denotes the power spectrum associated with setting the linear theory correlation function to zero on scales larger than the diameter of a patch $2R_p$. The other term, $N(m_1, m_2|M, V)$, denotes the average number of haloes of mass m_1 and m_2 in patches V which contain total mass M. Sheth & Lemson (1999) argue that, as a consequence of mass conservation,

$$N(m_1, m_2|M, V) \approx N(m_1|M, V)\,N(m_2|M - m_1, V - v_{m_1}), \tag{2.13}$$

where v_m is the volume associated with an m halo.

When $M \gg m$ and $V \gg v_m$, as it is in large overdense regions, then $N(m_1, m_2|M, V) \approx N(m_1|M, V)\,N(m_2|M, V)$ i.e, $N(m_1, m_2|M, V)$ is well approximated by the product of the individual mean values. In this case,

$$\begin{aligned}\frac{P_{2h-1p}(k|\delta)}{P_{\rm Lin}(k) - P_{\rm Lin}(k|R_p)} \approx \int_{M_{min}}^{M_{max}} dM\,\frac{n(M,V)}{M/\bar{\rho}} \\ \times\left[\int_0^M dm\,N(m|M,V)\,b(m)\right. \\ \left.\times\left(\frac{m}{\bar{\rho}_\delta}\right)|u(k|m)|\right]^2.\end{aligned} \tag{2.14}$$

In underdense regions, however, simply using the product of the individual mean values is expected to be a bad approximation.

The smooth curves in Figure 2.2 show that the model developed above provides a good description of the environmental dependence of the dark matter correlation function. In our comparisons, we have found that the full expression (equations 2.11–2.13) provides a substantially better description of P_{2h-1p} in the underdense regions, whereas the simpler approximation of equation (2.14) is adequate for the dense regions.

2.3.2 From dark matter to galaxies

In order to extend the model above to describe the environmental dependence of galaxy clustering, we will follow a procedure similar to that given in Section 1.4. When all environments have been averaged over, the mean number density of galaxies $\bar{n}_{gal}$ is given by replacing the weighting by m for the mean density by $\langle N_{gal}|m\rangle$, the mean number of galaxies in an m-halo. For instance, one could use equation (2.5) to write $\langle N_{gal}|m\rangle = 1 + \langle N_s|m\rangle$. Similarly, the 1-halo and 2-halo terms can be obtained.

If one had an estimate of the dark matter density field, then one could include the environmental dependence of galaxy clustering by writing the mean number of galaxies in an m-halo surrounded by a region V containing mass M as $\langle N_{gal}|m, M, V\rangle$. If this mean number did not depend on M and V, then the environmental dependence of galaxy clustering would be described by making the same replacements in the expressions for $P(k|\delta)$ as one makes for $P(k)$.

In practice, one observes galaxies, not dark matter, so one has an estimate of the galaxy density field $\delta_{gal} = N/\bar{n}_{gal}V - 1$, and not of the dark matter $\delta = M/\bar{\rho}V - 1$. Our previous expressions show what to do if the environment as defined by the mass density δ is known; describing galaxy clustering as a function of environment defined by the galaxies themselves, δ_{gal} rather than by the dark matter δ, is considerably more complicated.

2.3.2.1 Effects of scatter in the $\delta_{gal} - \delta$ relation However, a simple approximate model can be derived if one assumes that $\langle N_{gal}|m, M, V\rangle = \langle N_{gal}|m\rangle$ is a monotonic function of m, and that the scatter around this mean relation is small. The reason why is particularly easy to see if $\langle N_{gal}|m\rangle \propto m$ and there is no scatter around this relation. If the environment of each galaxy is quantified by the number N of other galaxies in a volume V around it, then rank ordering cells by N, which is an observable, is the same as rank ordering by M, which is not. This rank ordering allows us to describe the environmental dependence of galaxy clustering by making simple adjustments to the expressions of Section 2.3.1.

In particular, suppose that one has measured how the number density of galaxies surrounded by regions containing at least N_{min} other galaxies depends on N_{min}. This number

density is given by summing over the observable quantity N. However, this number density can also be written as

$$\begin{aligned} \bar{n}_{\delta-gal} = & \int_{M_{min}(N_{min})}^{\infty} dM\, n(M,V) \\ & \times \quad \int_0^M dm\, N(m|M,V)\, \langle N_{gal}|m\rangle, \end{aligned} \tag{2.15}$$

where $M_{min}(N_{min})$ is obtained by requiring that the value of this expression match that observed as N_{min} is varied. Once $M_{min}(N_{min})$ is known, the two-halo two-patch term can be written as

$$\begin{aligned} P^{gal}_{2h-2p}(k|\delta) &\approx b^2_{\delta-gal}\, P_{Lin}(k), \quad \text{where} \\ \bar{n}_{\delta-gal}\, b_{\delta-gal} &= \int_{M_{min}}^{\infty} dM\, n(M,V)\, B(M,V) \\ & \times \int_0^M dm\, N(m|M,V)\, \langle N_{gal}|m\rangle, \end{aligned} \tag{2.16}$$

with the analagous substitutions for the $(m/\bar{\rho}_\delta)$ terms in Equation (2.11) for $P_{2h-1p}(k|\delta)$. For similar reasons, the one-halo term in the centre plus Poisson satellites model is

$$\begin{aligned} P^{gal}_{1h}(k|\delta) = & \int_{M_{min}}^{\infty} dM\, n(M,V) \int_0^M dm\, N(m|M,V) \\ & \times \quad \frac{[2\langle N_s|m\rangle\, u(k|m) + \langle N_s|m\rangle^2\, |u(k|m)|^2]}{\bar{n}^2_{\delta-gal}}. \end{aligned} \tag{2.17}$$

These expressions, which follow from those in Section 2.3.1, are only accurate if the relation between the number of galaxies in a cell N (the observable) and the mass in the cell M is deterministic (i.e., has no scatter) and monotonic. There *will* be scatter in N at fixed M if $\langle N_{gal}|m\rangle$ is a monotonic but nonlinear function of m, even if there is no scatter around the $\langle N_{gal}|m\rangle$ relation. This scatter arises from the fact that there is scatter in the halo distribution at fixed M. (Write $N = \sum_i N(m_i)$ and $\sum_i m_i \equiv M$. This shows that N is independent of the distribution of the m_i only if $N(m_i) \propto m_i$.)

Scatter in N at fixed M means there is scatter in M at fixed N. The expressions above should provide reasonable approximations to the exact description if the mean mass M associated with a given value of N in a cell, $\langle M|N\rangle$, is a monotonic function with small scatter. In this context, 'small' means that the dependence of clustering on environment (e.g.,

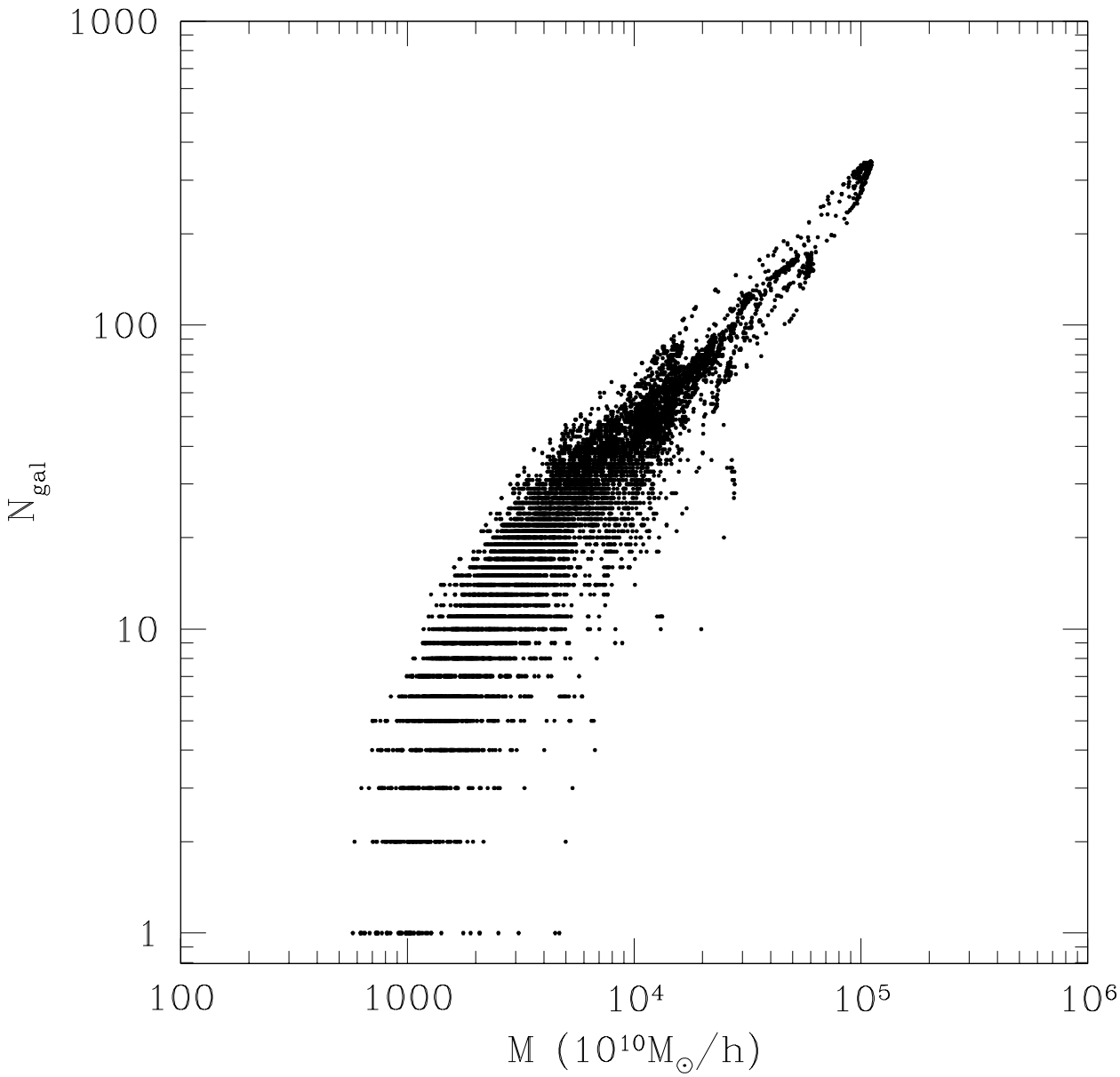

Figure 2.6: Number of galaxies within spheres of radii $5h^{-1}$ Mpc centered on each galaxy, versus the total mass in the sphere, for the mock galaxies described in Section 2.2.3. The small scatter in M at large N suggests that our model should be reasonably accurate.

the mix of haloes) does not change dramatically over the range of environments associated with the rms scatter in M around the mean $\langle M|N\rangle$. To see what this means, recall that, if the scale on which the environment is defined is large, then Equation (2.1) indicates that this dependence is proportional to $b(m)(M/\bar{\rho}V - 1)$. Massive haloes have $b(m) > 1$, with $b(m)$ a strongly increasing function of m, whereas low mass haloes have $b(m) < 1$, and the m-dependence is weak. Since the most massive haloes populate the densest cells, the model can tolerate larger scatter in the $M - N$ relation at small N than at large N.

Figure 2.6 shows the relation between the number of galaxies N and the mass M within a sphere of radius $5h^{-1}$Mpc centred on each galaxy, for the mock galaxies described in Section 2.2.3. The figure indicates that treating N as a monotonic deterministic function of M (and vice-versa) is a good approximation at least at high masses, even though the underlying relation between number of galaxies and halo mass, $\langle N_{gal}|m\rangle$, is nonlinear (Equation 2.5). Notice that the scatter around the mean $\langle M|N\rangle$ relation is particularly small at large N. Although the scatter increases at smaller N, we argued above that, in this regime, the effect of scatter is less important. Hence, Figure 2.6 suggests that the simple model developed in this section should be reasonably accurate.

Recall that Figure 2.5 shows measurements of the environmental dependence of galaxy clustering made in a mock galaxy catalog constructed using the GIF simulation following methods outlined in Section 2.2.3. The smooth curves in Figure 2.5 show that the model developed here is in reasonable agreement with the measured dependence of $\xi(r)$ on δ_{gal}.

2.3.2.2 The two contributions to the 2-halo term In constructing the model, we remarked that there were two types of contribution to the two-halo term. So one might wonder if there is some clear signature of the transition from one type of contribution to another. The mock galaxy sample used for Figure 2.5 does not result in a $\xi(r|\delta)$ with a clear inflection or break on the patch scale.

Figure 2.7 shows dot-plots for the underdense and dense samples where the x and y coordinates for the galaxies have been plotted, which is for a mock galaxy sample in the VLS simulations constructed with $M_{min} = 10^{12.72}h^{-1}M_{\odot}$ and $\alpha = 1.39$, chosen to be similar to galaxies more luminous than $M_r = -21$. The transition mentioned earlier is seen in

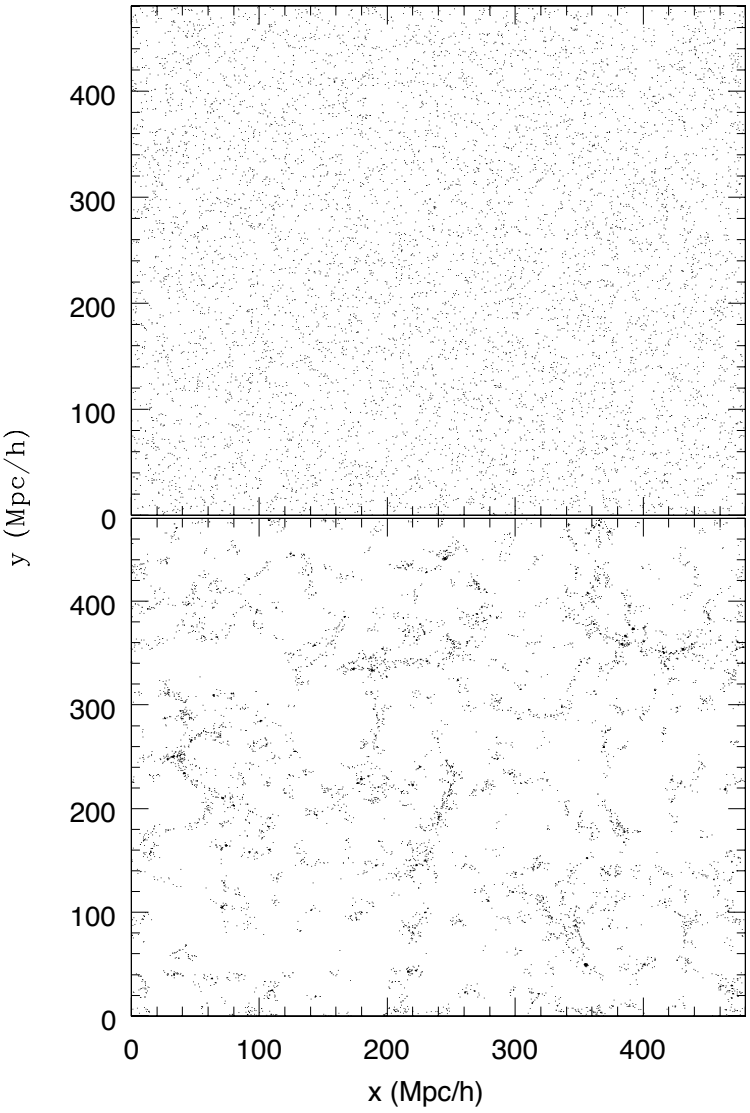

Figure 2.7: Dot plots of the mock galaxies in the VLS simulation, where the x vs y coordinate for the galaxies are plotted with $z < 50$. The top panel shows the one-third underdense sample, and lower panel shows the dense sample with the same number of points. The galaxies in the underdense sample fill the void regions which are present in the dense sample.

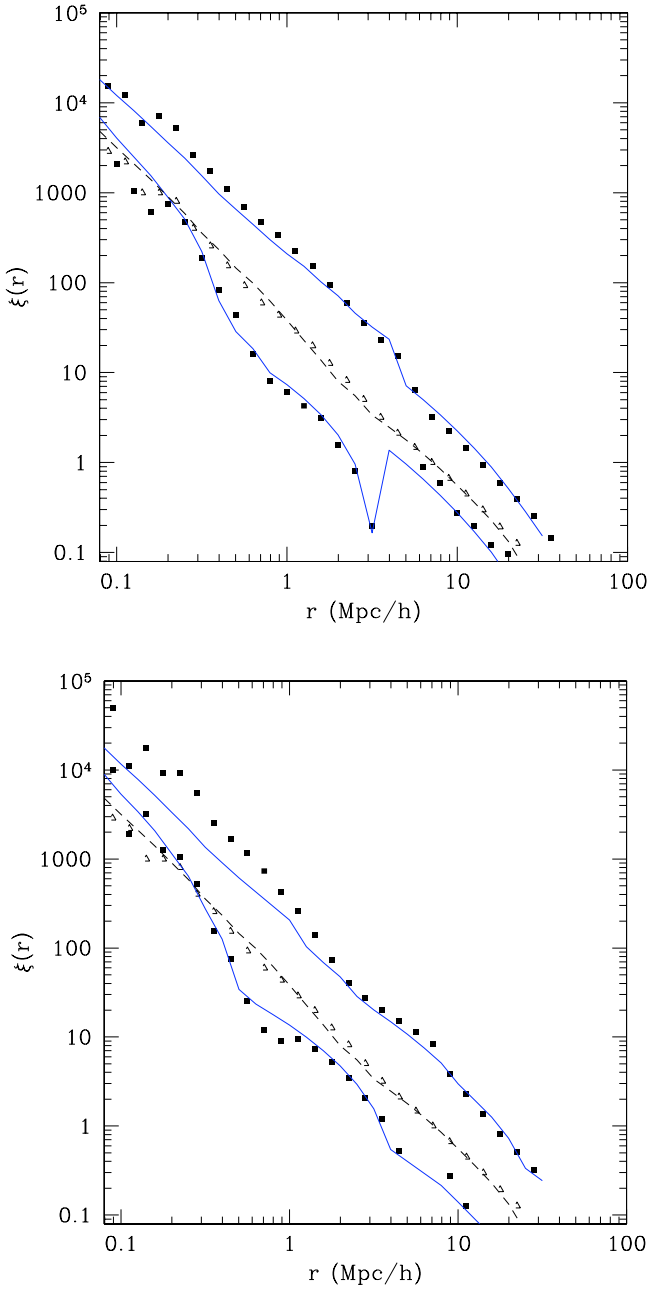

Figure 2.8: Correlation function for mock galaxies in the VLS simulations. Upper squares show $\xi(r|\delta)$ for the overdense sample and lower squares are for the underdense sample. In the top panel, the density was defined by counting galaxies within spheres having radius $5h^{-1}$Mpc centred on each object, and the bottom panel used $8h^{-1}$Mpc spheres. Curves show the associated analytic model which is able to describe the three different clustering regimes. The dashed curve and associated symbols in both panels represent the clustering for the full sample and are for the model and measurements respectively.

Figure 2.8 for these mock galaxies. The top and bottom panels show results from analyses in which the large scale environment was defined by the mass within spheres of radii $5h^{-1}$ and $8h^{-1}$Mpc, respectively. The inflection indicating the transition from $\xi_{2h-1p}(r|\delta)$ to $\xi_{2h-2p}(r|\delta)$ is more clearly seen for low density regions, and it is clearer when the scale which defines the environment is smaller. For these more luminous galaxies, using the simpler Equation (2.14) in P_{2h-1p} is an excellent approximation in dense regions, but grossly overestimates (by at least a factor of two) the clustering strength on intermediate scales in underdense regions. In underdense regions, Equations (2.11)–(2.13) are substantially more accurate. We would like to point out that the relation between N and M (that was mentioned in the previous section) within spheres of $8h^{-1}$Mpc, shows more scatter at large N. As the model is more sensitive to larger scatter at large N, we believe this causes the analytical prediction to depart from the simulation results at small scales, as shown in the lower panel of Figure 2.8.

It can be seen that at very small scales the correlation function for the total sample is smaller than that for the dense sample and is similar to that for the underdense sample. This can be explained by using the simple estimator (Eq 1.4) as follows; the dense, underdense and middle samples each have one-third the total number of data points, $n_d = n_u = n_m = 0.33n_t$, where the middle sample is the one-third sample containing all points not in the dense and underdense samples. Then the correlation function can be written as:

$$\begin{aligned} 1 + \xi_t(r) &= \frac{TT(r)/(n_t(n_t-1))}{RR(r)/(n_r(n_r-1))} \\ &= \frac{(D+U+M)(D+U+M)}{RR(r)(n_t(n_t-1))/(n_r(n_r-1))} \\ &= \frac{DD(r)+UU(r)+MM(r)+2DU(r)+2DM(r)+2UM(r)}{RR(r)(n_t(n_t-1))/(n_r(n_r-1))} \end{aligned} \tag{2.18}$$

At scales smaller than the patch size DU(r), DM(r) and UM(r) are zero, then the expression above simplifies to:

$$\begin{aligned} 1 + \xi_t(r) &= (1+\xi_d)/9 + (1+\xi_u)/9 + (1+\xi_m)/9 \\ \xi_t(r) &= (\xi_d(r) + \xi_u(r) + \xi_m(r))/9 - (2/3) \end{aligned} \tag{2.19}$$

Now, Figure 2.8 shows that at some small scale (≈ 0.2 Mpc/h) $\xi_t(r) = \xi_u(r)$ with amplitudes much greater than 1, this gives us $\xi_t(r) \approx (\xi_d(r) + \xi_m(r))/8$. The particles in the middle

sample are at average density, which means that we can assume $\xi_m(r)$ to be small compared to $\xi_d(r)$. This shows that $\xi_d(r)$ is roughly ten times $\xi_u(r)$ (or $\xi_t(r)$), which is seen to be the case in Figure 2.8 at around 0.2 Mpc/h. Furthermore, we can also see that at some small scale when $\xi_d(r) = \xi_u(r)$, $\xi_t(r) = (2\xi_d(r) + \xi_m(r))/9$. If $\xi_m(r)$ has an amplitude similar to either $\xi_u(r)$ or $\xi_d(r)$, then $\xi_t(r)$ is roughly one-third that of $\xi_u(r)$ or $\xi_d(r)$.

Figure 2.2 indicates that our model for the environmental dependence of dark matter clustering is accurate, and the agreement between the symbols and the curves in Figures 2.5 and 2.8 indicates that our approximate model for the environmental dependence of galaxy clustering also works quite well. Although it is reassuring that the model works so well—it suggests that we understand the physics which drives the environmental dependence of clustering—this analytic description is not needed to perform the test of environmental effects described earlier.

2.3.2.3 In redshift space Our test using mock catalogs is unrealistic in one respect: with real data, one only has redshift-space positions, so one cannot define the local density in real-space. However, if the scale on which the redshift-space density is defined is larger than the typical size of redshift-space distortions ($\sim 5h^{-1}$Mpc), then the real and redshift-space densities will not differ substantially. Hence, we do not expect our separation into environments based on the real-space density within $8h^{-1}$Mpc spheres to differ substantially from that which we would have obtained had we used redshift-space positions instead. This section discusses how the halo model calculation of environmental effects on clustering can be extended to include redshift-space effects useful for the next chapter. Our strategy is to combine the halo-model description of redshift space effects (Seljak 2001; White 2001) with the halo model description of environmental effects provided in this chapter.

In redshift space, two effects modify the real space expressions derived previously. One of these is a boost of power on large scales due to the instreaming of matter into overdense regions (Kaiser 1987); this affects the 2-halo terms. Using density conservation to linear order and making the distant observer approximation, the redshift-space galaxy density perturbation can be written as

$$\delta_g^{rs} = \delta_g + \delta_v \mu^2 \tag{2.20}$$

where $\mu = \hat{r} \cdot \hat{k}$, δ_g is the real space galaxy density perturbation and δ_v is the velocity divergence. This is related to the density perturbation δ_{dm} via $\delta_v = f\delta_{dm}$, where $f(\Omega) \equiv d\log\delta/d\log a \simeq \Omega^{0.6}$, and a is the scale factor.

The other effect is the suppression of power due to the virial motions within haloes; this affects the 1-halo term (Sheth 1996). The assumption of isotropic, isothermal, Maxwellian motions within halos means that the effect can be modeled as a convolution with a Gaussian. In particular, the density contrast in redshift space is

$$\delta_g^{rs} = \delta_g e^{-(k\sigma\mu)^2/2}. \tag{2.21}$$

Scoccimarro (2004) discusses why these descriptions (Equations 2.20 and 2.21) of redshift-space distortions are rather crude. For our purposes, they represent reasonable first approximations to a more sophisticated model.

The isotropized redshift space power spectrum is obtained by averaging $(\delta_g^{rs})^2$ over μ, m and M. In particular, the 1-halo term can be written as,

$$\begin{aligned} P_{1h}^{gal}(k|\delta) = & \int_{M_{min}}^{M_{max}} dM\, n(M,V) \int_0^M dm\, N(m|M,V) \\ & \times \frac{[\langle 2N_s|m\rangle\, u(k|m)\Re_1(k\sigma) + \langle N_s|m\rangle^2\, |u(k|m)|^2\Re_2(k\sigma)]}{\bar{n}^2_{\delta-gal}}, \end{aligned} \tag{2.22}$$

where

$$\Re_p(\alpha = k\sigma[p/2]^{1/2}) = \frac{\sqrt{\pi}}{2}\frac{\mathrm{erf}(\alpha)}{\alpha} \tag{2.23}$$

for $p = 1, 2$, and $\bar{n}_{\delta-gal}$ is the number density of galaxies surrounded by regions containing at least N_{min} other galaxies:

$$\begin{aligned} \bar{n}_{\delta-gal} = & \int_{M_{min}(N_{min})}^{M_{max}(N_{max})} dM\, n(M,V) \\ & \times \int_0^M dm\, N(m|M,V)\, \langle N_{gal}|m\rangle. \end{aligned} \tag{2.24}$$

As before, the two-halo term is more complex as it now has two types of contributions: pairs which are in the same patch (2h-1p), and pairs in different patches (2h-2p). The 2h-1p term can only be important on intermediate scales (i.e., those which are larger than the

diameter of a typical halo but smaller than the diameter of a patch). The 2h-2p term (which dominates at scales larger than the diameter of the patch) is:

$$P_{2h-2p}^{gal} = (F_g^2 + \frac{2}{3}F_v F_g + \frac{1}{5}F_v^2)P_{\rm Lin}(k|R_p), \tag{2.25}$$

where

$$\begin{aligned} F_v &= f\int_{M_{min}}^{M_{max}} dM\, n(M,V)\, B(M,V) \\ &\quad \times \int_0^M dm\, N(m|M,V)\, \frac{m}{\rho_\delta}\Re_1(k\sigma)u(k|m) \\ F_g &= \int_{M_{min}}^{M_{max}} dM\, n(M,V)\, B(M,V) \\ &\quad \times \int_0^M dm\, N(m|M,V)\, \frac{\langle N_{gal}|m\rangle}{\bar{n}_{\delta-gal}}\Re_1(k\sigma)u(k|m); \end{aligned} \tag{2.26}$$

Similarly, the 2h-1p term can be written as

$$\begin{aligned} P_{2h-1p}^{gal}(k|\delta) = \int_{M_{min}}^{M_{max}} dM\, n(M,V)(F_g'^2 + \frac{2}{3}F_v'F_g' + \frac{1}{5}F_v'^2) \\ \times \Big[P_{Lin}(k) - P_{Lin}(k|R_p)\Big], \end{aligned} \tag{2.27}$$

and,

$$\begin{aligned} F_v' &= f\int_0^M dm\, N(m|M,V)\, \frac{m}{\rho_\delta}\Re_1(k\sigma)b(m)\, u(k|m) \\ F_g' &= \int_0^M dm\, N(m|M,V)\, \frac{\langle N_{gal}|m\rangle}{\bar{n}_{\delta-gal}}\Re_1(k\sigma)b(m)\, u(k|m). \end{aligned} \tag{2.28}$$

Here $b(m)$ is the bias factor of haloes of mass m (from Sheth & Tormen 1999). The correlation function, $\xi(s)$, is obtained by taking the Fourier transform of the power spectrum $P(k)$.

2.4 DISCUSSION AND CONCLUSIONS

In hierarchical models, the clustering of dark matter particles should be a strong function of environment. This is a consequence of the fact that massive haloes populate the densest regions. If the properties of galaxies are determined entirely by the masses and formation histories of the haloes in which they sit, then the clustering of galaxies should also depend strongly on environment.

We discussed a method for testing this assumption. Section 2.2.3 described how to generate a mock galaxy catalog in which all correlations with environment are a consequence of the correlation between halo abundances and environment. Comparison of the environmental dependence of clustering in the mock catalog and in real data allows one to test the assumption that environmental effects are dominated by the statistical correlation between halo mass and environment.

Section 2.3 showed how to incorporate the assumption that environmental effects are determined entirely by the statistical correlation between halo mass and environment into the halo model description of large scale structure. While this model aids in understanding the effect, to perform the test with data, it is unnecessary—measurements of clustering in the mock catalogs of the sort described in Section 2.2.3 are sufficient. However, the methodology developed in this chapter provides a means for computing the clustering properties of galaxies affected by other environment dependent processes (assuming that some model for $\langle N_{gal}|m, M, V\rangle$ could be derived) that explicitly do not depend upon the surrounding halo mass. For instance, winds from galaxies in nearby clusters are assumed to depend on the larger scale environment and could affect the clustering statistics.

In particular, on scales larger than a Megaparsec or so, the two-point correlation function ξ of galaxies surrounded by high density regions is predicted to be larger than for galaxies in less dense regions. In addition, the slope of ξ for galaxies in dense regions should be shallower (cf. Figure 2.5). And if the distribution of galaxies around halo centers depends only weakly on environment (current analyses of galaxy clustering assume any such dependence is negligible), the effect on their clustering amplitude is small.

3.0 THE ENVIRONMENTAL DEPENDENCE OF GALAXY CLUSTERING IN THE SLOAN DIGITAL SKY SURVEY

3.1 INTRODUCTION

In Chapter 2, we described how the clustering of galaxies can be used to test the assumption that the correlations between galaxy properties and their environments are *entirely* a consequence of the correlations between haloes and their environments. This is a strong assumption which significantly simplifies interpretation of the observed luminosity dependence of galaxy clustering (e.g. Zehavi et al. 2005). It is also a standard assumption in current halo-model descriptions of galaxy clustering (see Cooray & Sheth 2002, for a review). The main goal of this chapter is to perform this test.

This chapter is arranged as follows: in Section 3.2 we show how galaxy clustering depends on environment in the SDSS (Adelman-McCarthy et al. 2005). In particular, we measure the pair correlation function in redshift space, $\xi(s|\delta_s)$, for a range of environments δ_s, as well as the projected quantity, $w_p(r_p|\delta_s)$; the latter is free of redshift-space distortions. These measurements are compared with similar measurements in carefully constructed mock catalogs, and from an analytic calculation based on the halo-model (Section 2.3.2.3). In both the mocks and the analytic calculation, correlations between galaxy properties and environment are entirely a consequence of the correlation between galaxy properties and halo masses, and between halo mass and environment. We summarize our results in Section 3.3, where we also discuss some implications.

3.2 THE ENVIRONMENTAL DEPENDENCE OF CLUSTERING

To measure the environmental dependence of clustering, we must decide on a measure of the environment. In Chapter 2 we showed that the precise choice of environment is not particularly important, in the sense that different choices lead to quantitative but not qualitative differences. We used N_R, the number of galaxies in a sphere of radius R centred on a galaxy, as a measure of that galaxy's environment, and presented results for $R = 5$ and $8h^{-1}$Mpc. That analysis was performed in real-space. Performing a similar analysis in redshift-space is complicated because the environmental effect we would like to test is due to correlations between halo masses with the real-space density. Therefore, we must find a definition of density in redshift space which is as faithful as possible to that in real-space.

Line-of-sight redshift-space distortions can make a sphere in real-space appear very different in redshift space. For instance, around a spherically symmetric cluster there are two main effects: one is due to coherent infall around the center of the cluster, which appears as a squashing effect along the line of sight in redshift space (Kaiser 1987). The second is the "finger of God" effect which is due to the virial motions of galaxies within the cluster (de Lapparent et al. 1986). This shows up as an elongation of the cluster along the line of sight. The squashing effect is relatively small, producing effects of order unity or less, whereas the finger-of-god distortions are more dramatic—elongations along the line of sight are typically about a factor of ten. Since clusters have radii of a Mpc or so, fingers of god can extend up to about 10 Mpc. Therefore, while counts in redshift space spheres of radii $5h^{-1}$Mpc are not expected to faithfully trace the counts in the corresponding real-space spheres, counts in spheres of radii $8h^{-1}$Mpc, N_8 should be more similar. For this reason, in what follows we use N_8 as a measure of the environment of each galaxy. (If we wished to push to smaller scales, we could identify all the fingers of god, and then 'decompress' them, by rescaling the distances along the line-of-sight so that they have the same extent as across the line-of-sight (e.g. Tegmark et al. 2004). But performing such a 'manicure' is beyond the scope of the present work.)

We use N_8 to divide the galaxy population into three equal-sized subsamples: the third with the largest values of N_8 are defined as being the dense subsample, and the third with

the smallest values of N_8 are the underdense subsample. The data samples were pruned to keep only those points that have 80% of the surrounding 8 Mpc/h sphere within the sample, effectively getting rid of outliers and in order to obtain accurate estimates of the local density. We then measure the correlation functions in these two subsamples using the Landy-Szalay estimator (Equation 1.5).

Our strategy is to make such measurements in a volume limited galaxy catalog, so that selection effects are minimized. We then compare with similar measurements in realistic mock catalogs. Throughout, we show results for a flat ΛCDM model for which $(\Omega_0, h, \sigma_8) = (0.3, 0.7, 0.9)$ at $z = 0$. Here Ω_0 is the density in units of critical density today, the Hubble constant at the present time is $H_0 = 100h$ km s^{-1} Mpc^{-1}, and σ_8 describes the rms fluctuations of the initial field, evolved to the present time using linear theory, when smoothed with a tophat filter of radius $8h^{-1}$ Mpc.

3.2.1 The SDSS galaxy sample

We perform our analysis on a volume limited catalog extracted from the SDSS DR4 database (Adelman-McCarthy et al. 2005). We chose galaxies brighter than $M_r < -21$, to match the analysis of Zehavi et al. (2005), whose results we use below. The resulting catalog contains about 75000 objects with accurate angular positions and redshifts, where the number density is 0.00117 h^3/Mpc^3. As discussed above, we define the environment of each object using the redshift-space information, and then measure the clustering in subsamples defined according to environment. Figure 3.1 and 3.2 show the dot plots of the subsamples with y vs x and z vs x shown respectively.

Uncertainties on our measurements were estimated by jack-knife resampling, in which the statistics were remeasured after omitting a random region, and repeated thirty times (approximately 1.5 times the total number of bins in separation for the results presented).

3.2.2 Mock galaxy samples

We have generated realistic mock galaxy samples as follows. We start with the Very Large Simulation (VLS; Yoshida, Sheth & Diaferio 2001) and follow the procedure given in Section

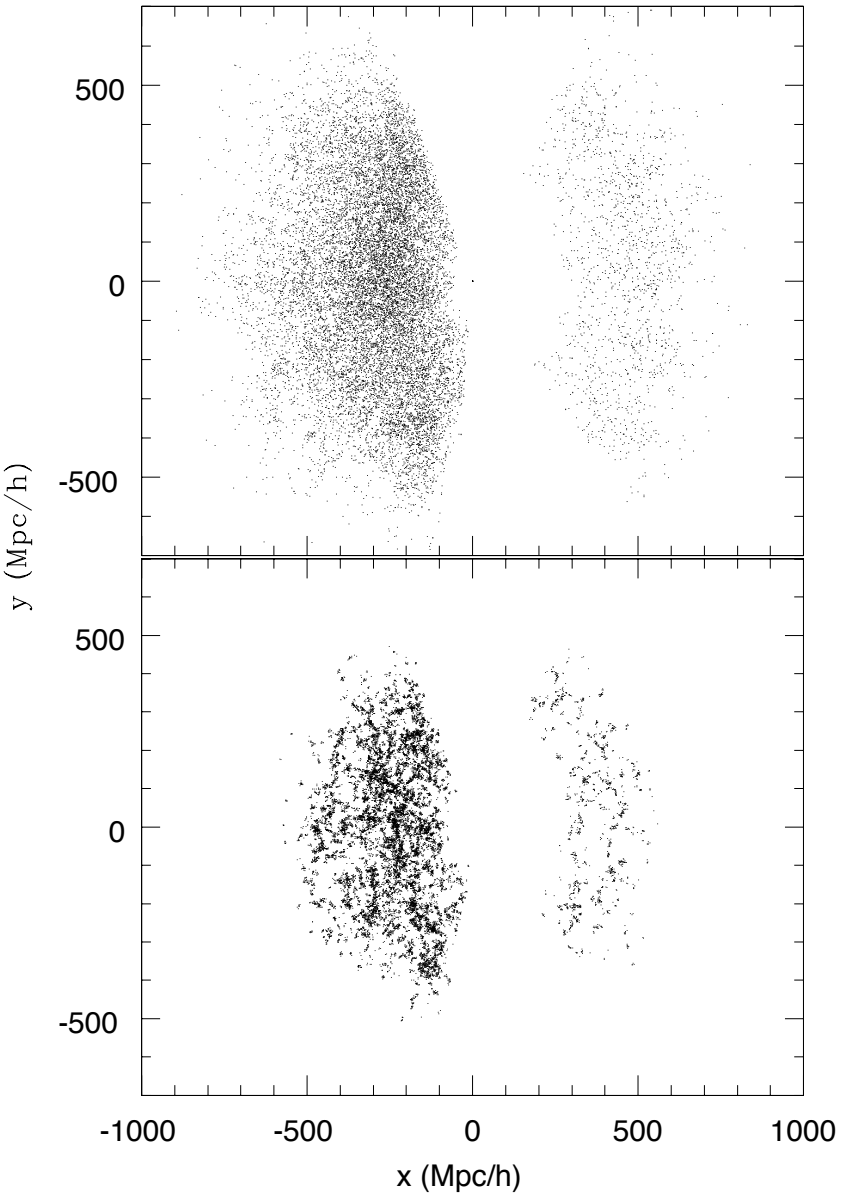

Figure 3.1: Dot plots of the SDSS subsamples, where the x vs y coordinates of the galaxies are plotted. The top panel shows the one-third underdense sample, and lower panel shows the dense sample with the same number of points. The panels depict 30% of the total number of galaxies in each sample. The underdense sample shows the presence of galaxies in the void regions that are seen in the dense sample.

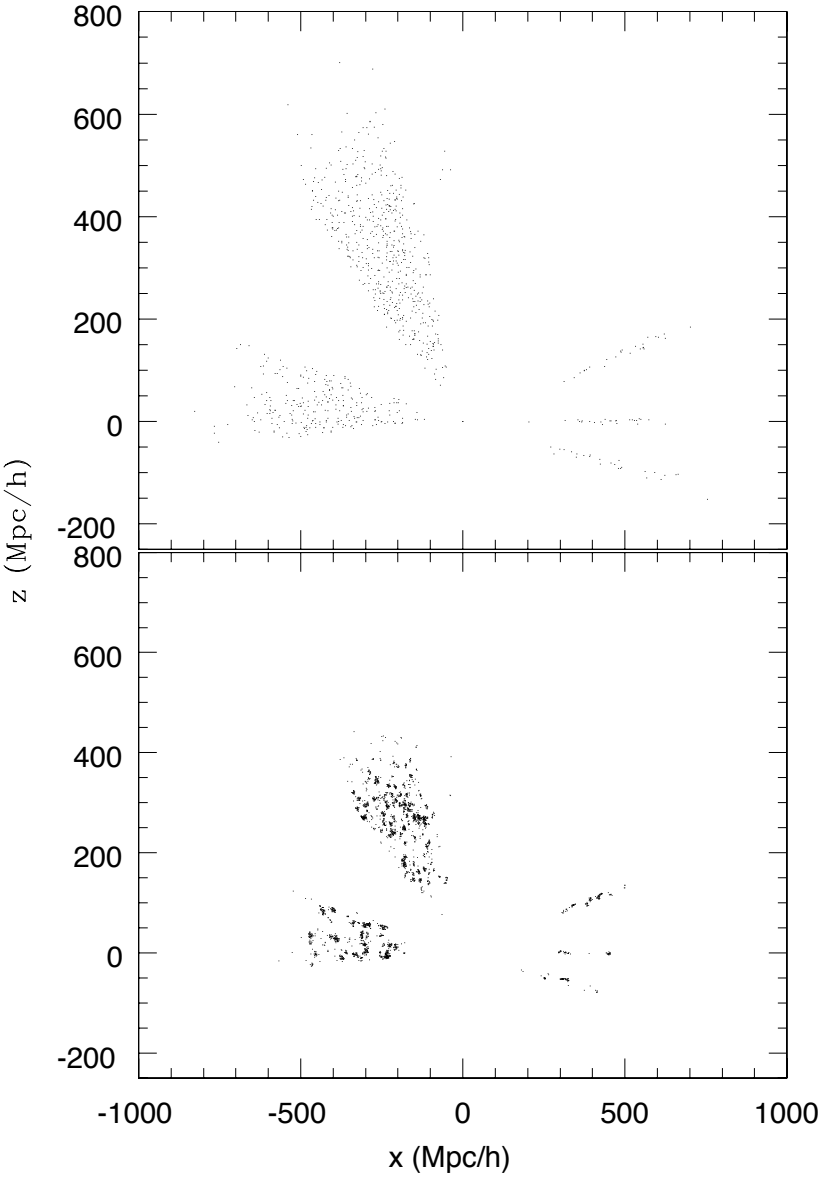

Figure 3.2: Dot plots of the SDSS subsamples, where the x vs z coordinates of the galaxies are plotted. The top panel shows the one-third underdense sample, and lower panel shows the dense sample with the same number of points, where each sample has 1% of the total number of galaxies.

2.2.3.

In this case, for SDSS galaxies more luminous than $M_r < -21$, $m_L = 10^{12.72} h^{-1} M_\odot$, $m_1 = 23 m_L$, and, $\alpha = 1.39$ (Zehavi et al. 2005). (A Poisson distribution for the number of satellites is motivated by the work of Kravtsov et al. 2004). We then assume that the satellites in a halo are distributed around the halo center similarly to the dark matter (e.g. Navarro et al. 1997).

To model redshift space effects, we must model the velocity vector of each mock galaxy. We do so by assuming that $v_{\rm gal} = v_{\rm halo} + v_{\rm vir}$, where $v_{\rm halo}$ is the halo motion provided by the simulation, and $v_{\rm vir}$ is obtained as follows. The central galaxy in a halo is assumed to be at rest with respect to the halo, so $v_{vir} = 0$. The virial motions of satellite galaxies are modelled by assuming that haloes are isotropic, virialized, and isothermal with Maxwellian velocities around the halo center. The one-dimensional velocity dispersion is $1000\,(r_{200} h/{\rm Mpc})/\sqrt{2}$, where r_{200} is the scale on which the enclosed mass is 200 times the critical density: $m = 200 \bar{\rho}_{\rm crit}\,(4\pi r_{200}^3/3)$. Following results in Sheth & Diaferio (2001), we assume that this virial term is independent of local environment.

In the distant observer approximation, the position in redshift space is $s = x + v_x/H_0$, where x is the real-space coordinate in the x-direction, v_x is the x-component of the peculiar velocity. and s is the redshift-space distance in the x-direction. The y- and z- components of the position are unchanged. The isothermal Maxwellian assumption means that the virial motions add Gaussian noise to the line-of-sight position of each satellite galaxy.

We then measure N_8 for each galaxy by counting the total number of galaxies within $8h^{-1}$Mpc. For the mock catalog, we can do this in both real- and redshift-space. Figure 3.3 compares these two estimates of the local density. They are not widely different, suggesting that the analysis in the next section will be useful. Figure 3.4 shows the dot plots of the x vs y coordinates of mock galaxies in the one-third underdense and dense subsamples.

3.2.3 Results

Figure 3.5 shows the redshift space correlation functions in dense and underdense regions measured in the mock catalogs (upper panel) and in the SDSS volume limited catalog (lower

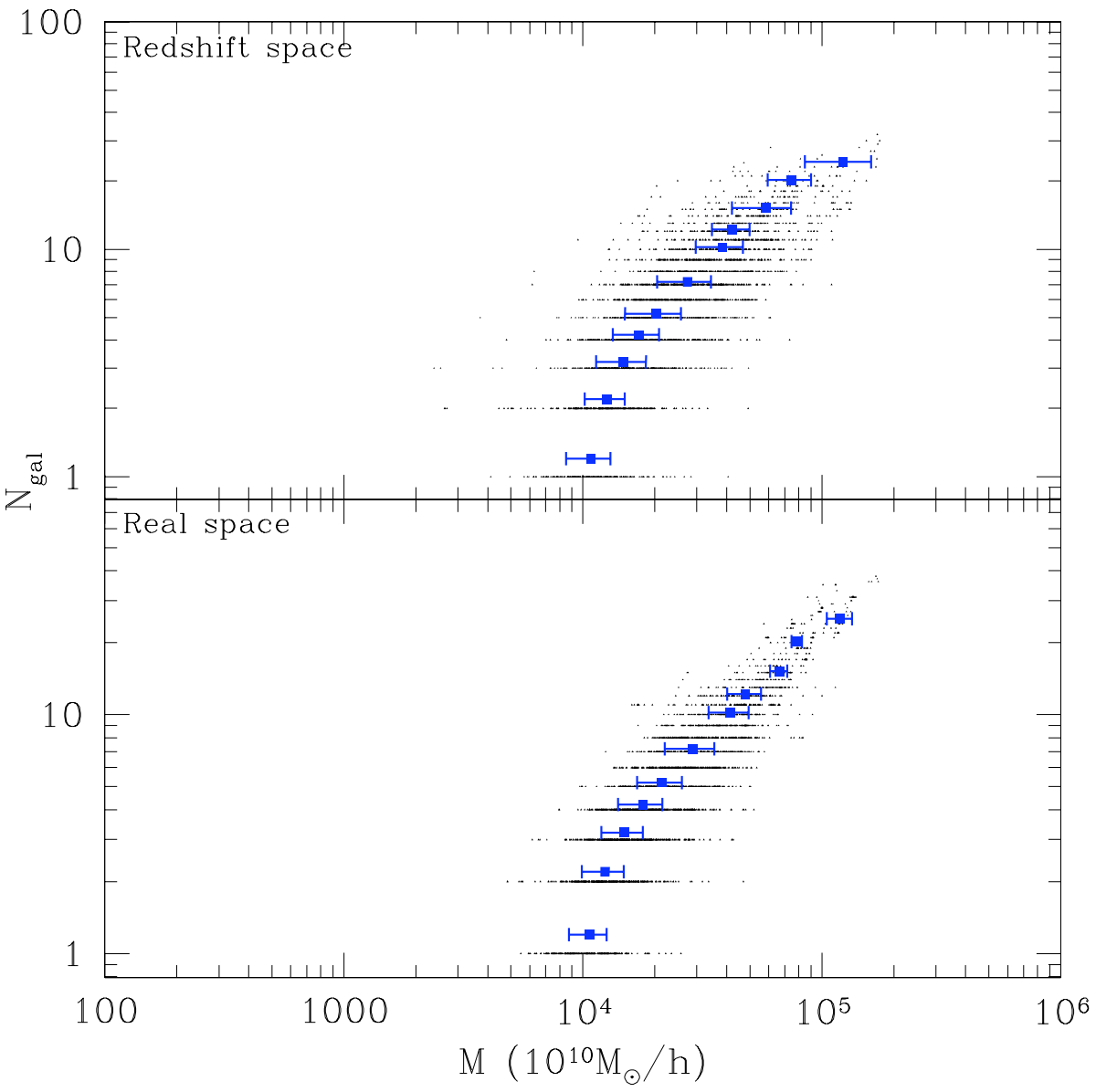

Figure 3.3: Comparison of local density estimates within real (bottom) and redshift (top) space spheres of radius $8h^{-1}$Mpc. The median (given by the squares) and quartile range of halo mass corresponding to certain number of galaxies is shown (for clarity these points have been shifted upwards).

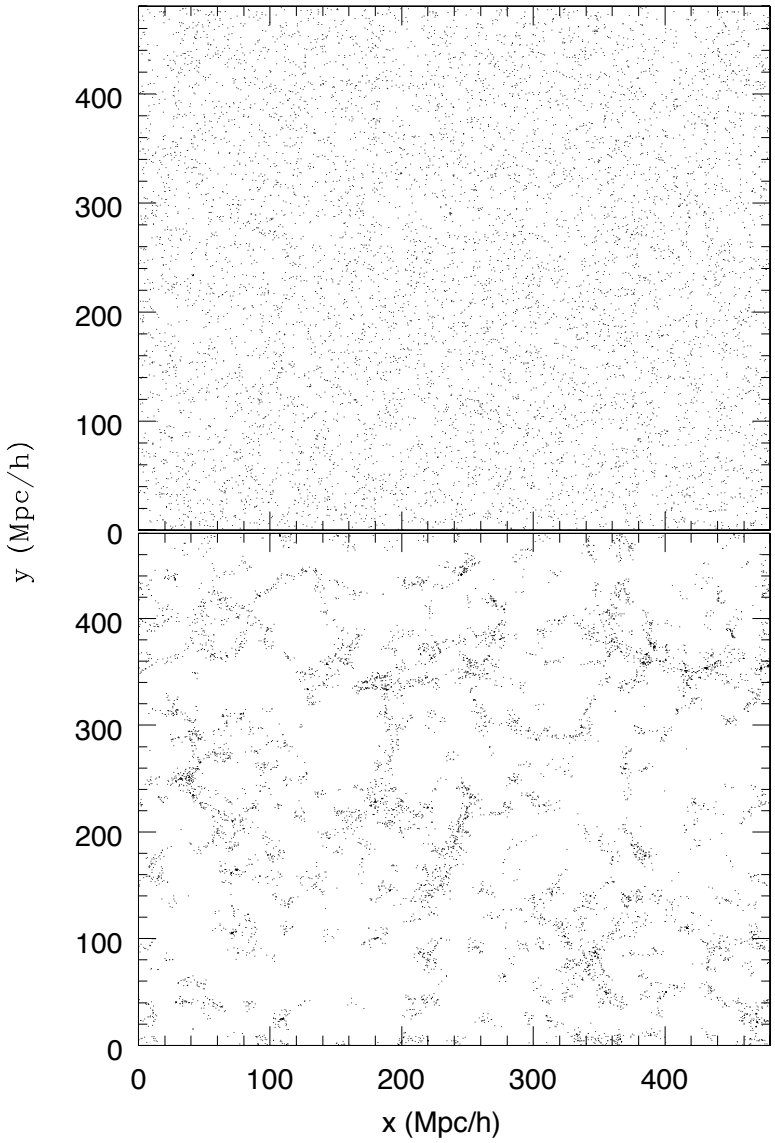

Figure 3.4: Dot plots of the mock galaxies in redshift space with their x coordinate vs y coordinate plotted for $z < 50$. The top panel shows the one-third underdense sample, and lower panel shows the dense sample with the same number of points.

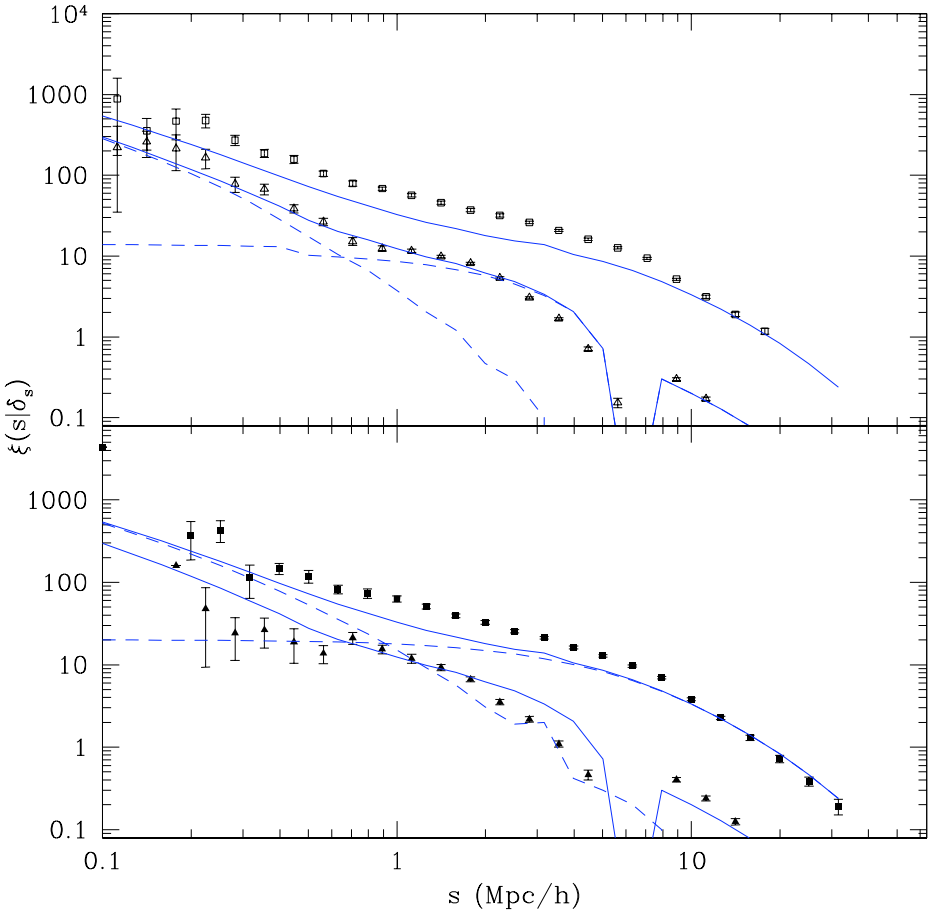

Figure 3.5: Environmental dependence of the galaxy correlation function in redshift space. Upper panel shows measurements in the mock catalog, and lower panel shows measurements for the SDSS given by symbols. In both cases, the galaxy catalog is a volume limited catalog to $M_r < -21$. In both panels, the environment of a galaxy was defined by counting the number of galaxies within a redshift-space sphere of radius $8h^{-1}$Mpc centred on it. The upper curve and corresponding square symbols in each panel shows $\xi(s|\delta_s)$ for the galaxies in the densest 1/3 of the sample, and the lower curve and triangle symbols is for the least dense 1/3. Solid curves show the analytic model for this environmental effect that is developed in Section 2.3.2.3. Dashed curves in the upper panel show the 1-halo (dominates on small scales) and the sum of the 2h-1p (dominates on intermediate scales) and 2h-2p (dominates on large scales) contributions to $\xi(s|\delta_s)$ of the less-dense sample. In the lower panel, the dashed curves show the corresponding contributions for the dense sample.

panel). In both panels, $\xi(s)$ for the dense sample is significantly larger than it is in the underdense sample. On large scales, this is because dense regions host the most massive haloes which in turn contain many galaxies; on smaller scales, the fact that the halo density profiles depend on halo mass also matters (see Chapter 2). The inflection or break on the patch scale (i.e. 8 Mpc/h) is caused by the transition from the 2h-1p term to the 2h-2p term and is clearly seen for the underdense sample. The 2h-1p and 2h-2p terms are both due to pairs of particles lying in separate haloes with the difference that in the former these pairs lie in the same patch, whereas in the latter they lie in different patches. For the underdense sample the low mass haloes that dominate have on average a virial radius of around 1 Mpc/h that is much smaller than the patch size. This causes the contribution of the 1-halo term (which is due to pairs in the same halo) to drop off sufficiently rapidly with increasing scale and thereby allow the 2h-1p contribution to show up. On the other hand, for the dense sample this inflection is not seen due to the larger virial radii of the high mass haloes, which causes the 2h-2p term to pick up by the time the 1-halo term dies off.

The solid curves in the two panels show the analytic calculation outlined in Section 2.3.2.3. They provide a reasonable description of the measurements in both panels. However, while the agreement is good on large scales, the curves underestimate the small-scale signal in dense regions. Since these smaller scales are the ones most affected by finger-of-god distortions, it may be that the discrepancy is due to inadequacies in the analytic treatment of redshift-space effects (see Scoccimarro 2004, for a discussion of the sorts of effects our analysis ignores).

To eliminate this source of uncertainty, we have also studied the projected quantity

$$w_p(r_p|\delta_s) = 2\int_0^\infty d\pi\, \xi(r_p, \pi|\delta_s); \tag{3.1}$$

where $r = \sqrt{r_p^2 + \pi^2}$. We integrate up to $\pi = 35\ h^{-1}$Mpc, which is large enough to include most correlated pairs. Figure 3.6 shows the results, both in the mock catalog (top panel) and in the SDSS (bottom panel). Now, the agreement with the analytical model is very good, suggesting that our analytic treatment of redshift-space distortions is inadequate. Once again, the inflection at the scale of the patch size for the underdense sample is caused due to the transition from one type of 2-halo term to the other.

Both for $\xi(s|\delta_s)$ and $w_p(r_p|\delta_s)$ the differences between the two environments are dramatic– they are measured with high statistical significance. Nevertheless, the analytic model, which only incorporates those correlations with environment which arise from the correlation between halo mass and environment, provides an excellent description of the measurements. This leaves little room for other environmental effects.

3.3 DISCUSSION AND CONCLUSIONS

One of the luxuries of the latest generation of large-scale sky surveys is that they contain sufficiently many objects that one can study subsamples of galaxies divided up in various ways. Here, we have focused on the clustering of galaxies in a volume limited sample drawn from the SDSS, and studied how the clustering of these galaxies depends on environment. We find that galaxies in dense regions are considerably more strongly clustered than those in less dense regions (Figures 3.5 and 3.6).

This is perhaps not so surprising—after all, a dense region is one in which many galaxies are crowded together. What is more surprising is that this dependence on environment is very well reproduced by numerical (Section 3.2.2) and analytic (Section 2.3.2.3) models in which the entire effect is due to the fact that galaxy properties correlate with the masses of their parent halos, and massive halos preferentially populate dense regions. Hierarchical models make quantitative predictions for this correlation between halo mass and environment, and so the agreement between our models and the measurements provides strong support for such models. In this respect, our results are consistent with those of Mo et al. (2004), Kauffmann et al. (2004), Blanton et al. (2004), Berlind et al. (2005) and Skibba et al. (2006); this is reassuring, since our methods are very different.

Our test of environmental effects is particularly interesting in view of recent work showing that, at fixed mass, haloes in dense regions form earlier (Sheth & Tormen 2004), and that this effect is stronger for low mass haloes (Gao, Springel & White 2005; Harker et al. 2006; Wechsler et al. 2006). Such a correlation is not part of our analytic model, nor is it included in our mock catalogs.

The agreement between our models and the measurements has an important consequence:

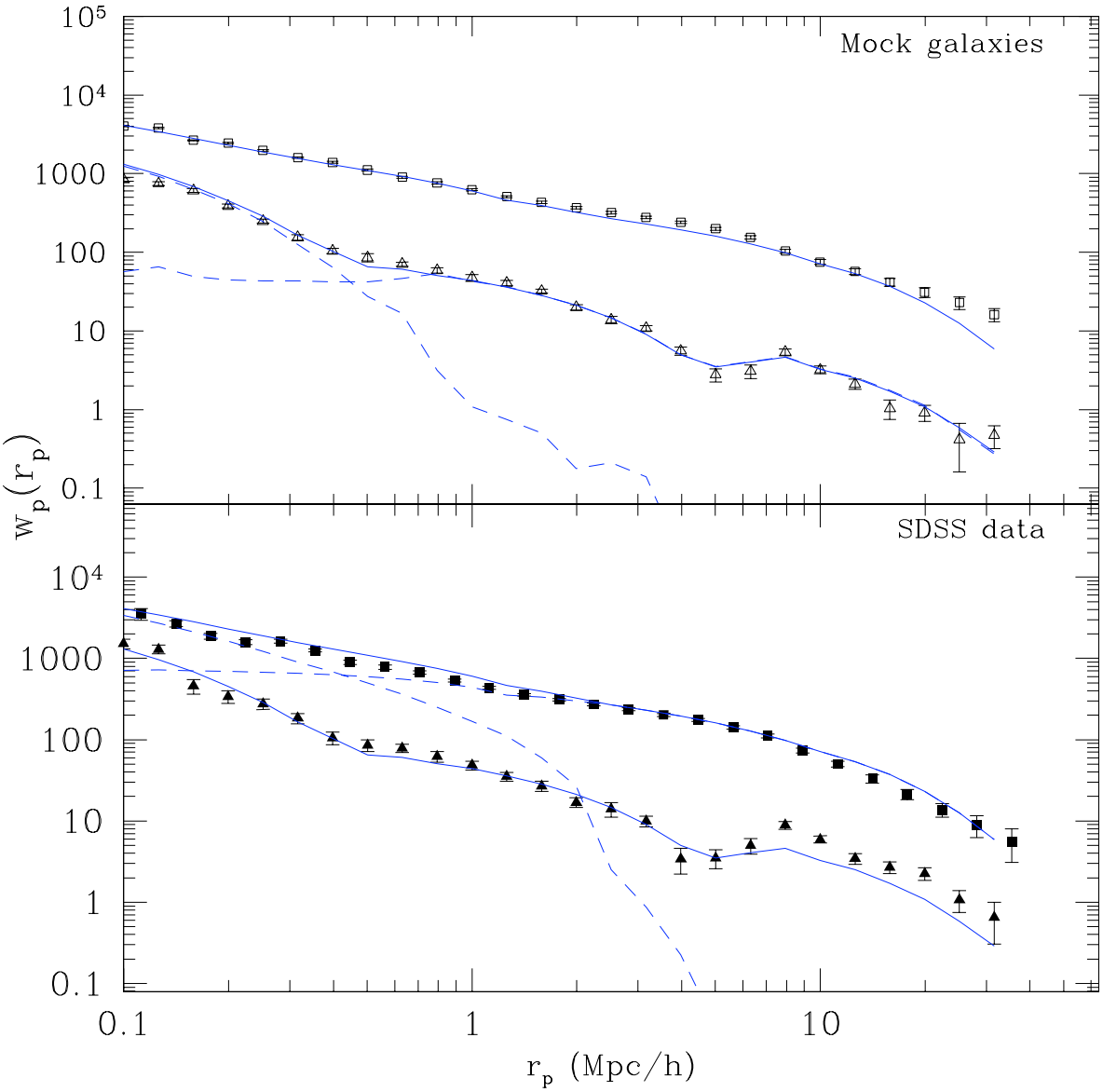

Figure 3.6: Similar to Figure 3.5, but now for the projected galaxy correlation function. The upper panel shows measurements in the mock catalog, and the lower panel is for the SDSS. Symbols and line-styles are the same as for previous Figure 3.5.

unless care is taken to study a population at fixed halo mass, our results indicate that observed correlations between gastrophysical effects (e.g. ram pressure stripping, strangulation, harassment) and environment are dominated by the fact that these effects also correlate with halo mass, and halo mass correlates with environment. More distant samples will allow us to study if these trends evolve.

4.0 THE AMPLITUDE OF FLUCTUATIONS AND THE ENVIRONMENTAL DEPENDENCE OF GALAXY CLUSTERING

4.1 INTRODUCTION

In Chapter 2 we described how the clustering of galaxies can be used to test the assumption that the correlations between galaxy properties and their environments are *entirely* a consequence of the correlations between haloes and their environments— the densest regions are populated by the most massive haloes. In Chapter 3 we showed that measurements of the environmental dependence of clustering in the Sloan Digital Sky Survey (SDSS) provide strong support for this assumption—the halo mass function in dense regions is top-heavy.

To perform the test, we used the currently popular flat ΛCDM background cosmology, with parameters $(\Omega, h, \sigma_8) = (0.3, 0.7, 0.9)$. The least well-constrained of these parameters is σ_8, which describes the initial fluctuation amplitude. So it is interesting to ask if measurements of galaxy clustering provide strong constraints on σ_8.

Current interpretations of galaxy clustering are based on the halo-model (Cooray & Sheth 2002). In the halo-model, all galaxies are associated with dark matter halos, and the number of galaxies in a halo is a function of halo mass. The measured number density and two point correlation function constrain how galaxies populate halos. If σ_8 is changed but all other cosmological parameters are kept fixed, then the abundances of halos change (e.g. Sheth & Tormen 1999). In response to such a change, the halo model attempts to obtain the same galaxy number density and observed clustering signal by changing how galaxies populate halos. For values of σ_8 in the range of interest ($0.7 \leq \sigma_8 \leq 1.0$), and for the range of galaxy populations which are most easily studied in the SDSS, it is always possible to populate halos with galaxies in such a way that the two point correlation function $\xi(r)$ remains essentially

invariant (Zheng et al. 2002; Yoo et al. 2006). In practice, it is the projected quantity, $w_p(r_p)$, which is measured. Since it is simply an integral over $\xi(r)$, the discussion above implies that no useful constraint on σ_8 can be derived from measurements of $w_p(r_p)$ alone.

On the other hand, combining measurements of $w_p(r_p)$ with measurements of the redshift space quantity $\xi(s)$ may well provide useful constraints (e.g. Yang et al. 2003; Tinker et al. 2005). However, this comes at a cost: modeling $w_p(r_p)$ requires no information about galaxy peculiar velocities, whereas modeling $\xi(s)$ does. For this reason, if peculiar velocity independent constraints on σ_8 can be obtained, they will be interesting indeed. Yoo et al. (2006) predict that the ratio of the cross correlation function and the galaxy-galaxy correlation function provide constraints on σ_8. The main goal of this chapter is to study if the environmental dependence of clustering, $w_p(r_p|\delta)$ combined with the galaxy-dark matter cross correlation function $\xi_{gm}(r|\delta)$, can be used to devise such a test for σ_8.

That such a test is possible is not unreasonable: all other cosmological parameters remaining fixed, larger values of the rms fluctuation amplitude σ_8 imply a larger range of environments. So the difference between the densest and least dense regions, and the difference between halo abundances in dense and underdense regions in particular, increases with increasing σ_8. Therefore, one might expect the environmental dependence of clustering to yield useful information about σ_8. For example, Tinker (2006) has shown that the conditional matter probability distribution function (PDF) reproduces distinct effects of varying σ_8, as shown in Figure 4.1.

Section 4.2 describes the test and shows that, in fact, the environmental dependence of clustering could yield useful information about the value of σ_8. Section 4.3 summarizes our results.

4.2 σ_8 AND ENVIRONMENTAL TRENDS

4.2.1 Test using $w_p(r_p|\delta)$

It makes little sense to use environmental effects to constrain σ_8 if a given value of σ_8 is already excluded because an acceptable fit to both the observed number density $n_{\rm gal}$ and

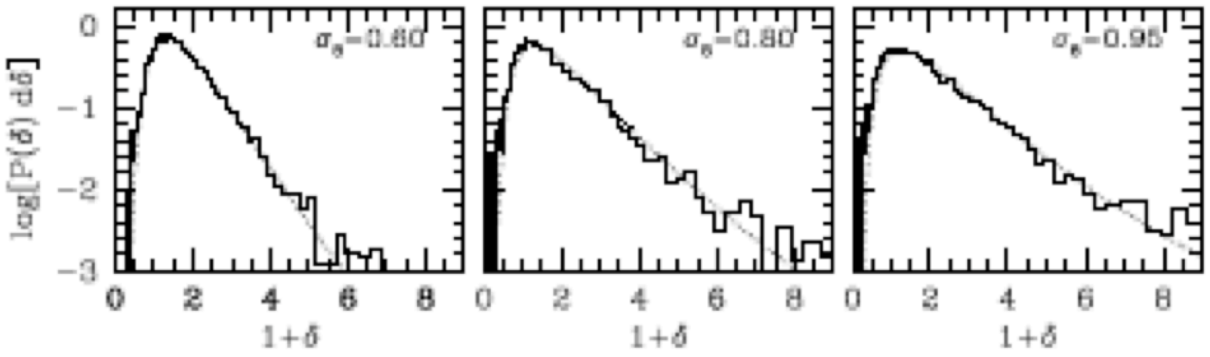

Figure 4.1: Comparison between the conditional matter PDFs measured from the simulations (solid histograms) and the analytic model of Tinker (2006) (dotted lines). The PDF for fixed mass ($10^{12.3}h^{-1}M_{\odot}$) and radius (10 h^{-1} Mpc), but changing σ_8 are shown. Figure has been taken from Tinker (2006).

clustering strength $\xi(r)$ cannot be found. Therefore, when performimg the environmental test, we first ensure that, given σ_8, the model can reproduce both $n_{\rm gal}$ and $\xi(r)$.

The halo model provides a simple way to see how this is accomplished. Current parameterizations of how galaxies populate halos assume that the mean number of galaxies in halos is a monotonic function of halo mass that decreases sharply at sufficiently low halo masses. A particularly simple, but nevertheless successful model, sets

$$\langle N_{\rm gal}|m\rangle = 1 + \left(\frac{m}{M_1}\right)^{\alpha} \qquad \text{if} \quad m \geq M_{\rm min}, \tag{4.1}$$

and $\langle N_{\rm gal}|m\rangle = 0$ otherwise. Suppose that parameters have been found which provide a good fit to the observed abundance and clustering of a given galaxy sample for a given fiducial value of σ_8. Models with smaller σ_8 have fewer massive halos. If the number of galaxies in massive halos is unchanged relative to the fiducial case, then to keep the same number density of galaxies one must populate lower mass halos (since these are more abundant) than in the fiducial case. However, this will lead to disagreement with the observed clustering strength, because low mass halos in a low σ_8 model will be much less clustered than higher mass halos in a larger σ_8 model. Thus, in addition to populating lower mass halos, low

Table 4.1: Dependence on σ_8 of halo-model parameters for galaxies with $M_r < -21$.

σ_8	$M_{\rm min}/h^{-1}M_\odot$	$M_1/h^{-1}M_\odot$	α
0.7	4.46×10^{12}	7.98×10^{13}	1.40
0.8	4.71×10^{12}	9.58×10^{13}	1.31
0.9	4.85×10^{12}	1.11×10^{14}	1.25
1.0	4.95×10^{12}	1.23×10^{14}	1.19

σ_8 models must put more galaxies in massive halos, since doing so will raise the clustering strength back up to the observed level. Table 4.1, from Yoo et al. (2006), shows the result of this exercise for SDSS galaxies with $M_r < -21$.

In the last chapter we saw that the halo model parameters of Zehavi et al. (2005) were in excellent agreement with the measurements for galaxies with $M_r < -21$, this is shown in Figure 4.2. The following Figure 4.3 shows the same measurements in comparison to the halo-model based prediction corresponding to $\sigma_8 = 0.9$ from Table 4.1 of Yoo et al. (2006). It is interesting to note that the measurements for this sample of galaxies are significantly different from the model predictions, especially for the underdense sample, despite the good agreement for the full sample. This suggests that in an attempt to match the model with measurements in underdense and dense regions, we could possibly obtain stringent fits on the HOD parameters.

Figure 4.4 shows that the parameter choices from Yoo et al. (2006) result in similar correlation functions $w_p(r_p)$, all of which are in good agreement with the measurement for the full sample. Having verified that these parameter values provide a good description of the SDSS clustering signal as σ_8 is varied, we are ready to perform our test. Namely, we now check if these same parameter values can also provide a good description of the environmental dependence of clustering. Figure 4.4 compares the halo-model based predictions for different values of σ_8 described in detail in Chapter 2, using the halo-model parameters given in

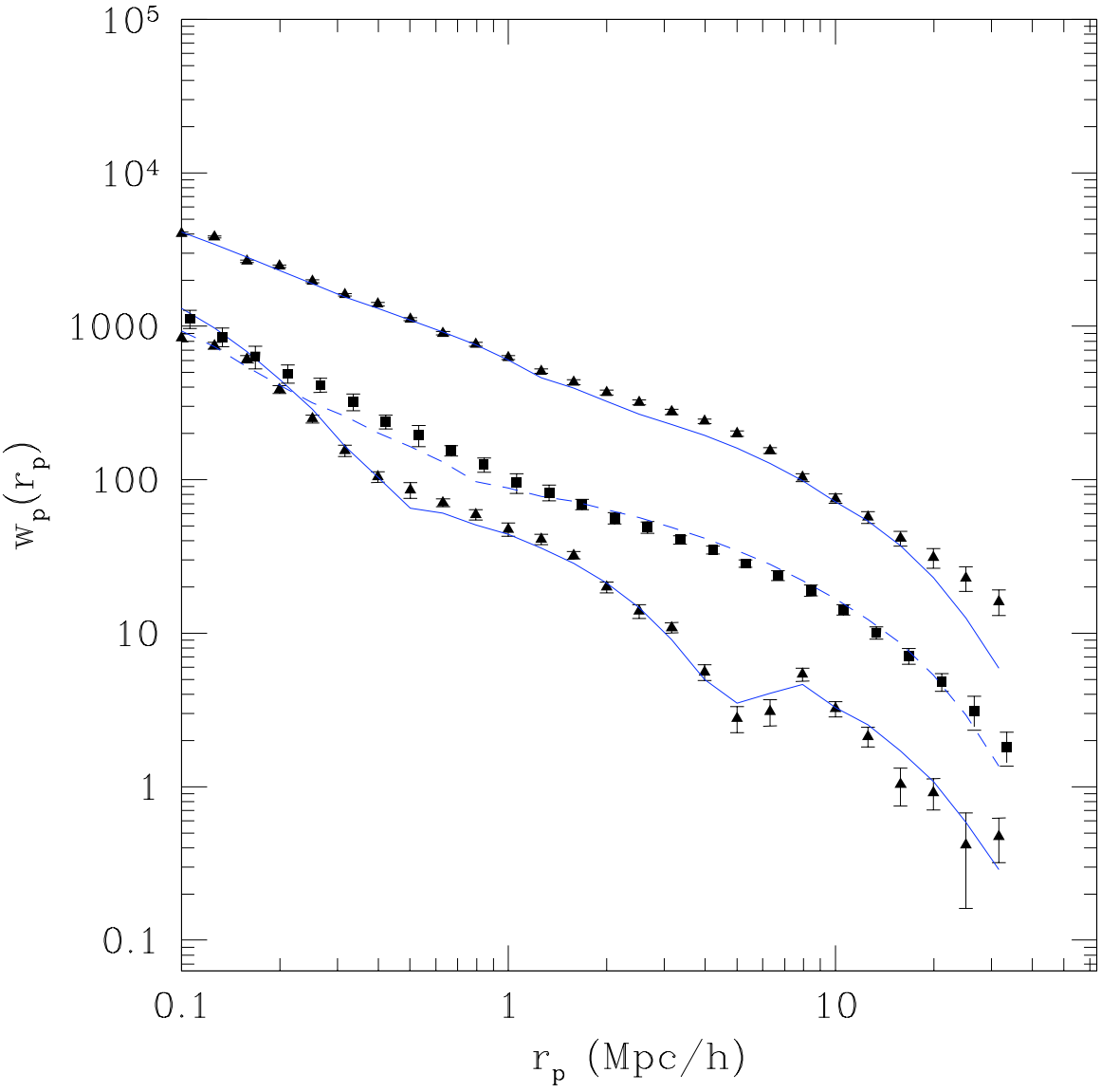

Figure 4.2: The halo model predictions associated with the HOD parameters from Zehavi et al. (2005) in comparison to the measured projected correlation function for $\sigma_8 = 0.9$ and for galaxies with $M_r < -21$. The upper (lower) triangles represent the measurements for the one-third overdense (underdense) samples, and the squares show the full sample. The associated solid and dashed curves show the corresponding model predictions.

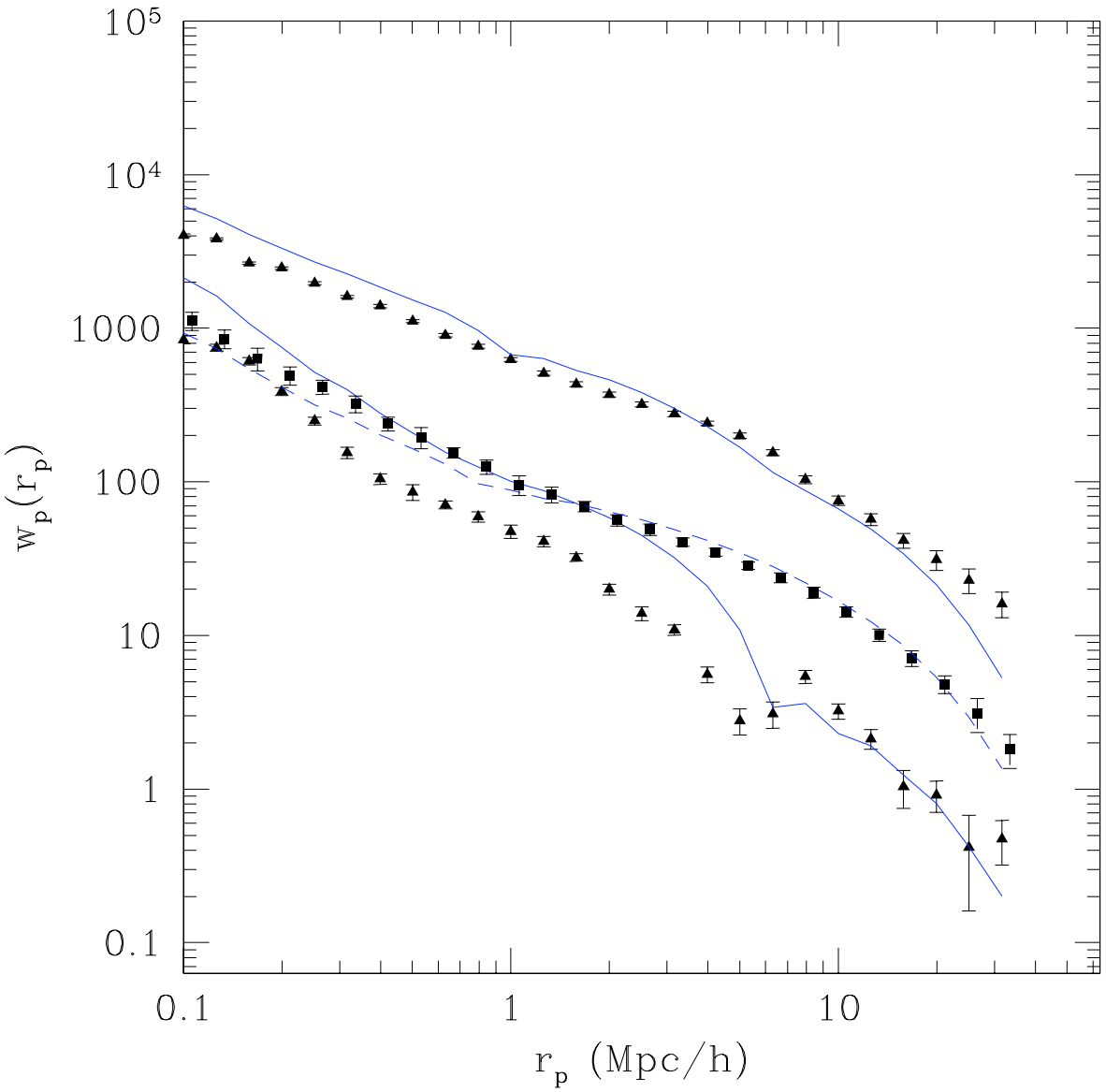

Figure 4.3: The halo model predictions associated with the HOD parameters from Yoo et al. (2006) in comparison to the measured projected correlation function for $\sigma_8 = 0.9$ and for galaxies with $M_r < -21$. The symbols represent the same measurements as in Figure 4.2. The associated solid and dashed curves show the model predictions.

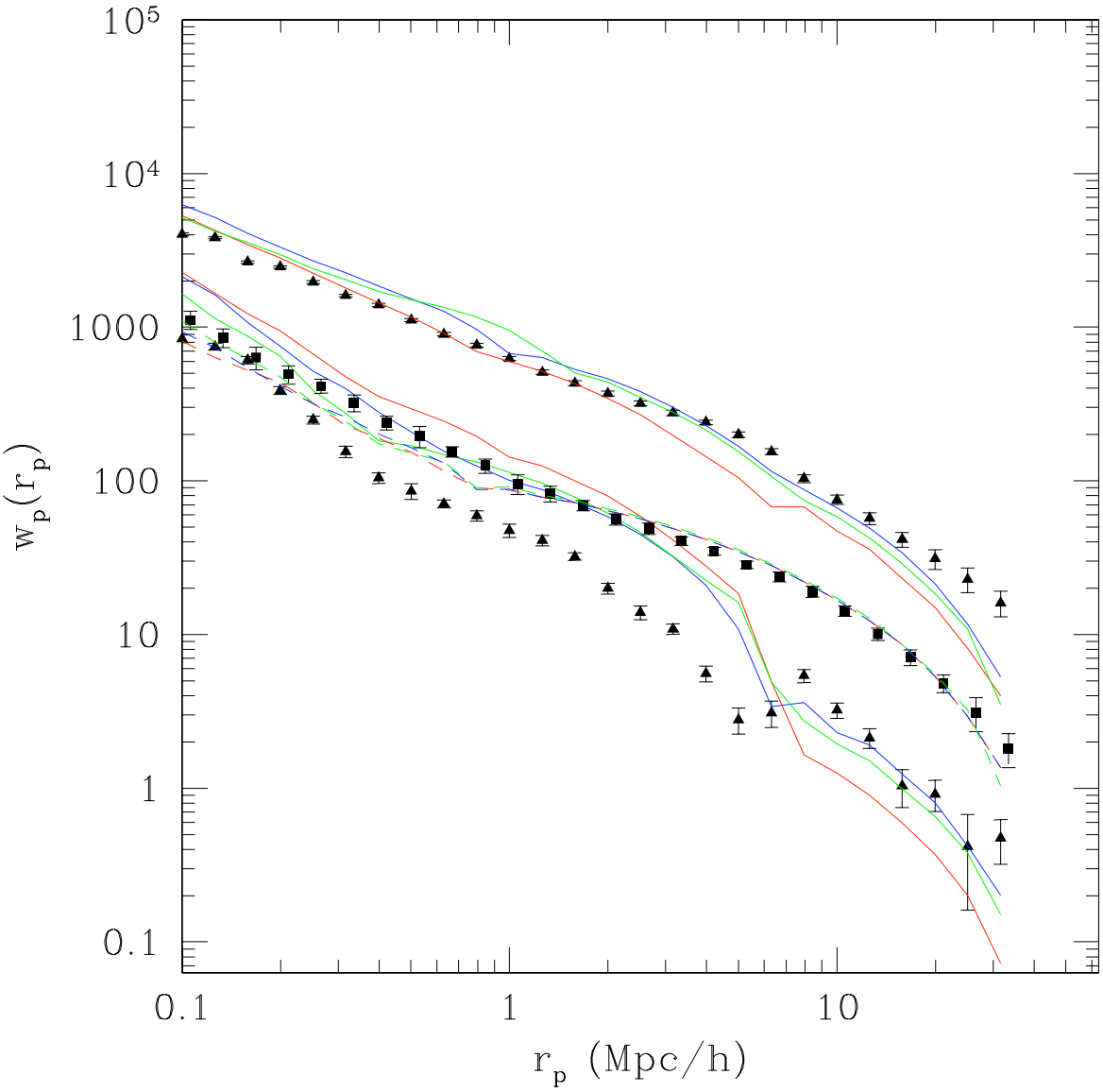

Figure 4.4: Comparison of the measured $w_p(r_p)$ (given by the symbols) with halo-model fits as σ_8 is varied for galaxies with $M_r < -21$. The square symbols show results for the full sample, whereas the upper (lower) triangle symbols show results for the one-third overdense (underdense) samples. The red, blue, green and black dashed lines are for $\sigma_8 = 0.7$, 0.8, 0.9 and 1.0 respectively. The solid lines show the projected correlation function for the dense (upper curves) and underdense samples (lower curves), with the same color coding as before.

Table 4.1. Here δ_s is defined by counting N_8, the number of galaxies in an $8h^{-1}$Mpc redshift-space sphere centered on each galaxy. The top and bottom sets of measurements show the correlation functions of the galaxies with the largest and the smallest values of N_8 (the top and bottom 33%). In both cases, the curves for the various σ_8 values are quite different. This means that for a sufficiently large number of data, useful constraints on σ_8 could possibly be obtained, provided the error bars are smaller than the difference between two of the curves for the same sample. Given that the results shown in Figure 4.4 were obtained with a sample of 75,000 galaxies, an improvement of a factor of a few in the estimated uncertainties could be obtained by utilizing a sample of a few 100,000 data points (simply based on Poisson statistics).

4.2.2 Test using the galaxy-dark matter cross correlation function

With simple modifications to the halo model based description described in Chapter 2 we can obtain the galaxy-dark matter cross correlation function. To bring out the salient features, we present the model below.

Given that the density profile of a halo depends on its mass, but not on the surrounding environment, the one-halo term is

$$P_{gm\,1h}(k|\delta) = \int_{M_{min}}^{M_{max}} dM\, n(M,V) \times \int_{M_{min}}^{M} dm\, N(m|M,V) \left(\frac{m}{\bar{\rho}_\delta}\right) \left(\frac{1+\langle N_s|m\rangle\, u(k|m)}{\bar{n}_{\delta-gal}}\right) |u(k|m)|. \tag{4.2}$$

where $\bar{n}_{\delta-gal}$ is given in Equation 2.15.

The two-halo term has two types of contributions given by the 2-halo–1-patch and 2-halo–2-patch terms. The 2-halo-2-patch term is well approximated by

$$\frac{P_{gm\,2h-2p}(k|\delta)}{P_{\rm Lin}(k|R_p)} = \left[\int_{M_{min}}^{M_{max}} dM\, n(M,V)\, B(M,V) \int_0^M dm\, N(m|M,V) \frac{m}{\bar{\rho}_\delta}\, u(k|m)\right] \times \left[\int_{M_{min}}^{M_{max}} dM\, n(M,V)\, B(M,V) \int_{M_{min}}^{M} dm\, \frac{1+\langle N_s|m\rangle\, u(k|m)}{\bar{n}_{\delta-gal}} N(m|M,V)\right] \tag{4.3}$$

where $P_{\rm Lin}(k|R_p)$ denotes the power spectrum associated with setting the linear theory correlation function to -1 on scales smaller than the diameter of a patch $2R_p$. This truncation

has little effect on small $kR_p \ll 1$, where $P_{\rm Lin}(k|R_p) \approx P_{\rm Lin}(k)$. The factor $B(M,V)$ describes the bias associated with the clustering of the patches, and depends on the abundance of such patches.

The 2-halo–1-patch term can be written as

$$P_{gm\,2h-1p}(k|\delta) = \int_{M_{min}}^{M_{max}} dM\, n(M,V) \times \int_0^M dm_1 \left(\frac{m_1}{\bar{\rho}_\delta}\right) |u(k|m_1)| \times \int_{M_{min}}^{M-m} dm_2 \left(\frac{1+\langle N_s|m_2\rangle u(k|m_2)}{\bar{n}_{\delta-gal}}\right) \times N(m_1,m_2|M,V)\; U(k|m_1,m_2,M)^2. \tag{4.4}$$

Here U denotes the normalized Fourier transform of the spatial distribution of m_1 and m_2 haloes within a patch and is given in Equation 2.12.

The other term, $N(m_1,m_2|M,V)$, denotes the average number of haloes of mass m_1 and m_2 in patches V which contain total mass M (see eq. 2.13).

When $M \gg m$ and $V \gg v_m$, as it is in large overdense regions, then $N(m_1,m_2|M,V) \approx N(m_1|M,V)\,N(m_2|M,V)$ i.e, $N(m_1,m_2|M,V)$ is well approximated by the product of the individual mean values. In this case,

$$\frac{P_{gm\,2h-1p}(k|\delta)}{P_{\rm Lin}(k) - P_{\rm Lin}(k|R_p)} \approx \int_{M_{min}}^{M_{max}} dM\, \frac{n(M,V)}{M/\bar{\rho}} \times \left[\int_0^M dm_1\, N(m_1|M,V)\, b(m_1) \left(\frac{m_1}{\bar{\rho}_\delta}\right) |u(k|m_1)|\right] \times \left[\int_{M_{min}}^M dm_2\, N(m_2|M,V)\, b(m_2) \left(\frac{1+\langle N_s|m_2\rangle\, u(k|m_2)}{\bar{n}_{\delta-gal}}\right)\right]. \tag{4.5}$$

In underdense regions, however, simply using the product of the individual mean values is expected to be a bad approximation.

Figure 4.5 shows the model for different values of σ_8 showing that once again useful constraints on σ_8 can be obtained, particularly in underdense regions.

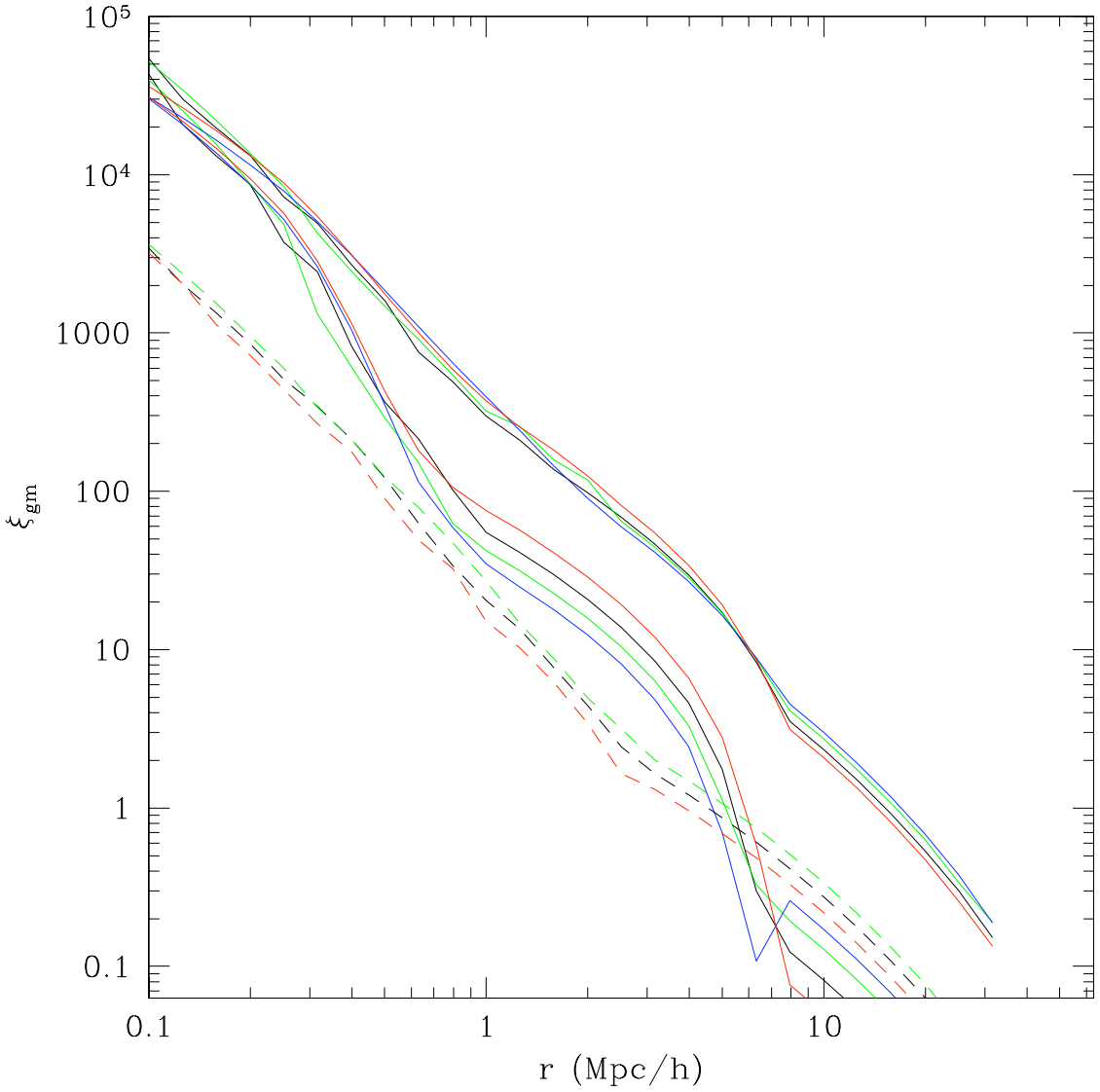

Figure 4.5: The environmental dependent galaxy-dark matter cross correlation function for different values of σ_8 and galaxies with $M_r < -21$. The red, black, green and blue curves are for $\sigma_8 = 0.7$, 0.8, 0.9 and 1.0 respectively. The solid lines depict the model for underdense regions (lower curves) and overdense regions (upper curves). The dashed lines show ξ_{gm} for the full sample.

4.3 DISCUSSION

We tested the hypothesis that the environmental dependence of clustering offers a test of the amplitude of fluctuations, σ_8. The method in Section 4.2.1 has the attractive feature of not requiring peculiar velocity information. The different sets of curves in Figure 4.4 indicates that the environmental dependence of clustering can provide useful constraints on σ_8. It can be seen that in underdense regions the ratio $w_p(r_p = 1)/w_p(r_p = 10)$ is a strong function of σ_8, but only a weak function in dense regions. The galaxy - dark matter cross correlation function in Figure 4.5 shows a similar result.

Why is this so? Table 4.1 suggests that, when σ_8 is smaller, galaxies populate a wider range of halo masses ($M_{\rm min}$ is smaller), and they preferentially populate higher mass halos (α is larger). Thus, the smaller spread in large-scale environments associated with smaller σ_8 is compensated-for by increasing the spread in small-scale environments. Likewise, when σ_8 is larger, then the larger spread in environments is compensated-for by decreasing the range of halo masses which galaxies populate, and by decreasing the relative number of galaxies in larger mass halos. As a result, when σ_8 is small and turning our attention to the underdense regions dominated by the low mass haloes, the larger number of galaxies in these haloes bumps up the correlation function at small scales. On the other hand, the correlation function is lower at large scales due to the effectively smaller bias factor of the low mass haloes. This causes the environmental dependence of clustering to be a mildly varying function of σ_8.

5.0 SUMMARY AND FUTURE PROSPECTS

5.1 SUMMARY OF RESULTS

In the work presented in this thesis we have undertaken detailed studies of the environmental dependence of dark matter and galaxy clustering. This has involved statistical studies of N-body simulations, mock galaxy catalogs, and Sloan Digital Sky Survey data along with the development of an analytical framework. The correlation function was used as the proxy to quantify clustering. The main results and conclusions can be summarized as follows:

5.1.1 Results from simulations and data

- Dark matter N-body simulations: Dark matter samples from the GIF (256^3 particles in 141 Mpc/h cubic box) and VLS (512^3 particles in a 479 Mpc/h cubic box) ΛCDM simulations were used to obtain dense and underdense subsamples. The dense subsample shows stronger clustering than the underdense sample on scales greater than a few hundred kpc. At large scales the amplitude of the dense correlation function is approximately a factor 10 larger than that of the underdense sample, suggesting that the halo mass function is top-heavy in dense regions. Around a few hundred kpc there is an interesting crossing over, with the underdense sample exhibiting stronger clustering at scales smaller than this. With the help of a toy model we have shown that this is due to the overdensity relative to the local background and is not caused by the mass dependence of the concentration.
- Mock galaxy samples: The halo catalogs of the GIF and VLS simulations were used to generate mock galaxy samples with the help of the HOD model in real space (Chapter

2), and in redshift space (Chapter 3). Once again, the dense subsample of galaxies shows a stronger correlation function than the underdense sample at all scales. At scales of the order of the patch size (5-8 Mpc/h), the correlation function of the underdense sample exhibits a characteristic dip, and its overall behavior is characterized by 3 regimes with different slopes.

- SDSS data: Volume limited samples of the SDSS for a bright set of galaxies ($M_r < -21$) were used. The results obtained were consistent with what was found for the mock galaxy samples with the main results being the measurable offset of the dense and underdense samples at all scales.

5.1.2 Results from the analytical framework

We presented a novel approach to modeling the environmental dependence of clustering within the framework of the halo model. The model was shown to accurately represent the clustering in real (Chapter 2) and redshift space (Chapter 3). The presence of 3 distinct regimes, prominently displayed in the clustering of the underdense sample, can be successfully interpreted in terms of the 1 halo (at small scales), 1 patch- 2 halo (intermediate scales) and 2 patch- 2 halo (large scales) terms. The traditional 2 halo term is split up into two terms, one arising from pairs of particles in different halos lying within the same patch, and the other from pairs of particles lying in different patches.

The agreement between the analytical model and measurements from simulations and data provides strong support for the hierarchical structure formation scenario. Within this scenario, density fluctuations at different scales are correlated. In turn, the correlations between gastrophysical effects (such as ram-pressure stripping, harassment, strangulation etc.) and environment are dominated by the halo mass function. The analysis has verified two main assumptions of the hierarchical model, i.e. that the halo mass function is top-heavy in dense regions, and that at fixed mass the halo properties are independent of environment.

Finally, in Chapter 4, we show that for sufficiently large data samples, our model could place reasonable constraints on the amplitude of fluctuations σ_8.

5.2 FUTURE PROSPECTS

The present work has been carried out at redshift close to zero. In the future we plan to extend our studies to address the evolutionary behavior of the environmental dependence of clustering with the help of high-z surveys such as the COSMic evOlution Survey (COSMOS, Scoville et al. 2006), Vimos VLT Deep Survey (VVDS, Le Fevre et al. 2005) etc. The VVDS team has shown the strong redshift dependence of the color-density relation, with blue galaxies predominantly occupying high density regions at high redshifts ($z > 1.2$) compared to red galaxies, the opposite of what is seen at low redshifts (Cucciati et al. 2006). Also, the bias is shown to have a strong evolutionary behavior (Marinoni et al. 2005). Our analysis will help ascertain whether these effects are entirely statistical in nature or whether gastrophysics starts playing a role at high redshifts.

Furthermore, the model will help to predict dark matter distributions which can also be compared to gravitational lensing measurements (using the CFHTLS [1] etc.). This would allow for a mapping of dark matter in dense and underdense regions and further help to constrain models of galaxy formation.

Finally, it is becoming increasingly more important to constrain cosmological parameters with the help of high precision measurements from galaxy surveys such as 2dFGRS, SDSS, VVDS, Deep2, DES (Colless et al. 2001; Davis et al. 2003). This thesis shows that by restricting attention to luminous galaxies, the analysis of future surveys is simplified by the assumption that the environmental trends are almost entirely due to the correlations between halo mass and environment.

[1] http://www.cfht.hawaii.edu/Science/CFHLS/

APPENDIX A

THE COMPOSITION OF THE UNIVERSE

The universe is presently believed to be composed of 73% dark energy, 23% non-baryonic dark matter, and 4% ordinary baryonic matter.

A.1 DARK ENERGY

Dark energy is thought to be an energy field which is homogeneous with low density (of the order of $10^{-29} g/cm^3$), and it interacts only through something like a repulsive gravitational force. The first evidence for its existence came from observations of distances to Type IA supernovae (Riess et al. 1998; Perlmutter et al. 1999). Along with measurements of the CMB, gravitational lensing and large scale structure, this has led to the currently favored Lambda Cold Dark Matter (ΛCDM) cosmological model. Two competing theoretical models for Dark Energy are the cosmological constant and quintessence.

The cosmological constant was first proposed by Einstein in 1917 as an inclusion into his theory of General Relativity in order to allow for a static universe, as the tensor structure of the field equations admitted the presence of a linear term in the metric. Friedmann, a Russian mathematician, believed that this was an improper fix and put forth the Big Bang theory in 1922. When Hubble showed evidence for the expansion of the universe through studies of nearby galaxies in 1929, Einstein regretted proposing the cosmological constant and went on to make the all-too famous statement that it was his 'greatest mistake'. In recent

years, this constant has been used to explain the accelerated expansion of the universe by virtue of its negative pressure (Riess et al. 1998). However, quantum field theories predict a cosmological constant from quantum vacuum energy that is about 120 orders of magnitude too large, which has led to a major outstanding problem (Straumann 1999).

Quintessence is thought to arise from particle-like excitations in a type of dynamical field (Caldwell, Dave & Steinhardt 1998). It can vary in space and time and predicts a slower expansion for the universe than the cosmological constant. When the universe evolved from a radiation dominated to a matter dominated phase, quintessence took precedence over the dark matter density, thereby causing dark energy to dominate. Special cases of dark energy are phantom energy (Caldwell 2002), and k-essence (Armendariz-Picon et al. 2000).

A.2 DARK MATTER AND GALAXIES

The first evidence for dark matter came from measurements of clusters of galaxies in 1933. The Swiss astrophysicist Fritz Zwicky took measurements of the Coma cluster and found velocities of galaxies about 10-100 times larger than expected (summary of results in Zwicky 1937). Applying the virial theorem he was able to compare his mass estimate with one based on the total number and brightnesses of galaxies, finding a mass that was 400 times more than what was expected, if only luminous matter contributed to the mass. This led him to postulate that there was some 'missing mass' that gives rise to the observed high velocities of the galaxies on the edge of the cluster. However, this was not considered very strong evidence as the data could be misleading due to foreground galaxies which were not part of the cluster or were leaving the cluster. Stronger evidence came at the beginning of the 1970s when Vera Rubin observed that most stars in spiral galaxies orbit at roughly the same speed, giving rise to the galactic rotation problem. In other words, the galactic rotation curve (i.e. the velocity of rotation versus the distance from the galactic center) is constant at large radii (several kpc). If the rotation curve were only due to the visible matter, it should fall off with increasing distance from the galactic center instead of remaining constant. Our own Milky Way galaxy is estimated to have roughly 10 times as much dark matter as visible matter.

In 2005, an international team of researchers led by a group at Cardiff University claimed

to have found the first dark matter galaxy. This galaxy, VIRGOHI21, lies roughly 50 million light years away in the Virgo cluster (Minchin et al. 2005). Using radio observations they found that this particular galaxy contains no stars but a lot of hydrogen gas. The dark matter mass content is predicted to be 1000 times that of the ordinary matter, even though this galaxy is 1/10th the size of the Milky Way. Further confirmation of this dark galaxy and others like it would provide strong evidence in favor of dark matter.

A.2.1 But, what is dark matter?

Like dark energy, we do not know the nature of the elusive dark matter. It is only detectable through its attractive gravitational interaction, and is difficult to find as it does not seem to interact with the electromagnetic force. Nonetheless, possible dark matter candidates include:

Ordinary matter - planets, dim or failed stars, neutron stars, black holes. Very low mass stars (known as brown dwarfs), could be found in large amounts along with Jupiter-sized and lower mass planets. But, these objects do not appear to exist in numbers needed to explain the missing mass (Bahcall et al. 1994). A star that has finished its hydrogen fuel ends its life as a white dwarf slowly cooling to become a black dwarf. However, the timescale to cool to a black dwarf is many times longer than the age of the universe. Therefore, such stars do not account for the mass in dark matter (Chabrier et al. 1996). Gravitational lensing data rules out the possibility of galaxy-sized black holes (Turner 2000). Also, MACHO microlensing experiments show that there are not enough of the aforementioned dark objects (Alcock et al. 1997). Furthermore, big bang nucleosynthesis measurements predict that there are not enough baryons to produce the vast amounts of dark matter (Burles & Tytler 1998).

Non-Baryonic or exotic matter - Matter that is not composed of quarks and leptons, i.e. 'exotic matter', is the next type of candidate. Neutrinos are an interesting candidate as they have very low masses and do not interact via the electromagnetic force nor the strong force, and hence are extremely difficult to detect. These particles travel with relativistic speeds, so they would give rise to what is called Hot Dark Matter (HDM). However, current bounds on the neutrino mass show that they only make up a fraction of the dark matter

(Turner 2000). Furthermore, models using HDM do not give rise to the types of structures seen in the universe (Davis et al. 1985).

Currently, the most favored scenario for structure formation is the Cold Dark Matter (CDM) model, where the term cold is used because the speed of the particles are non-relativistic. Comparison of CDM simulations with observational data have shown consistency, lending strong support to the CDM paradigm (Davis et al. 1985). The best motivated particles for the popular CDM model are particles that were already invoked by particle physicists for other reasons: axions and WIMPs (weakly interacting massive particles) (Sadoulet 1999). Axions are particles invoked to resolve the strong CP violation problem (the violation of the combined conservation laws associated with charge conjugation 'C' and parity 'P' by the weak nuclear force). WIMPs instead are predicted by supersymmetry (SUSY) theories, the most favored candidate being the lightest SUSY particle, the neutralino (Sumner 2002). Even though the ΛCDM (CDM universe with a cosmological constant) paradigm has been highly successful at explaining observations, it too has its share of shortcomings. One of them is the prediction of an excess of substructure (Moore et al. 1999), but solutions to this problem have been suggested (Avila-Reese et al. 2001; Alam et al. 2002).

Alternative theories - Rather than having matter explain the missing mass, the alternative is to obtain a stronger gravitational force at large distances. The Newtonian approximation to the force has been successfully tested at small scales, of the order of our solar system. Some argue that assuming that the same law applies at even larger scales is presumptuous. Instead, by invoking Modified Newtonian Dynamics (MOND) proposed by Milgrom in 1983, gravity could become stronger at large distances. The construction of a relativistic MOND theory however has been problematic and is unable to be reconciled with gravitational lensing measurements (Zhytnikov & Nester 1994; Edery 1999), although solutions to these problems have been proposed in Bekenstein (2005).

A.2.2 The SDSS

This thesis made comparisons between the theoretical models and data from the Sloan Digital Sky Survey (SDSS). The SDSS is one of the most ambitious astronomical survey

projects to be undertaken, i.e. the mapping of one-quarter of the visible sky. It will measure the positions, apparent brightnesses, and colors for more than 100 million celestial objects; redshifts will be obtained for about a million galaxies and 100,000 quasars, allowing for a wide as well as deep mapping. The survey uses a 2.5 meter telescope at Apache Point, New Mexico, specially equipped with two powerful instruments: a 120 Megapixel camera that images 1.5 sq degrees of area at a time, and a pair of spectrographs that can measure the spectra of more than 600 galaxies and quasars at a time.

So far the SDSS has completed its first phase of operations over 5 years, SDSS-I. The survey has imaged 8,000 sq degrees of sky in 5 bandpasses, detecting 200 million celestial objects and spectra for 675,000 galaxies, 90,000 quasars and 185,000 stars. The annual data releases were spread over the past four years, the latest release being DR-4 (Adelman-McCarthy et al. 2005). The SDSS results are the most accurate measurements of how density in the Universe fluctuates on scales of millions of light-years and is shown in Figure A.1. The data that has been used in Chapter 3 of this thesis comes from SDSS DR 4.

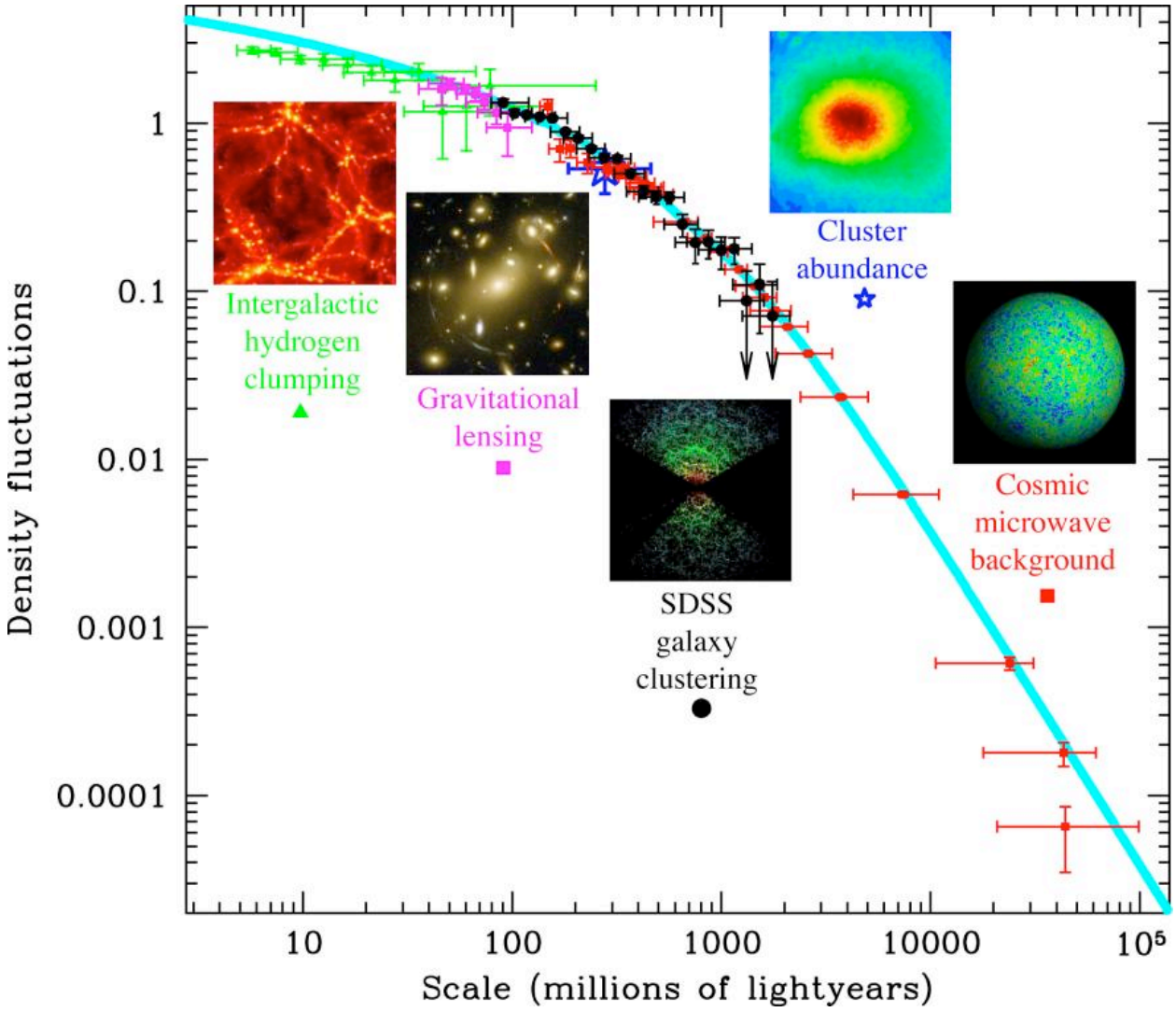

Figure A.1: The scales at which various cosmological probes operate. The blue curve shows the theoretical prediction for a universe composed of 5% atoms, 25% dark matter and 70% dark energy. Image credit: SDSS.

APPENDIX B

THE COSMOLOGICAL MODEL

B.1 THE THEORETICAL MODEL

The present standard cosmological model is based on the cosmological principle and the general theory of relativity. The cosmological principle tells us that the universe is seen to be homogenous and isotropic at large scales, whereas its evolution is governed by gravity and easily described by the general theory of relativity. The inhomogeneities that arise due to the structures at small scales can be explained by a perturbation field superimposed on the homogenous and isotropic fabric of space-time.

The assumptions of homogeneity and isotropy are equivalent to the requirement that the metric tensor takes the form of the Robertson-Walker metric (Weinberg 1972),

$$ds^2 = dt^2 - a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2(d\theta^2 + \sin\theta d\phi) \right], \tag{B.1}$$

where $a(t)$ is the cosmic scale factor, and k is a constant.

According to General Relativity, the geometry of the space-time background is related to its matter-energy content by virtue of the Einstein field equations,

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{B.2}$$

where $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ is the curvature tensor, R is the Ricci scalar, Λ is the cosmological constant and $T_{\mu\nu}$ is the stress-energy tensor.

In accordance with the cosmological principle, the stress-energy tensor of the universe is taken to be that of a perfect fluid (such that a comoving observer would observe an isotropic universe),

$$T_{\mu\nu} = pg_{\mu\nu} + (p + \rho)u_\mu u_\nu, \tag{B.3}$$

where p and ρ are the pressure and energy density measured by an observer in a locally inertial frame that happens to be moving with the fluid at the instant of measurement, while u_μ is the fluid four-velocity. Inserting the perfect fluid tensor and the Robertson-Walker metric into the Einstein field equations above, we can apply the stress-energy conservation law ($\delta_\nu T^{\mu\nu} = 0$) to get the Friedmann equations, which govern the dynamics of the cosmic scale factor $a(t)$:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2}; \qquad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{B.4}$$

The solutions of these equations are known as Lemaitre-Friedmann-Robertson-Walker (LFRW) universes, and describe the time evolution of the scale factor $a(t)$. The standard cosmological model of how the universe has evolved to its present state is well explained by the Big Bang theory, which is the generally accepted theory that the universe emerged from a dense and hot state about 13.7 billion years ago. There is strong observational evidence to support this theory described in the next section.

The LFRW universes have a set of free parameters easily described in terms of the density parameters and the expansion rate of the universe. By introducing the fractional matter density $\Omega_m(t) = (8\pi\rho(t)G)/(3H^2(t))$, the fractional density due to the curvature of space-time $\Omega_k(t) = -k/(a^2H^2(t))$, and the fractional vacuum energy density $\Omega_\Lambda(t) = \Lambda/(3H^2(t))$, then the first Friedmann equation in B.4 simplifies to;

$$\frac{\dot{a}}{a} = H_0 E(z) = H_0[\Omega_m(1+z)^3 + \Omega_k(1+z)^4 + \Omega_\Lambda]^{1/2}, \tag{B.5}$$

where z is the redshift ($1 + z = a_0/a(t)$), and H_0 and a_0 are the present-day values of the Hubble's constant and of the scale factor respectively. This shows that the evolution of a homogenous and isotropic universe depends only on the density parameters and Hubble's constant.

The three key models are highlighted in figure B.1: the Standard Cold Dark Matter (SCDM), dominated by non-relativistic non-baryonic matter, $(\Omega_m, \Omega_\Lambda) = (1,0)$; an open universe with no cosmological constant, $(\Omega_m, \Omega_\Lambda) = (0.3,0)$; and a flat universe with a Λ-term, the ΛCDM universe defined by $(\Omega_m, \Omega_\Lambda) = (0.3,0.7)$.

B.2 OBSERVATIONAL EVIDENCE

One important piece of supporting evidence for the standard model of cosmology comes from the expansion of the universe. Galaxies are seen to have a recessional velocity that is proportional to their distance from us, in accordance with Hubble's law (Hubble 1929). This aspect is fundamental for the determination of galactic redshifts. Theoretical predictions of the nucleosynthesis of light elements (such as deuterium, helium-3, and helium-4) within the first few seconds of the universe agree very well with current observations (see Holtmann et al 1996; Kurki-Suonio 2001, and references therein). For example, theory predicts about a quarter of the universe consists of helium-4, which is in excellent agreement with stellar abundance determinations.

At the time the universe was born, everything existed in the form of a very hot cosmic plasma; electrons, protons, neutrons, and radiation strongly coupled to each other. As the universe expanded and cooled the electrons and protons combined to form Hydrogen which led to the photons being 'freed' thereby defining a 'surface of last scattering'. This happened roughly 300,000 years after the Big Bang and is depicted in Figure B.2. Due to the vast expansion of the universe since then, these photons are now observed as a much cooler (about 2.7 K) background radiation in the form of microwaves. The temperature map of the cosmic microwave background (CMB) shows fluctuations of the order of one part in one-hundred thousand and allows scientists to study the early stages of structure formation.

Observational evidence for the CMB came in 1990 with findings from the Cosmic Background Explorer satellite (COBE), and later in Feb 2003 when the Wilkinson Microwave Anisotropic Probe (WMAP) produced a high-resolution map capturing the oldest light in the universe. The probe also determined the age of the universe to a high level of accuracy, with just a 1 % margin of error, placing the value at 13.7 billion years and found that the

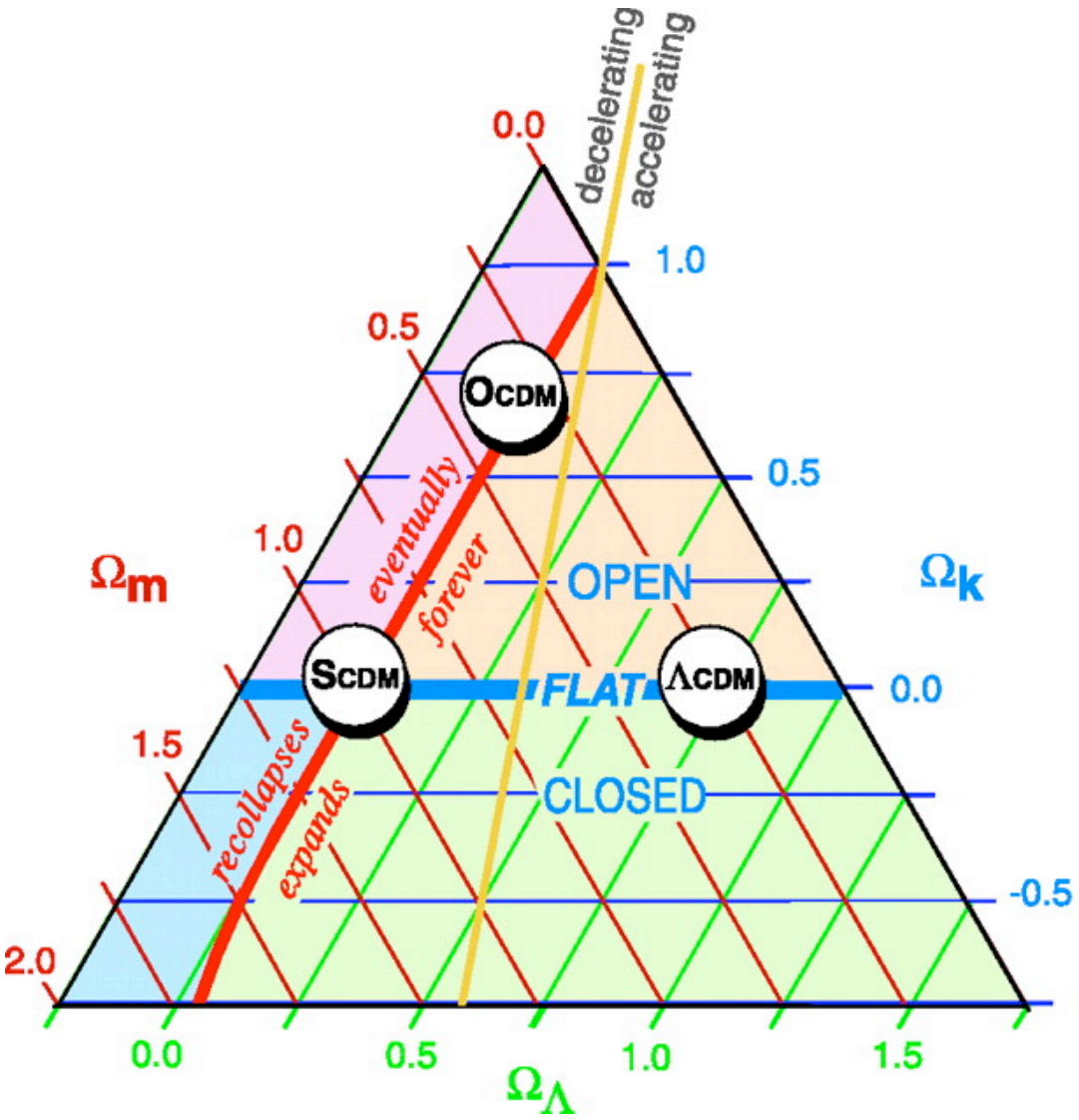

Figure B.1: The cosmic triangle. The horizontal line depicts a flat universe separating an open universe from a closed one. The red line separates a universe that will expand forever from another that eventually recollapses. The yellow line separates a universe with an expansion rate that is decelerating from one that is accelerating. Each point within the triangle satisfies the closure relation where the sum of the density parameters is one. The figure is taken from Bahcall (1999) and used for illustrative purposes.

first stars lit up about 200 million years after the Big Bang. The most recent WMAP observations as of March 2006 have obtained even tighter constraints on the basic parameters of the ΛCDM model, such as Hubble's constant (70 km/s/Mpc) and the spectral index (0.95). Hubble's constant determines the rate of expansion of the universe, whereas the spectral index determines how the density fluctuations change with scale.

The standard Big Bang model faces two major problems, for one it is unable to explain how fluctuations that started off as being causally disconnected (as seen in CMB observations) are seen today as a radiation map that is almost isotropic in all directions. Secondly, the flatness problem arises from the fact that the universe today is seen to be spatially very close to flat ($\Omega_m + \Omega_\Lambda + \Omega_k \simeq 1$). This would entail that any deviation from flatness must have been immeasurably small at the time of the Big Bang, otherwise even a small deviation would have grown very large by now due to the expansion of a universe filled with matter and radiation (Weinberg 1972). By invoking the theory of Inflation (Guth 1981) these two problems as well as several others are overcome by describing a phase of exponential expansion that took place when the universe was 10^{-35} seconds old and lasted for about 10^{-32} seconds. Through continuous expansion, the universe cooled and subsequently moved from a radiation dominated era about 10,000 years later to one where matter became the main component. The matter was able to interact gravitationally thereby becoming more and more clumpy. The initial tiny density perturbations were able to grow eventually leading to the formation of the first quasars and galaxies about a billion years after the Big Bang, and since then larger structures have been forming, such as galaxy clusters and superclusters. The different morphologies and distribution of these structures give valuable insights into the complex evolutionary behavior of the universe.

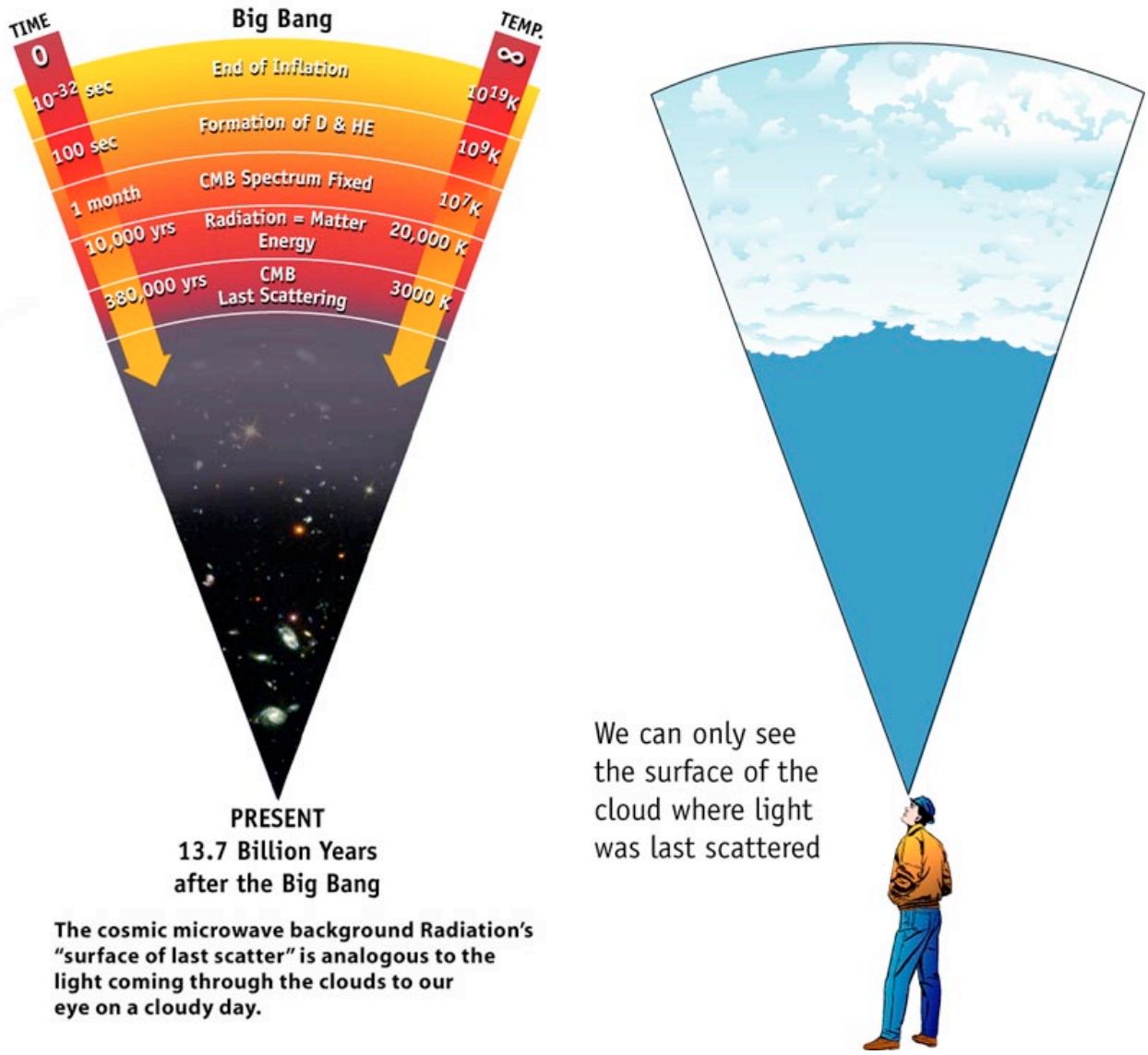

Figure B.2: The Cosmic Microwave Background and timeline of the evolution of the universe. (Image credit: NASA)

APPENDIX C

CONSISTENCY CHECKS

This Appendix shows that the expressions in Section 2.3.1 do reduce to the standard expressions in Section 1.4 upon averaging over all environments.

The mass density is

$$\begin{aligned}\bar{\rho}_\delta &= \int_{M_{min}}^{M_{max}} dM\, n(M,V) \int_0^M dm\, N(m|M,V)\, m \\ &\to \int_0^\infty dm\, m \int_m^\infty dM\, n(M,V)\, N(m|M,V) \\ &= \int_0^\infty dm\, \frac{dn(m)}{dm}\, m \int_m^\infty dM\, f(M,V|m) \\ &= \int_0^\infty dm\, \frac{dn(m)}{dm}\, m = \bar{\rho}\end{aligned} \tag{C.1}$$

when $M_{min} \to 0$ and $M_{max} \to \infty$ (this limit corresponds to averaging over all environments). In this limit, the one-halo term is

$$\begin{aligned}P_{1h}(k|\delta) \;\to\; & \int_0^\infty dM\, n(M,V) \\ & \times \int_0^M dm\, N(m|M,V) \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2 \\ =\; & \int_0^\infty dm \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2 \\ & \times \int_m^\infty dM\, n(M,V)\, N(m|M,V) \\ =\; & \int_0^\infty dm\, \frac{dn(m)}{dm} \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2.\end{aligned} \tag{C.2}$$

Similarly, the two halo term becomes

$$\begin{aligned}\frac{P_{2h-2p}(k|\delta)}{P_{Lin}(k|R_p)} \;\rightarrow\; & \left[\int_0^\infty dM\, n(M,V)\, B(M,V)\right. \\ & \left.\times \int_0^M dm\, N(m|M,V) \left(\frac{m}{\bar{\rho}}\right)\, u(k|m)\right]^2 \\ = \; & \left[\int_0^\infty dm\, \frac{m}{\bar{\rho}}\, u(k|m) \int_m^\infty dM\, N(m|M,V)\right. \\ & \left.\times \quad n(M,V)\, B(M,V)\right]^2. \end{aligned} \tag{C.3}$$

This can be simplified as follows. The halo overdensity is

$$\delta_h(m|M,V) = \frac{dN(m|M,V)/dm}{V\, dn(m)/dm} - 1 \equiv b(m|M,V)\, \delta. \tag{C.4}$$

When $V \gg 1$, then $\delta \ll 1$, and M is almost surely much larger than the typical halo mass, so $M \gg m$ for most values of m. In this limit,

$$\begin{aligned}\frac{dN(m,\delta_c|M,V)/dm}{V\, dn(m)/dm} &= (1+\delta)\, \frac{dn[m,\delta_c - \delta_0(\delta)]/dm}{dn(m)/dm} \\ &\rightarrow (1+\delta)\, \left(1 - \delta_0(\delta) \frac{d\ln dn(m)/dm}{d\delta_c}\right) \\ &\rightarrow 1 + \delta - \delta\, \frac{d\ln dn(m)/dm}{d\delta_c}, \end{aligned} \tag{C.5}$$

where we have used the fact that $\delta_0 \approx \delta \ll 1$. Hence, we can approximate $\delta_h(m) \approx b(m)\delta$ where $b(m) \approx 1 - d\ln n(m,\delta_c)/d\delta_c$ (equation 2.2) is no longer a function of V. $B(M)$ is related to $n(M,V)$ in a similar fashion.

Thus, the second integral in the expression above becomes

$$\begin{aligned}
&\int_m^\infty dM\, n(M,V)\, B(M,V)\, N(m|M,V) \\
&= \int_m^\infty dM\, n(M,V) \left[1 - \frac{d\ln n(M,V)}{d\delta_0}\right] N(m|M,V) \\
&= \frac{dn(m)}{dm} - \int_m^\infty dM\, \frac{dn(M,V)}{d\delta_0}\, N(m|M,V) \\
&= \frac{dn(m)}{dm} - \frac{d}{d\delta_0}\int_m^\infty dM\, n(M,V)\, N(m|M,V) \\
&\qquad + \int_m^\infty dM\, n(M,V)\, \frac{dN(m|M,V)}{d\delta_0} \\
&= \frac{dn(m)}{dm} - \frac{dn(m)}{d\delta_0} - \frac{d}{d\delta_c}\int_m^\infty dM\, n(M,V)\, N(m|M,V) \\
&= \frac{dn(m)}{dm}\left[1 - \frac{d\ln dn(m)/dm}{d\delta_c}\right] \\
&= \frac{dn(m)}{dm}\, b(m). \qquad \text{(C.6)}
\end{aligned}$$

Hence

$$\frac{P_{2h-2p}(k|\delta)}{P_{Lin}(k|R_p)} \to \left[\int_0^\infty dm\, \frac{dn(m)}{dm}\, \frac{m}{\bar{\rho}}\, b(m)\, u(k|m)\right]^2. \qquad \text{(C.7)}$$

If we do not truncate the two-halo term on small scales, as a crude approximation for P_{2h-1p}, i.e., if we simply set $P_{Lin}(k|R_p) \to P_{Lin}(k)$, then this agrees with Equation (1.9).

If we do truncate, then we must check if the large-scale limit of $P_{2h-1p}+P_{2h-2p}$ agrees with Equation (1.9). In the limit of large patches and all environments, the two-halo one-patch term becomes

$$\begin{aligned}
&\frac{P_{2h-1p}(k|\delta)}{P_{\rm Lin}(k) - P_{\rm Lin}(k|R_p)} \to \int_0^\infty dM\, \frac{n(M,V)(1+\delta)}{M/\bar{\rho}} \\
&\times \left[\int_0^M dm\, \frac{dn(m)}{dm}\, \frac{m}{\bar{\rho}}\, b(m)\, u(k|m)\, V[1 + b(m)\delta]\right]^2.
\end{aligned}$$

On large scales $\delta \ll 1$ so we can ignore $b\delta$ compared to unity. Since $M/\bar{\rho} = V(1+\delta)$,

$$\begin{aligned}
\frac{P_{2h-1p}(k|\delta)}{P_{\rm Lin}(k) - P_{\rm Lin}(k|R_p)} &\approx \int_0^\infty dM\, Vn(M,V) \\
&\qquad \times \left[\int_0^M dm\, \frac{dn(m)}{dm}\, \frac{m}{\bar{\rho}}\, b(m)\, u(k|m)\right]^2 \\
&\approx \int_0^\infty dM\, Vn(M,V) \\
&\qquad \times \left[\int_0^\infty dm\, \frac{dn(m)}{dm}\, \frac{m}{\bar{\rho}}\, b(m)\, u(k|m)\right]^2 \\
&= \left[\int_0^\infty dm\, \frac{dn(m)}{dm}\, \frac{m}{\bar{\rho}}\, b(m)\, u(k|m)\right]^2 \qquad \text{(C.8)}
\end{aligned}$$

where the second line follows from the fact that if M is much larger than the mass of the largest halo, then the upper limit of the integrals over m can safely be changed from M to ∞, and the final line uses the fact that the integral over M which remains is the same as the integral over the counts-in-cells probability distribution, and so equals unity. This shows explicitly that, in this limit, $P_{2h-1p} + P_{2h-2p}$ agrees with Equation (1.9).

BIBLIOGRAPHY

Abbas U., Sheth R.K., 2005, MNRAS, 364, 1327

Abbas U., Sheth R.K., 2006, MNRAS in press (astro-ph/0601407)

Adelman-McCarthy J.K. et al., 2005, astro-ph/0507711

Alam S.M.K., Bullock J.S., Weinberg D.H., 2002, ApJ, 572, 34

Alcock C., et al., 1997, ApJ, 486, 697

Armendariz-Picon C., Mukhanoz V., Steinhardt P.J., 2000, PRL, 85, 4438

Avila-Reese V., et al., 2001, ApJ, 559, 516

Avila-Reese V., Colin P., Gottloeber S., Firmani C., Maulbetsch C., 2005, ApJ, 634, 51

Bahcall J.N., Flynn C., Gould A., Kirhakos S., 1994, ApJL, 435, 51

Bahcall N. A., et al., 1999, Science, 284, 1481

Balogh M.L., Morris S.L., 2000, MNRAS, 318, 703

Balogh M.L. et al., 2002, MNRAS, 337, 256

Balogh M.L. et al., 2004, MNRAS, 348, 1355

Baugh C.M., 1996, MNRAS, 280, 267

Bekenstein J.D., 2005, Phys. Rev. D, 71, 069901

Benson A.J. et al., 2001, MNRAS, 327, 1041

Berlind A.A., Blanton M.R., Hogg D.W., Weinberg D.H., Dave R., Eisenstein D.J., Katz N., 2005, ApJ, 629, 625

Berlind A.A., Weinberg D.H., 2002, ApJ, 575, 587

Blanton M.R., Eisenstein D.J., Hogg D.W., Zehavi I., 2004, astro-ph/0411037

Bond J.R., Kofman L.A., Pogosyan D., 1996, Nature, 380, 603

Bottino A., et al., 1994, Astropart. Phys. 2, 67

Burles S., Tytler D., 1998, ApJ, 507, 732

Butcher H., Oemler A., 1984, ApJ, 285, 426

Caldwell R.R., Phys. Lett. B, 2002, 545, 23

Caldwell R.R., Dave R., Steinhardt P.J., 1998, Phys. Rev. Letters, 80, 1582

Chabrier G., Segretain L., M'era D., 1996, ApJ, 468, 21

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

Colless M., et al., 2001, MNRAS, 328, 1039

Cooray A., PhD thesis, astro-ph/0105440

Cooray A., Sheth R.K., 2002, Phys. Rep., 372, 1

Croton D.J., et al., 2005, MNRAS, 356, 1155

Cucciati O. et al., 2006, A&A submitted (astro-ph/0603202)

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

Davis M., Efstathiou G., Frenk C.S., White S.D.M., 1985, ApJ, 292, 371

Davis M., et al., 2003, Proc. SPIE, 4834, 161

de Lapparent V., Geller M.J., Huchra J.P., 1986, ApJ, 302, L1

Dressler A., 1980, ApJ, 236, 351

Edery A., 1999, Phys. Rev. Lett., 1999, 83, 3990

Einstein A., 1917, Sitz. Preuss. Akad. Wiss., 142

Farouki R., Shapiro S.L., 1980, ApJ, 241, 928

Friedmann A., 1922, Z. Physik 10, 377

Gao L., Springel V., White S. D. M., 2005, MNRAS, 363, L66

Geller M.J., Huchra J.P., 1989, Science, 246, 897

Gomez P.L. et al., 2003, ApJ, 584, 210

Gottloeber S., Klypin A., Kravtsov A.V., 2001, ApJ, 546, 223

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

Gunn J.E., Gott J.R.I., 1972, ApJ, 528, 118

Guth A.H., 1981, Phys. Rev. D, 23, 347

Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039

Hogg D.W. et al., 2004, ApJ, 601, L29

Holtmann E., Kawasaki M., Moroi T., 1996, Phys. Rev. Lett., 77, 3712

Hubble E., 1929, PNAS, 15, 168

Jenkins A.R., et al. (for the Virgo consortium), 1998, ApJ, 499, 20

Kaiser N., 1984, ApJ, 284, L9

Kaiser N., 1987, ApJ, 227, 1

Kauffmann G., Nusser A., Steinmetz M., 1997, MNRAS, 286, 795

Kauffmann G., Colberg J.M., Diaferio A., White S.D.M., 1999, MNRAS, 307, 529

Kauffmann G. et al., 2004, MNRAS, 353, 713

Klypin A.A., Primack J., Holtzman J., 1996, ApJ, 466, 13

Kravtsov A. et al., 2004, ApJ, 609, 35

Kurki-Suonio H., 2001, astro-ph/0112182

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

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

Lemson G., Kauffmann G., 1999, MNRAS, 302, 111

Le Fevre O. et al., 2005, A&A, 439, 845

Ma C.-P., Fry J.N., 2000, ApJ, 531, L87

Marinoni C. et al., 2005, A&A, 442, 801

McClelland J., & Silk J., 1978, ApJS, 36, 389

Milgrom M., 1983, ApJ, 270, 365

Minchin R., et al., 2005, ApJL, 622, 21

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

Mo H. J., Yang X., van den Bosch F.C., Jing Y.P., 2004, MNRAS, 349, 205

Monaco P., PhD thesis, astro-ph/9710085

Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nature, 379, 613

Moore B., et al., 1999, ApJ, 524, 19

Navarro J.F., 2003, RSPTA, 361, 2515

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

Neyman J., Scott E.L., 1952, ApJ, 116, 144

Norberg P. et al., 2001, MNRAS, 328, 64

Norberg P. et al., 2002, MNRAS, 332, 827

Peacock J., 1997, MNRAS, 284, 885

Peacock J., Dodds S.J., 1994, MNRAS, 267, 1020

Peacock J., Smith R.E., 2000, MNRAS, 318, 1144

Peebles P.J.E., 1974, ApJ, 189, L51

Peebles P.J.E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ

Perlmutter S. et al., 1999, AJ, 1999, 517, 565

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

Riess A.G. et al., 1998, AJ, 116, 1009

Rubin V., Ford Jr.W.K., 1970, ApJ, 159, 379

Sadoulet B., 1999, Rev. Mod. Phys., 71, S197

Scargle J.D., 1981, ApJS, 45, 1

Scherrer R.J., Bertschinger E., 1991, ApJ, 381, 349

Scoccimarro R., Sheth R.K., Hui L., Jain B., 2000, ApJ, 546, 20

Scoccimarro R., 2004, PhRvD, 70, 3007

Scoville N.Z., et al., 2006, ApJ submitted

Seljak U., 2000, MNRAS, 318, 203

Seljak U., 2001, MNRAS, 325, 1359

Sheth J.V., PhD thesis, astro-ph/0602433

Sheth R.K., 1996, MNRAS, 279, 1310

Sheth R.K., 1998, MNRAS, 300, 105

Sheth R.K., 2005, MNRAS, 364, 796

Sheth R.K., Diaferio A., 2001, MNRAS, 322, 901

Sheth R.K., Jain B., 1997, MNRAS, 285, 231

Sheth R.K., Jain B., 2003, MNRAS, 345, 62

Sheth R.K., Lemson G., 1999, MNRAS, 304, 767

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

Sheth R.K., Tormen G., 2002, MNRAS, 329, 61

Sheth R.K., Tormen G., 2004, MNRAS, 350, 1385

Sheth R.K., Abbas U., Skibba R.A., in Diaferio A., ed, 2004, Proc. IAU Coll. 195, Outskirts of galaxy clusters: intense life in the suburbs, CUP, Cambridge, p. 349

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

Sheth R.K., Hui L., Diaferio A., Scoccimarro R., 2001, MNRAS, 325, 1288

Skibba R., Sheth R.K., Connolly A.J., Scranton R., 2006, MNRAS submitted (astro-ph/0512463)

Smith R.E., Watts P.I.R., Sheth R.K., 2006, MNRAS, 365, 214

Springel V. et al., 2005, Nature, 435, 629

Somerville R., Primack J.R.S., 1999, MNRAS, 310, 1087

Straumann N., 1999, EJP, 20, 419

Sumner T., 2002, Living rev. Rel., http://www.livingreviews.org/lrr-2002-4

Tegmark M., et al., 2004, ApJ, 606, 702

Tinker J.L., Weinberg D.H., Zheng Z., Zehavi I., 2005, ApJ, 631, 41

Tinker J.L., 2006, MNRAS submitted (astro-ph/0604217)

Turner M.S., 2000, Phys. Rep., 333-334, 619

Wechsler R.H., Zentner A.R., Bullock J.S., Kravtsov A.V., 2005, ApJ submitted (astro-ph/0512416)

Weinberg S., 1972, Gravitation and Cosmology, Wiley & Sons, New York

White M., 2001, MNRAS, 321, 1

White S.D.M., Frenk C.S., 1991, ApJ, 379, 52

White S.D.M., Rees M.J., 1978, MNRAS, 183, 341

Yahagi H., Nagashima M., Yoshii Y., 2004, ApJ, 605, 709

Yang X., Mo H.J., van den Bosch F.C., 2003, MNRAS, 339, 1057

Yoo J., Tinker J. L., Weinberg D. H., Zheng Z., Katz N., Dave R., 2005, ApJ submitted (astro-ph/0511580)

York D.G. et al., 2000, AJ, 120, 1579

Yoshida N., Sheth R.K., Diaferio A., 2001, MNRAS, 328, 669

Zehavi I. et al., 2005, ApJ, 630, 1

Zheng Z., Tinker J.L., Weinberg D.H., Berlind A.A., 2002, ApJ, 575, 617

Zhytnikov V.V., Nester J.M., 1994, Phys. Rev. Lett., 73, 2950

Zwicky F., 1937, ApJ, 86, 217