Dissertation

submitted to the

Combined Faculties of the Natural Sciences and Mathematics

of the Ruperto-Carola-University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Marcello Cacciato

born in: Enna (Italy)

Oral examination: July 27th, 2009

Galaxy-Dark Matter Connection:
from Astrophysics to Cosmology

Referees: Prof. Dr. Matthias Bartelmann
Prof. Dr. Frank C. van den Bosch

Zusammenfassung

Die Beobachtung der Gravitationslinseneffekte von Galaxien auf andere Galaxien, das sogenannte “Galaxien-Galaxien-Linsensignal” (G-G-Signal), ist ein ideales Werkzeug, um die Verteilung der Dunklen Materie auf den Skalen von Galaxien zu untersuchen. Die nötige Fehlergenauigkeit bei der Messung dieses Signals kann nur dadurch erreicht werden, dass Bilder von unzähligen Vordergrundsgalaxien (“Linsen”) aufaddiert werden und dann die durchschnittliche Elliptizität der Hintergrundsgalaxien (“Objekte”) ermittelt wird. Allerdings verhindert die Mittelung über die Vordergrundsobjekte jede *direkte* astrophysikalische Interpretation, weswegen ein verlässliches Modell für die Verteilung von Galaxien in Dunkle-Materie-Halos verfügbar sein muss, um kosmologische Schlüsse aus dem mittleren Linsensignal zu ziehen. Hier stellen wir ein realistisches Modell für die Statistik von Halo-Besetzungszuständen vor, die auf der bedingten Leuchtkraftfunktion aufbaut. Diese gibt eine statistische Beschreibung der Anzahl von Galaxien mit einer bestimmten Leuchtkraft, die in Halos von einer bestimmten Masse zu finden ist. Da sie *a priori* durch die Leuchtkraftabhängigkeit der groräumigen Galaxienverteilung eingegrenzt ist, erlaubt es die bedingte Leuchtkraftfunktion, das G-G-Signal ohne zusätzliche Feintarierung vorherzusagen. Unser Modell ermöglicht ein intuitives Verständnis der verschiedenen Terme, die zu diesem Signal beitragen. Die Auswirkungen der verschiedenen Annahmen, die in das Modell einflieen, werden detailliert untersucht. Unsere theoretischen Vorhersagen stimmen gut mit Beobachtungen aus der SDSS-Durchmusterung überein. Da die Vorhersagen des Modells sich sehr stark mit unterschiedlichen Annahmen über die zugrundeliegende Kosmologie verändern, stellen wir mit der gemeinsamen Analyse von Galaxienverteilung und G-G-Signal eine neuartige Methode fr die Bestimmung kosmologischer Parameter wie $\Omega_{\rm m}$ und σ_8 vor. Wir zeigen, dass Unsicherheiten und systematische Fehler im Modell seine Vorhersagen nicht signifikant beeinflussen. Daher ist unsere neuartige Methode ein geeignetes Mittel, um kosmologische Parameter komplementär zu bisherigen Verfahren zu bestimmen.

Abstract

Galaxy-galaxy (g-g) lensing represents an ideal technique to constrain the dark matter distribution on galaxy scales. The required accuracy in the signal can be achieved only by stacking many foreground galaxies and averaging the ellipticity of the resulting background galaxies. Unfortunately, the stacking procedure complicates any astrophysical interpretation. In order to extract information from the composite g-g lensing signal, a reliable model of the way galaxies populate dark matter haloes is required. We use a realistic description of the halo occupation statistics based on the conditional luminosity function. It provides a statistical prescription for the number of galaxies with a given luminosity living in dark matter haloes of a given mass. Being "a priori" constrained by the luminosity dependence of the galaxy clustering, it can be used for predicting the g-g lensing signal without any additional tuning. Our model allows a thorough understanding of the different terms contributing to the signal. We carefully explore the effect of the assumptions entering the model. Our theoretical predictions are in very good agreement with SDSS data. Furthermore, we use the sensitivity of this technique to the underlying cosmological model to study the feasibility of a joint analysis of galaxy clustering and g-g lensing as a novel technique to constrain the values of cosmological parameters such as $\Omega_{\rm m}$ and σ_8. We show that uncertainties and systematics in the model do not significantly affect the model predictions. We conclude that a combined analysis of galaxy clustering and g-g lensing can be used as a novel, complementary and competitive technique to constrain cosmological parameters.

Contents

CONTENTS

List of Figures

LIST OF FIGURES

List of Tables

LIST OF TABLES

1

Overview

Mankind has always been wondering about the origin and the history of the universe. However, only in the last century, it has become possible to address in an empirical manner specific questions about the nature of extragalactic phenomena. Particularly in the last decades, the increasing technical capability in observations has led to deep insights in extragalactic astronomy and cosmology. As a result, a concordance model has emerged within which astrophysical observations can be successfully explained and accurate theoretical predictions can be provided. As we aim towards an ever increasing understanding, every answer to astrophysical problems raises yet more fundamental questions about the physics of the universe. In this perspective, it is extremely fascinating that the study of the luminous matter yields to the discovery and the analysis of the intrinsic properties of the "dark" constituents of the universe.

The history of this connection goes back to 1933 when Zwicky realized that the mass resulting from the sum of the masses of individual galaxies in the Virgo cluster was not sufficient to explain the mass of the cluster inferred from a dynamical study of these galaxies. This *"missing mass"* problem gave birth to the inference of the presence of dark matter in the universe. Since then, our understanding of the physics in galaxy clusters has surely developed (see Abell & Eastmond 1975; Bahcall 1977; Biviano 2000 and reference therein for detailed reviews). However, the precise nature of the dark matter still eludes us. Intensive experimental searches have being carried out for *"direct"* (e.g. ANAIS, Cebrián et al. 2000; CDMS and XENON, Baudis 2006; DAMA/LIBRA, Bernabei et al. 2008; EDELWEISS, Sanglard et al. 2005; ZEPLIN-III, Lebedenko et al. 2008; MACHO, Bennett et al. 1991; OGLE, Udalski et al. 1994) and *"indirect"* (e.g. PAMELA, Bonvicini et al. 2001) detection of dark matter. Although such a detection has not yet happened, there are favored particle candidates, namely weakly-interacting massive particles (WIMP) that are believed to have survived since the very early phases of the universe. However, still major progresses in understanding the extreme physics of that cosmological era are required to infer what particles might have existed then, and how many would have survived. Thus, the search for dark matter particles has made astrophysics the test bed of the new horizons of particle physics.

1. OVERVIEW

In this thesis, we address the study of dark matter in the universe in a pure astrophysical perspective. In particular, we study the properties of dark matter distribution rather than the nature of such particles. Moreover, since we are interested in the properties of dark matter distribution on galaxy scales and beyond, we define the scope of this thesis as *galaxy-dark matter connection*.

For different types of galaxies, different methods have been developed in order to constrain the properties of this connection. For example, it has been noted that the rotation velocity of many spiral galaxies stays roughly constant out to large radii rather than declining, as would be expected if the visible stars provided all the gravitating mass. This phenomenon was observed through both optical and extended neutral hydrogen (HI) rotation curves (e.g. Bosma 1978; Faber & Gallagher 1979; Rubin, Ford and Thonnard 1980; Rubin et al. 1982). HI observations have the advantage that they can be extended up to large radii, so the rotation curves can probe the mass distribution far from the center of a galaxy. A flat rotation curve at large radii is interpreted as a strong evidence of the presence of a dark matter halo around spiral galaxies. For elliptical galaxies, dynamic tracers as planetary nebulae and globular clusters have been successfully used to probe the existence of dark matter haloes around these kind of galaxies. Recently, multiple images due to strong gravitational effects have been also used to probe the properties of the dark matter distribution around ellipticals (see e.g. Treu & Koopmans 2002 and reference therein). Furthermore, for the case of galaxy clusters two methods have been extremely powerful in determining the properties of the dark matter distribution: the analysis of the X-ray emission of the hot intra-cluster gas and the study of (both weak and strong) gravitational lensing effects. While the first method relies on the assumption of hydrostatic equilibrium of the gas in the potential well of the dark matter halo, gravitational lensing is not based on any assumption about the physical state of the luminous matter. However, typical galaxy cluster mass reconstructions based on lensing have the shortcoming of probing the distribution of mass inside a cylinder rather than in a spherical region around the galaxy cluster.

With strong support from all these observational probes, a theoretical framework has been developed for galaxy formation in a cold dark matter scenario. It is generally accepted that every galaxy resides in a a dark matter halo. Moreover, we refer to central galaxies to those located more or less at rest at the center of the halo's potential well, whereas we refer to satellite galaxies as those orbiting within the halo. Note also that, in jargon, one refers to galaxy groups as to sets of a handful of galaxies sharing the same dark matter halo. Instead, if the satellite galaxy population is numerous (i.e. hundred to thousands of galaxies) one refers to galaxy clusters. The line-of-sight (hereafter los) velocity dispersion of satellite galaxies reflects the depth of the halo's potential well, and is thus a measure for the halo's mass. In the case of rich galaxy clusters, the number of satellite galaxies can be sufficient to properly sample the los velocity distribution of its dark matter halo (Carlberg et al. 1997) and this method can be used, in addition to the X-ray and lensing studies, to estimate the halo mass. In the case of galaxy groups, the number of (detectable) satellite galaxies is generally too small to obtain a reliable measure of the los velocity dispersion. However, under the assumption that central galaxies with similar properties (i.e. luminosity) are hosted by haloes of similar masses, one can stack central galaxies and combine the kinematic information of their associated satellites to improve the

statistics. Pioneering efforts in this direction were made by Erickson et al. 1987, Zaritsky et al. 1993, and Zaritsky et al. 1997. With the advent of large galaxy redshift surveys such as the two degree field galaxy redshift survey (hereafter 2dFGRS; see Colless et al. 2001) and the Sloan Digital Sky Survey (hereafter SDSS; see York et al. 2000), it has become possible to apply this method to much larger samples of satellite galaxies. Consequently, sophisticated techniques have been introduced to properly account for all uncertainties related with this methodology (see e.g. McKay et al. 2002; Prada et al. 2003; Brainerd & Specian 2003; van den Bosch 2004; Conroy 2007; Becker 2007; Norberg et al. 2008; More et al. 2009a,b).

The advent of galaxy redshift surveys has also allowed one to obtain accurate measurements of the spatial distribution of galaxies. In particular, it has been possible to determine the clustering of galaxies as a function of their properties, such as luminosity, morphology and color (e.g. Guzzo et al.2000; Norberg et al.2001, 2002; Zehavi et al.2005; Wang et al.2007). Since galaxies are believed to form and reside in dark matter haloes, the clustering strength of a given population of galaxies can be compared to that of dark matter haloes as predicted by numerical simulations or the extended Press-Schechter formalism. Such a comparison reveals a wealth of information about the way galaxies populate dark matter haloes (e.g. Jing, Mo & Börner 1998; Peacock & Smith 2000; Berlind & Weinberg 2002; Yang, Mo & van den Bosch 2003; van den Bosch, Yang & Mo 2003; van den Bosch et al.2007). Unfortunately, this method of constraining the link between galaxies and dark matter haloes using galaxy clustering has one important shortcoming: the halo occupation statistics inferred from the observed clustering properties depend on the adopted cosmological model. More precisely, models based on different cosmologies can fit the clustering data almost equally well by simply relying on different halo occupation statistics or, equivalently, different mass-to-light ratios. In order to break this degeneracy independent constraints on the mass-to-light ratios are required (e.g. van den Bosch, Mo & Yang 2003; Tinker et al.2005). One method that can provide these constraints is galaxy-galaxy lensing (hereafter g-g lensing), which probes the mass distributions (and hence the halo masses) around galaxies.

Galaxy-galaxy lensing measures the tangential shear distortions in the shapes of background galaxies (hereafter sources) induced by the mass distribution around foreground galaxies (hereafter lenses). The first attempt to detect g-g lensing was made by Tyson et al.(1984), but because of the relatively poor quality of their data they were unable to detect a statistically significant signal. The first clear detection was obtained only twelve years later (Brainerd et al. 1996). With wider and deeper surveys becoming available, however, g-g lensing has been detected with very high significance, and as function of various properties of the lensing galaxies (e.g. Griffiths et al.1996; Hudson et al.1998; McKay et al.2001; Guzik & Seljak 2002; Hoekstra et al.2003, 2004; Sheldon et al.2004, 2007a,b; Mandelbaum et al.2006; Heymans et al.2006; Johnston et al.2007; Parker et al.2007; Mandelbaum, Seljak & Hirata 2008). Unfortunately, a proper interpretation of these data in terms of the link between galaxies and dark matter haloes has been hampered by the fact that the lensing signal can typically only be detected when stacking the signal of many lenses. Since not all lenses reside in haloes of the same mass, the resulting signal is a non-trivial average of the lensing signal due to haloes of different

masses. Most studies to date have assumed that the relation between the luminosity of a lens galaxy and the mass of its halo is given by a simple power-law relation with zero scatter (see Limousin et al.2007 for a detailed overview). However, it has become clear, recently, that the scatter in this relation between light and mass can be very substantial (More et al.2009b, and references therein). As shown by Tasitsiomi et al.(2004), this scatter has a very significant impact on the actual lensing signal, and thus has to be accounted for in the analysis. In addition, central galaxies (those residing at the center of a dark matter halo) and satellite galaxies (those orbiting around a central galaxy) contribute very different lensing signals, even when they reside in haloes of the same mass (e.g. Natarajan, Kneib & Smail 2002; Yang et al.2006; Limousin et al.2007). This has to be properly accounted for (see e.g. Guzik & Seljak 2002), and requires knowledge of both the satellite fractions and of the spatial number density distribution of satellite galaxies within their dark matter haloes.

Over the years, numerous techniques have been developed to interpret galaxy-galaxy lensing measurements (Schneider & Rix 1997; Natarajan & Kneib 1997; Brainerd & Wright 2002; Guzik & Seljak 2001). Several authors have also used numerical simulations to investigate the link between g-g lensing and the galaxy-dark matter connnection (e.g., Tasitsiomi et al.2004; Limousin et al.2005; Natarajan, De Lucia & Springel 2007; Hayashi & White 2007). It has become clear from these studies that g-g lensing in principle contains a wealth of information regarding the mass distributions around galaxies; in addition to simply probing halo masses, g-g lensing also holds the potential to measure the shapes, concentrations and radii of dark matter haloes, and the first observational results along these lines have already been obtained (Natarajan et al.2002; Hoekstra et al.2004; Mandelbaum et al.2006; Limousin et al.2007; Mandelbaum et al.2008).

In this thesis we use an analytical model, similar to that developed by Seljak (2000) and Guzik & Seljak (2001), to predict the g-g lensing signal as a function of the luminosity of the lenses starting from a model for the halo occupation statistics that is constrained to fit the abundances and clustering properties of the lens galaxies. A comparison of these predictions with the data thus allows us to test the mass-to-light ratios inferred from the halo occupation model. The occupation statistics are described via the conditional luminosity function (CLF; see Yang et al.2003), which specifies the average number of galaxies of given luminosity that reside in a halo of given mass. This CLF is ideally suited to model g-g lensing, as it allows one to properly account for the scatter in the relation between luminosity and halo mass, and to split the galaxy population in centrals and satellites. We demonstrate how these different galaxy populations contribute to the lensing signal in different luminosity bins, and show the impact on our results of the uncertainties related to the expected concentrations of dark matter haloes and the radial number density distributions of satellite galaxies.

Assuming a flat ΛCDM cosmology with parameters supported by the third year data release of the Wilkinson Microwave Anisotropy Probe (WMAP, see Spergel et al.2007), we obtain a CLF that accurately fits the abundances and clustering properties of SDSS galaxies. Using our analytical model, we show that this same CLF also accurately matches the g-g lensing data obtained from the SDSS by Seljak et al.(2005) and Mandelbaum et al.(2006) *without any additional tuning of the model parameters*. However, if we repeat the exercise of constraining the CLF parameters using the abundance

and clustering of galaxies for a cosmology with slightly higher ($\sim 20\%$) matter density and power-spectrum normalization, the model that fits the clustering data can no longer reproduce the g-g lensing data. This means that one cannot arbitrarily modify the halo occupation statistics and simultaneously reproduce all the statistical properties of galaxies for any cosmology. More specifically, this result clearly shows that a joint analysis of clustering and g-g lensing is extremely sensitive to the underlying cosmological model. Thus, we study to which extent such a joint analysis may be used as a possible novel and competitive method to constrain cosmological parameters.

In this thesis, we only consider inflationary CDM cosmologies with the initial conditions for structure formation set by density perturbations which follow a power spectrum assumed to have a power law form. In addition, we assume that the vacuum energy is described by a cosmological constant. Furthermore, the geometry of the universe is assumed flat as suggested by the location of the first peak in the angular power spectrum of cosmic microwave background (CMB) temperature fluctuations (e.g., Balbi et al.2000; Lange et al.2001; Pryke et al.2002; Netterfield et al.2002; Ruhl et al.2002). Such a cosmological model can be characterized by 5 parameters: the energy densities (in terms of the critical density) of baryons, $\Omega_{\rm b}$ and of cold dark matter, $\Omega_{\rm dm}$, the Hubble parameter, $h = H_0/(100\ {\rm km\,s^{-1}\,Mpc^{-1}})$, the spectral index, n, and normalization, σ_8, of the initial power spectrum. We define $\Omega_{\rm m} \equiv \Omega_{\rm b} + \Omega_{\rm dm}$ as the matter density and we thus have that the $\Omega_\Lambda = 1 - \Omega_{\rm m}$ defines the cosmological constant energy density.

Recent years have seen a tremendous improvement in the constraints on the cosmological parameters listed above. Clearly, a concordance cosmological model has emerged with well constrained values for $(\Omega_{\rm m}, \Omega_\Lambda, \Omega_{\rm b}, h, n)$. High redshift supernovae (SN) Ia indicate that $\Omega_\Lambda \sim 0.7$ (e.g., Riess et al.1998; Perlmutter et al.1999, Astier et al. 2006). The contribution from the baryons is determined either from primordial nucleosynthesis models and measurements of the primeval deuterium abundance for which $\Omega_{\rm b}h^2 = 0.020 \pm 0.001$ (Burles, Nollett & Turner 2001) or via CMB anisotropy measurements (Hanany et al.2000; de Bernardis et al.2002; Pryke et al.2002; Netterfield et al.2002, Dunkley et al. 2009) which suggest $\Omega_{\rm b}h^2 = 0.022 \pm 0.004$. The HST Key project has constrained the Hubble constant to $h = 0.72 \pm 0.08$ (Freedman et al.2001). Using this as a prior, the CMB data yield values for $\Omega_{\rm m}$ within the range $[0.2, 0.4]$ (depending on the assumptions in the analysis, see e.g., Lewis & Bridle 2002; Dunkley et al. 2009) and the analysis of SN Ia gives $\Omega_{\rm m} = 0.26 \pm 0.03$ (Astier et al. 2006) in agreement with combined analyses of SN Ia, CMB and baryonic acoustic oscillation (BAO) in galaxy correlation function data which yield $\Omega_{\rm m} = 0.249 \pm 0.018$ (Percival et al.2007, Komatsu et al. 2009) and $\Omega_{\rm m} = 0.27 \pm 0.02$ (Eisenstein et al. 2005). Additional (and equally consistent) constraints on the amount of matter in the universe were found by using the galaxy power spectrum (Padmanabhan et al. 2007, Blake et al. 2007, Spergel et al. 2007). The spectral index of the initial power spectrum has been constrained by CMB anisotropy measurements to $n = 0.96^{+0.014}_{-0.015}$ (see Dunkley et al.2009 and reference therein), in excellent agreement with the inflation paradigm which predicts values of n close, but not necessarily equal, to unity.

It must be noticed that in such a *concordance cosmology* the value of the normalization parameter, σ_8, has not yet been determined with high accuracy. In the literature, one can find different attempts to

determine σ_8. Perhaps, the two most widely applied methods are the observed abundance of clusters of galaxies (Bahcall et al.2003; Pierpaoli et al.2002; Reiprich & Böhringer 2002; Seljak 2002; Viana, Nichol & Liddle 2002) and the cosmic shear measured from weak lensing (Hamana 2002; Hoekstra, Yee & Gladders 2004; van Waerbeke et al. 2002; Bacon & Taylor 2003). Unfortunately, both methods are actually dependent on a combination of both σ_8 and Ω_m. Moreover, the cluster abundance method requires accurate mass estimates of individual clusters, whereas the weak lensing method relies on difficult cosmic shear measurements (e.g. Jarvis et al.2003 and references therein). Current estimates for σ_8 range from ~ 0.6 (e.g., Borgani et al.2001; Reiprich & Böhringer 2002; Seljak 2002; Viana, Nichol & Liddle 2002; Spergel et al. 2007) to ~ 1.0 (e.g., Bacon et al.2002; Fan & Bahcall 1998; Pen 1998; Pierpaoli, Scott & White 2001; van Waerbeke et al.2002, Spergel et al. 2003). Note that this range is much wider than the typical statistical uncertainty derived by the individual methods. In particular, note that the WMAP team itself has suggested $\sigma_8 = 0.84 \pm 0.04$ (after the first year data release), $\sigma_8 = 0.76 \pm 0.04$ (after the third year data release), and $\sigma_8 = 0.796 \pm 0.036$ (as result of the fifth year data analysis).

As shown in chapter 6, the method proposed in this thesis allows us to determine the value of both Ω_m and σ_8 with high accuracy. Since galaxy clustering and galaxy-galaxy lensing probe the galaxy-dar matter connection in complementary ways, the degeneracy between these two parameters is almost entirely lifted. The method reveals itself as a new tool to constrain cosmological parameters. Note that it is complementary to all the other cosmological probes because it is based on the purely non-linear regime of the structures in the local universe. In fact, it uses the information encoded on galaxy scales through galaxy clustering and g-g lensing. In addition, the analytical character of the CLF allows us to study the systematics involved in the method and to verify that they do not strongly affect the results.

This thesis is organized as follows. We present the properties of the Sloan Digital Sky Survey (SDSS) in chapter 2. Furthermore, in the same chapter, we introduce the data used throughout the thesis: the galaxy luminosity function, the projected galaxy correlation function, the galaxy correlation lengths as a function of luminosity and the excess surface density of galaxies selected according to their absolute magnitude. Note that all these data have been obtained with SDSS. In chapter 3, we review the basic theory of structure formation in the universe. In chapter 4, we introduce the analytical model used throughout the thesis to interpret the data. In particular, note that in this chapter we introduce the CLF used to describe the connection between galaxies and dark matter haloes. In chapter 5, we constrain the CLF parameters establishing then the galaxy-dark matter connection in a quantitative way. Finally, In chapter 6, we study the sensitivity of such an analysis to the underlying cosmology and we introduce a novel method to constrain cosmological parameters based on a joint analysis of galaxy clustering and g-g lensing. We summarize the findings of this thesis in chapter 7, in which we also provide indications for possible future applications.

Some of the results presented in this thesis have been submitted to a referred journal and accepted for publication[1]:

- **Cacciato M.**, van den Bosch F.C., More S., Li R., Mo H.J., Yang X., 2009, MNRAS, 394, 929
- Li R., Mo H.J., Fan Z., **Cacciato M.**, van den Bosch F.C., Yang X., More S., 2009, MNRAS, 394, 1016

[1]The analysis presented in chapter 6 is in preparation for submission to a refereed journal

2

Observational Data

In this chapter, we introduce the data that we will use throughout the thesis. Note that the entire thesis is based on data obtained with the Sloan Digital Sky Survey (SDSS). Therefore, in the first section we describe the general properties of this survey, whereas in the remaining sections we describe the galaxy luminosity function, the two-point galaxy correlation function for galaxies of different luminosity, the luminosity dependence of the correlation length, and the galaxy-galaxy lensing signal. One great advantage of using SDSS is that the same galaxies can be used for abundance, clustering, and lensing analysis. This ensures homogeneity in the selection effects of the catalogue.

2.1 Observing galaxies with SDSS

The Sloan Digital Sky Survey is an ongoing project to map approximately a quarter of the sky using a dedicated instrument located at the Apache Point Observatory in southern New Mexico. The telescope provides a nearly 3°, virtually undistorted field of view. The SDSS imaging camera (Gunn et al. 1998) is a mosaic of 30 2048 × 2048 pixel CCDs used for the primary imaging. The imaging CCDs are arranged in six columns of 5 CCDs each. Each of the 5 CCDs in a column views the sky through a different broadband filter. The five filters (u, g, r, i, z, see Fukugita et al. 1996 for details) span a range from the atmospheric cutoff at 300nm to the limit of CCD sensitivity at 1100nm.

SDSS imaging data are obtained in drift scan mode. In general, the telescope is driven along a great circle on the sky in such a way that objects pass directly down a column of CCDs. This allows essentially simultaneous observations to be obtained in each of the five passbands and provides very efficient survey observing (the shutter never closes). Because the CCD columns are separated by nearly a CCD width, a single SDSS observation of a strip of sky contains large gaps. The gaps in a single "strip" are then filled in on a subsequent night to obtain a completely filled "stripe". Moreover, because of the Galactic extinction, only certain region of the sky are covered (see fig. 2.1). Successful drift scanning with a wide field system requires an optical design with very low distortion, which has the added benefit of removing an important shape systematic facilitating gravitational lensing studies.

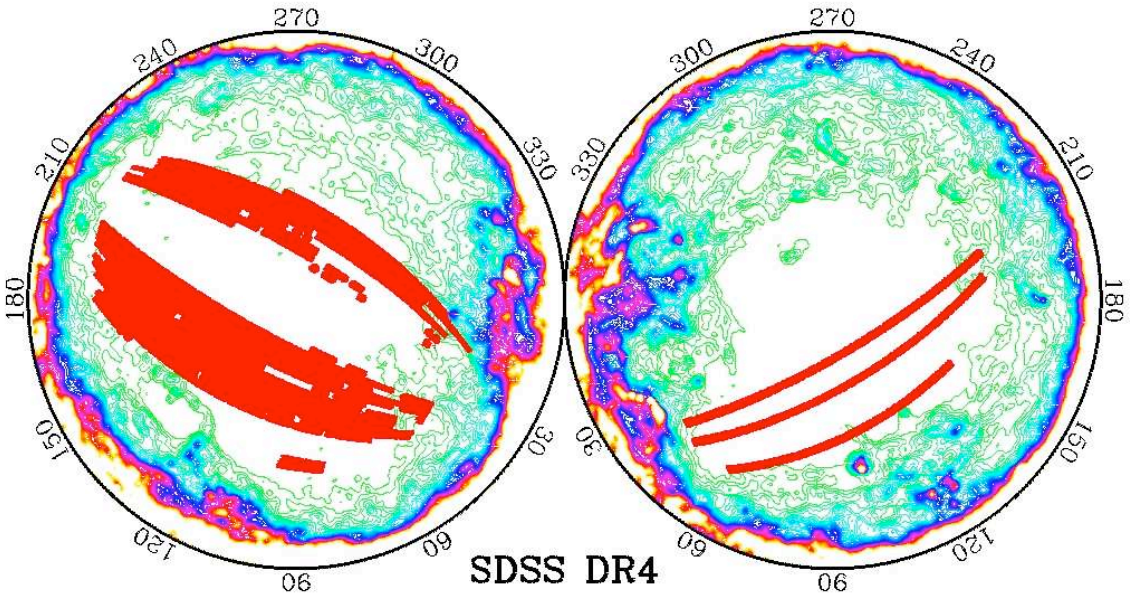

Figure 2.1: The sky coverage of the SDSS DR4 galaxies, overlaid on the Galactic extinction contours of Schlegel, Finkbeiner & Davis (1998). Note that the SDSS probes regions of low galactic extinction.

Detailed descriptions of the data acquisition, reduction and calibration can be found in Uomoto et al. (2000) and in Pier et al. (2000).

SDSS has obtained deep, multi-color images covering a quarter of the sky and it has created 3-dimensional maps containing more than 930,000 galaxies and more than 120,000 quasars. SDSS data have been provided to the scientific community in various releases. In this thesis, we use the version based on SDSS Data Release 4 (Adelman-McCarthy et al. 2006). Fig. 2.2 shows a slice through the SDSS 3-dimensional map of the distribution of galaxies. Earth is at the center, and each point represents a galaxy. Galaxies are colored according to the ages of their stars, with the redder showing galaxies that are made of older stars. The outer circle is at a distance of $z \sim 0.15$. The region between the wedges was not mapped by the SDSS because of Galactic extinction.

2.2 Galaxy Luminosity Function

The galaxy luminosity function specifies the way in which galaxies are distributed with respect to their luminosity. More precisely, one can define $\Phi(^{0.1}M_r)\,\mathrm{d}^{0.1}M_r$ as the number density of galaxies with absolute magnitude in the interval $[^{0.1}M_r, {}^{0.1}M_r + \mathrm{d}^{0.1}M_r]$. Here $^{0.1}M_r$ indicates the r-band magnitude K+E corrected to $z = 0.1$ following the procedure of Blanton et al.(2003b). Throughout this thesis, we use the luminosity function (hereafter LF) of Blanton et al (2003a) uniformly sampled

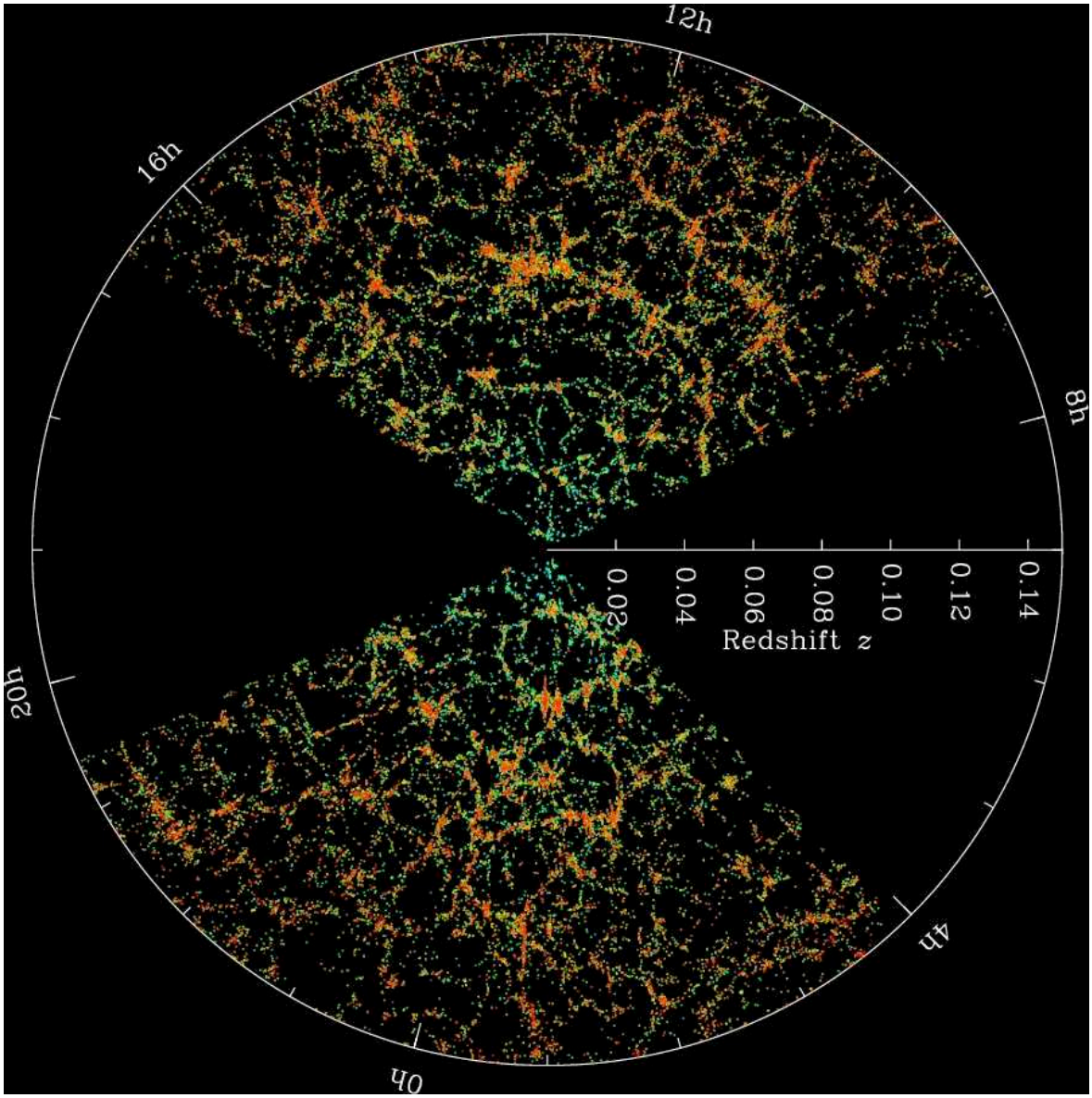

Figure 2.2: Slices through the SDSS 3-dimensional map of the distribution of galaxies. Earth is at the center, and each point represents a galaxy. Galaxies are colored according to the ages of their stars, with the redder points showing galaxies that are made of older stars. The outer circle is at a distance of $z \sim 0.15$. The region between the wedges was not mapped by the SDSS because of Galactic extinction. Both slices contain all galaxies within -1.25 and 1.25 degrees declination. [Ref. www.sdss.org]

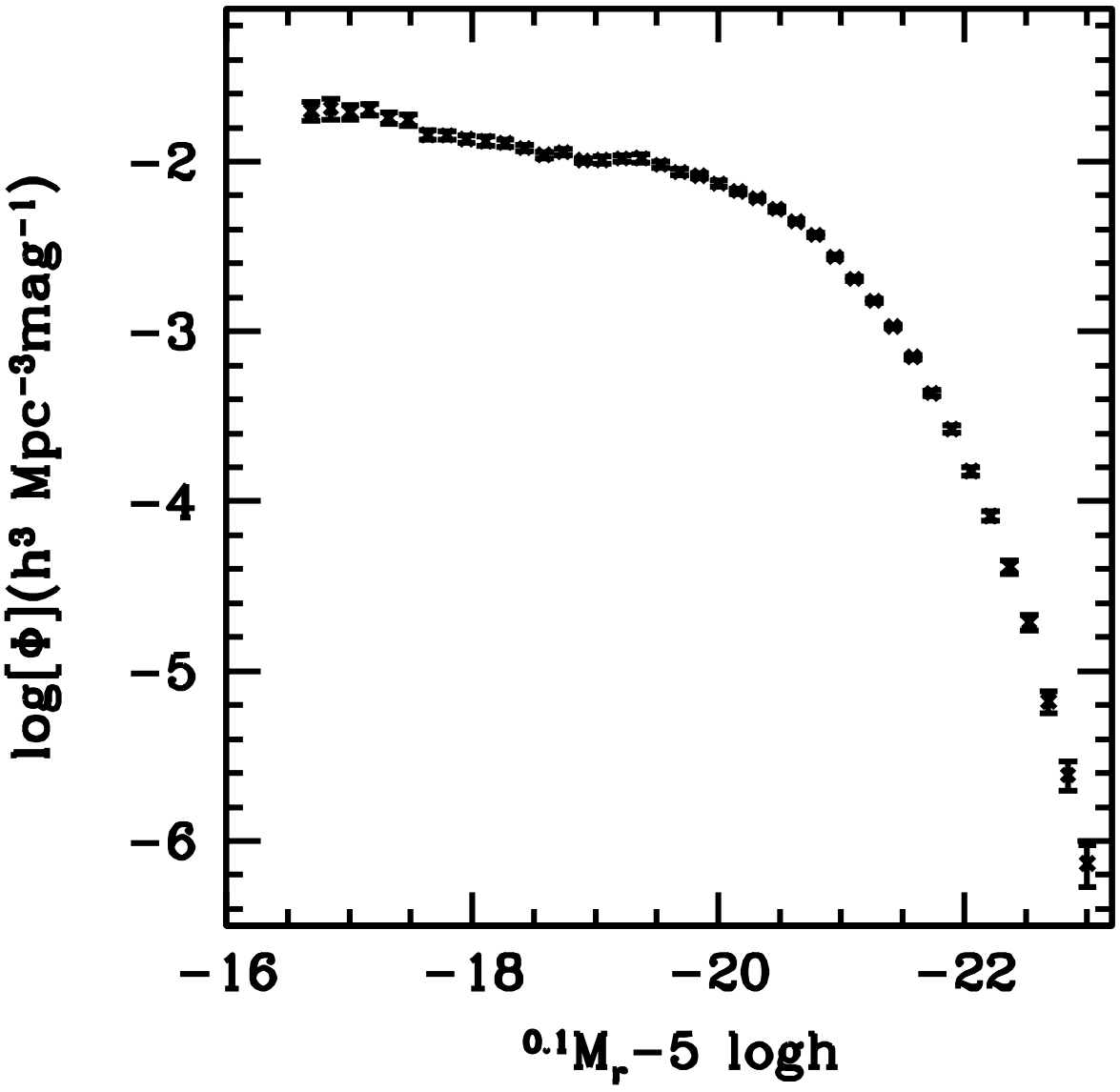

Figure 2.3: The galaxy luminosity function, Φ, derived from SDSS (see Blanton et al. 2003a).

at 41 magnitudes covering the range $-23.0 \leq {}^{0.1}M_r - 5\log h \leq -16.4$ as shown in fig. 2.3. Here, we briefly describe the procedure to obtain the galaxy luminosity function, $\Phi({}^{0.1}M_r)$ from SDSS data. Each galaxy has a measured magnitude m_r and an associated uncertainty Δm_r. The magnitude ${}^{0.1}M_r$ may be constructed from the apparent magnitude m_r and the redshift z:

$$ {}^{0.1}M_r - 5\log h = m_r - \mathrm{DM}(z, \Omega_\mathrm{m}, \Omega_\Lambda, h) - K(z)\,. \tag{2.1} $$

Note that one needs to calculate the distance modulus, $\mathrm{DM}(z, \Omega_\mathrm{m}, \Omega_\Lambda, h)$, for a given distance and cosmology (for example using the method suggested by Hogg 1999) and the K-correction at the any given z. In order to fit the distribution of the absolute magnitudes and redshifts of galaxies one uses a maximum likelihood method which allows for a generic luminosity function shape (see Blanton et al. 2003b for detail). If expressed as a function of the luminosity, L, instead of the absolute magnitude, the measured galaxy luminosity function turns out to be well described by the functional form

$$ \Phi(L) = \left(\frac{\Phi^*}{L^*}\right)\left(\frac{L}{L^*}\right)^{\alpha} \exp(-L/L^*)\,, \tag{2.2} $$

where L^* is a characteristic luminosity above which the function drops exponentially, α is the faint end slope, and Φ^* is a normalization. The use of this functional form to describe galaxy number density was first introduced by Schechter (1976) and this is the reason why in the astronomical community this function is known as *Schechter function*.

2.3 Galaxy Correlation Function

One of the principal scientific objectives of SDSS is to map the large scale structure traced by optical galaxies with unprecedented precision and over a wide range of scales. The large sample of galaxies observed by SDSS allows one to study the clustering of galaxies with different properties, such as luminosity, color, spectral type, morphology, and stellar mass (see e. g. Guzzo et al.2000; Norberg et al.2001, 2002; Zehavi et al.2005; Wang et al.2007). All these studies show that galaxies of different kinds have different clustering properties. This indicates that galaxies are biased tracers of the dark matter distribution. Therefore, one has to be careful when using the observed galaxy distribution to infer the mass distribution in the universe to ultimately provide tests of galaxy formation theory.

In this thesis, we use galaxy clustering data obtained via a sample of 284489 galaxies from SDSS. These galaxies are distributed over a range $0.01 < z < 0.3$ and have absolute magnitude in the range $-23 < {}^{0.1}M_r < -18$. We quantify the clustering properties using the 2-point correlation function (see chapter 3 and 4 for a theoretical definition of this quantity). Different method can be used to estimate the 2-point galaxy correlation function. We adopt the definition introduced by Davis & Peebles (1983),

$$ \xi_{\mathrm{gg}}(R, r_\pi) = \frac{DD(R, r_\pi)}{DR(R, r_\pi)} - 1 \tag{2.3} $$

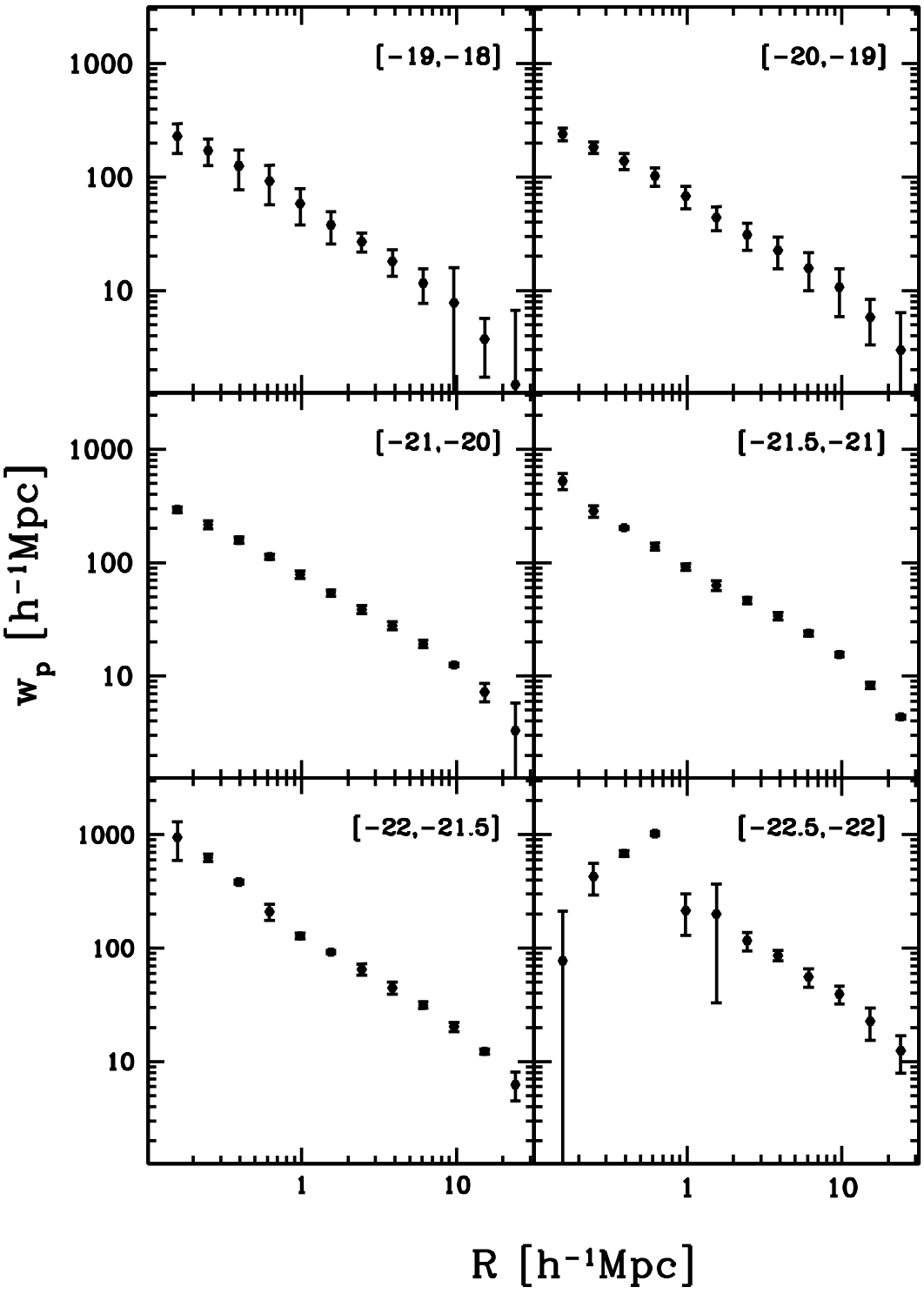

Figure 2.4: The projected correlation function, w_p, calculated from SDSS (see Wang et al. 2007).

where DD is the count of galaxy pairs at distance (R, r_π), whereas DR is the count of galaxy-random pairs. Note that R and r_π are the separation of pairs perpendicular and parallel to the line of sight. For this thesis, the *projected 2-point correlation function*, $w_p(R)$, has been calculated as follows

$$w_p(R) = 2\int_0^\infty \xi_{\rm gg}(R, r_\pi)\mathrm{d}r_\pi = 2\sum_k \xi_{\rm gg}(R, r_{\pi,k})\Delta r_{\pi,k}\,, \tag{2.4}$$

where the summation has been done for $1 < k < 40$ corresponding to $r_\pi = 0.5h^{-1}$ Mpc to $r_\pi = 39.5h^{-1}$ Mpc. Furthermore, the projected 2-point correlation function has been evaluated for galaxies in different luminosity bins.

Since SDSS is a flux limited survey, it is affected by the *Malmquist bias*, i.e. brighter galaxies are observed up to higher redshift than fainter galaxies. When splitting the galaxy sample into luminosity bins, this translates into having sub-samples of galaxies with different mean redshift. Table 2.2 displays the different luminosity bins together with mean redshift for every luminosity bin. Fig. 2.4 shows the 2-point projected galaxy correlation function obtained by Wang et al. (2007). Note that the overall normalization of the signal increases with increasing luminosity. It is, in fact, well established that brighter galaxies are more strongly clustered. Note also that, although a power law may describe the data reasonably well, one can argue that small deviations from a simple power law are observed. As we will see in chapter 4, these deviations are expected as a feature of the transition between the small and the large scale clustering. Finally, we would like to comment on the data of the brightest bin. We interpret the clustering at small scales as a lack of bright galaxies in the sample used from Wang et al. (2007). In fact, since extremely bright galaxies are expected to reside at the center of extremely large dark matter haloes, it is expected that on scales smaller than twice the typical radius of these haloes the correlation is suppressed simply due to lack of galaxies.

2.3.1 The correlation length

On relatively large scales ($1 \lesssim r \lesssim 10h^{-1}$Mpc), the 2-point galaxy correlation function is described reasonably well by a power law

$$\xi_{\rm gg}(r) = \left(\frac{r}{r_0}\right)^\alpha \tag{2.5}$$

where r_0 and α are respectively the correlation length and the slope of the correlation function. We stress that r_0 is defined such that $\xi(r_0) = 1$. The value of r_0 can thus be used to quantify the strength of the correlation function. In this thesis, we use the correlation lengths of galaxies selected according to luminosity. In particular, we use the results obtained by Wang et al.(2007) for six volume limited samples selected from the SDSS DR4. For completeness, these data are listed in Table 2.1. In Fig.(2.5), we show these correlation lengths as a function of the r-band magnitudes. Note that the error bars along the x-axis represent the width of the magnitude bins.

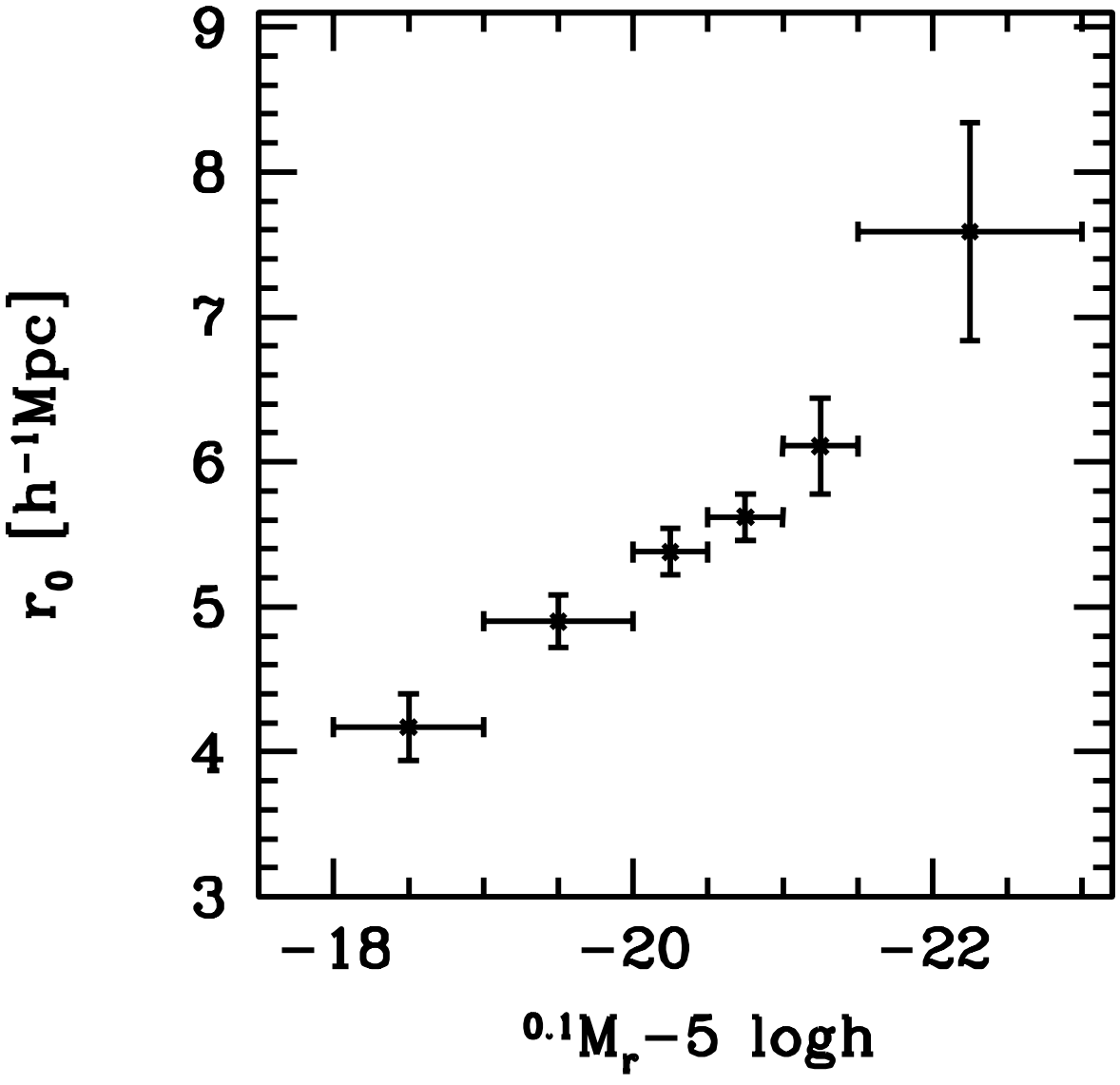

Figure 2.5: The luminosity dependence of the correlation length, r_0, calculated from SDSS (see Wang et al. 2007).

Table 2.1: Correlation lengths.

Sample	$^{0.1}M_r - 5\log h$	$< z >$	r_0
V1	$(-23.0, -21.5]$	0.173	7.59 ± 0.75
V2	$(-21.5, -21.0]$	0.135	6.11 ± 0.33
V3	$(-21.0, -20.5]$	0.109	5.62 ± 0.16
V4	$(-20.5, -20.0]$	0.089	5.38 ± 0.16
V5	$(-20.0, -19.0]$	0.058	4.90 ± 0.18
V6	$(-19.0, -18.0]$	0.038	4.17 ± 0.23

Galaxy-galaxy clustering correlation lengths of Wang et al.(2007). Column (1) lists the ID of each volume limited sample, following the notation of Wang et al. (2007). Columns (2) and (3) indicate the absolute magnitude range and the mean redshift of each sample, while column (4) lists the correlation length plus its standard deviation (in h^{-1} Mpc), obtained by fitting a power-law to the projected correlation function over the radial range $[0.98, 9.6]h^{-1}$ Mpc. See Wang et al.(2007) for details.

2.4 Galaxy-Galaxy Lensing

Galaxy-galaxy lensing measures the tangential shear distortions, $\gamma_{\rm t}$, in the shapes of background galaxies (hereafter sources) induced by the mass distribution around foreground galaxies (hereafter lenses). Since the tangential shear distortions due to a typical lens galaxy (and its associated dark matter halo) are extremely small, and since background sources have non-zero intrinsic ellipticities, measuring $\gamma_{\rm t}$ with sufficient signal-to-noise requires large numbers of background galaxies. In general, however, the number density of detectable background sources is insufficient for a reliable measurement of $\gamma_{\rm t}$ around individual lenses. This problem is circumvented by stacking many lenses according to some observable property. For example, Mandelbaum et al. (2006) measured $\gamma_{\rm t}$ as a function of the transverse comoving distance R by stacking thousands of lenses in a given luminosity bin $[L_1, L_2]$.

The tangential shear as a function of the projected radius R around the lenses is related to the excess surface density (ESD) profile, $\Delta\Sigma(R)$, according to

$$\Delta\Sigma(R) = \gamma_{\rm t}(R)\Sigma_{\rm crit}\,. \tag{2.6}$$

where the so-called critical surface density, $\Sigma_{\rm crit}$, is a geometrical parameter given by

$$\Sigma_{\rm crit} = \frac{c^2}{4\pi G}\frac{\omega_{\rm S}}{\omega_{\rm L}\omega_{\rm LS}(1+z_L)}, \tag{2.7}$$

with $\omega_{\rm S}$, $\omega_{\rm L}$ and $\omega_{\rm LS}$ the comoving distances to the source, the lens and between the two, respectively, and with z_L the redshift of the lens.

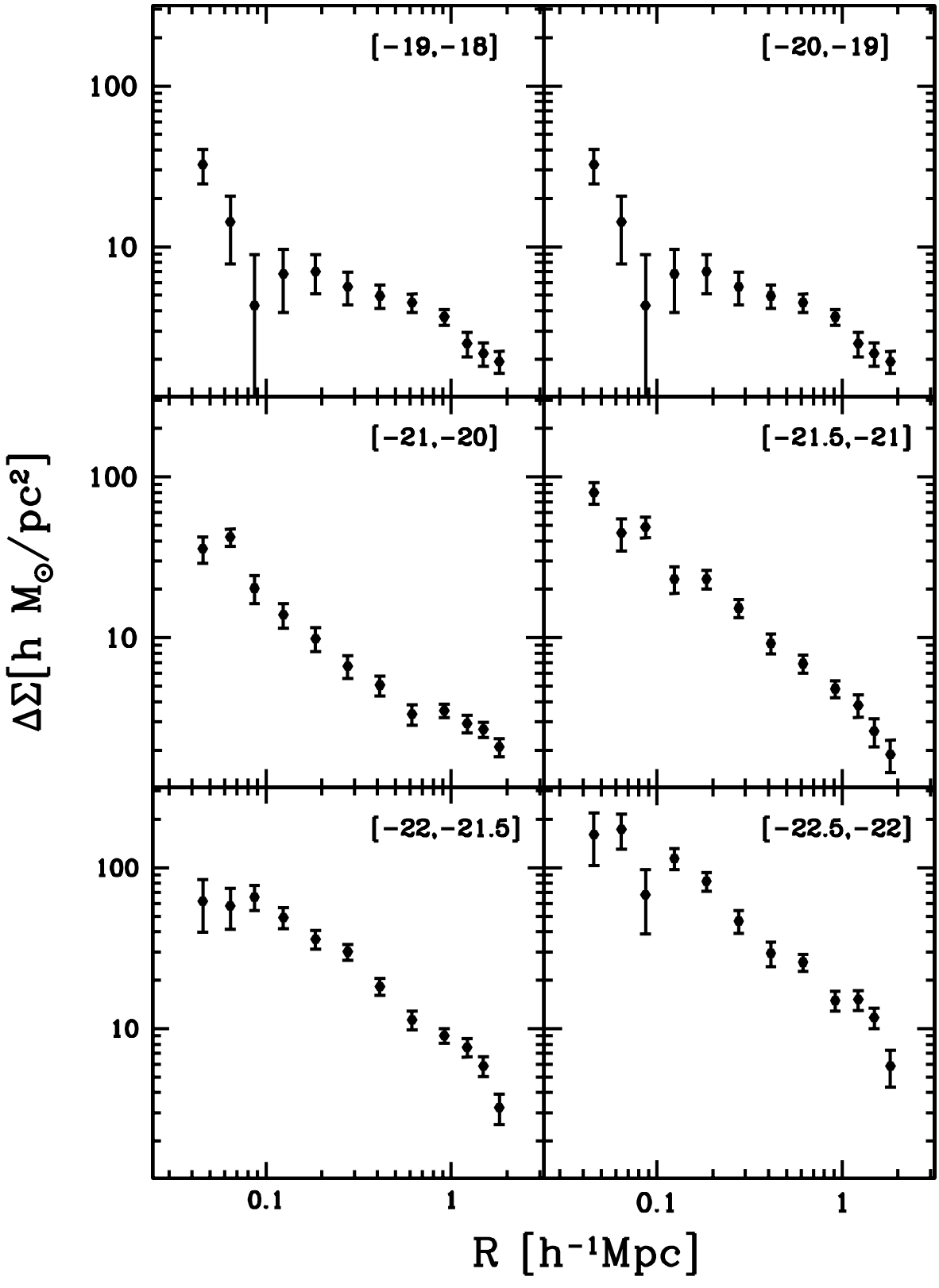

Figure 2.6: The excess surface density, $\Delta\Sigma$, calculated from SDSS (see Mandelbaum et al. 2006)

The g-g lensing data used here is described in Seljak et al.(2005) and Mandelbaum et al.(2006) and has been kindly provided to us by R. Mandelbaum. The catalogue of lenses consists of 351, 507 galaxies with magnitudes $-17 \geq {}^{0.1}M_r - 5\log h \geq -23$ and redshifts $0.02 < z < 0.35$ taken from the main galaxy catalogue of the SDSS Data Release 4 (Adelman-McCarthy et al.2006). This sample is split in luminosity bins (see Table 2.2), and for each of these luminosity bins the excess surface density profiles, $\Delta\Sigma(R)$, have been determined from measurements of the shapes of more than 30 million galaxies in the SDSS imaging data down to an apparent r-band magnitude of $r = 21.8$. The resulting data cover the range between 0.04 and 2 h^{-1} Mpc. Data are shown as solid dots with errorbars in Fig. 2.6. We refer the reader to Mandelbaum et al.(2005b, 2006) for a detailed description of the data and of the methods used to determine the ESD profiles. In what follows, we comment only on the properties of the g-g lensing data of interest for this thesis. First, notice that the shear map is obtained by measuring the ellipticity of background galaxies as a function of the angular separation in the sky. This map can be converted in $\gamma_t(R)$ and then in $\Delta\Sigma(R)$ only if the average redshift of the stacked foreground galaxies (lenses) is known. This is one of the reasons why SDSS represents a milestone in g-g lensing measurements. In fact, as it will be clear from the description of the model used to analyze the signal, a proper interpretation of g-g lensing data would be extremely difficult without the knowledge of the redshift of the lenses. Second, since not all the galaxies with the same luminosity reside in haloes of the same mass, the signal resulting from the stacking procedure is a non-trivial average of the lensing signal due to haloes of different masses. From an observational point of view, an attempt to limit this effect consists in using luminosity bins which are as narrow as possible. However, it must be noticed that even if one could use infinitesimally narrow bins, a scatter in the luminosity-halo mass relation is expected. Thus, in order to properly interpret the g-g lensing signal one has to account for this scatter as, for example, done in the model presented in chapter 4.

Table 2.2: Luminosity bins and mean redshifts

ID	$^{0.1}M_r - 5\log h$	$\langle z \rangle$
(1)	(2)	(3)
L1	$(-18.0, -17.0]$	0.032
L2	$(-19.0, -18.0]$	0.047
L3	$(-20.0, -19.0]$	0.071
L4	$(-21.0, -20.0]$	0.10
L5f	$(-21.5, -21.0]$	0.14
L5b	$(-22.0, -21.5]$	0.17
L6f	$(-22.5, -22.0]$	0.20

Luminosity bins of the lenses. Column (1) lists the ID of each luminosity bin, following the notation of Mandelbaum et al.(2006). Column (2) indicates the magnitude range of each luminosity bin (all magnitudes are K+E corrected to $z = 0.1$). Note that clustering data refer to the luminosity bins $L2 - L6$f. Column (3) indicates the mean redshift of the galaxies in each luminosity bin. See Mandelbaum et al.(2006) and Wang et al. (2007) for details.

3

Theoretical Framework

In this chapter, we present the paradigm within which we interpret theoretical predictions and observational data throughout the thesis. In particular, in section § 3.1, we introduce two analytical tools which will be extensively used throughout the thesis: the correlation function and the power spectrum. We then describe the main assumptions of the hierarchical structure formation scenario (§ 3.2) and we define the linear and non linear regime of the growth of perturbations in the density field. After a brief description of the principles of the formation of dark matter haloes, we present the theoretical predictions for their internal structure (§ 3.3), their spatial abundance (§ 3.4), and their clustering properties (§ 3.5).

3.1 Statistical tools: Correlation Function & Power Spectrum

In the current paradigm, we believe that structures in the universe form by the growth of tiny inhomogeneities in the dark matter density field. The origin of these inhomogeneities is usually related to inflationary models of the early stages of the universe. In this thesis, we are not concerned about the physical origin of the fluctuations but we simply assume that their evolution can be described by gravitational instability. Let us define the quantity $\delta(\mathbf{r})$ as fluctuations relative to the mean background density, $\bar{\rho}$:

$$\delta(\mathbf{r}) = \frac{\rho(\mathbf{r}) - \bar{\rho}}{\bar{\rho}}\,, \tag{3.1}$$

where $\rho(\mathbf{r})$ defines the density distribution of dark matter in the universe. For a statistical analysis of the fluctuations in the density field, it is convenient to describe them as a superposition of many modes. Throughout the thesis we assume that the universe has a flat geometry. Since, in a flat space, plane waves are a complete set of functions, we decompose the density fluctuations in Fourier modes:

$$\delta(\mathbf{r}) = \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp[-i\mathbf{k}\mathbf{r}]\,, \tag{3.2}$$

where $\mathbf{k}$ is the wave-vector of components (k_x, k_y, k_z). This decomposition is exact if one assumes that the field is periodic over a box of side L. If we let the box become arbitrarily large, the sum will go over to an integral that incorporates the density of states in $k-$space. We thus have

$$\delta(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \delta_{\mathbf{k}} \, \exp[-i\mathbf{kr}] \, \mathrm{d}^3\mathbf{k} \, . \tag{3.3}$$

Since the cosmic density field is believed to be a random field generated by some random processes, one is interested in its statistical properties. One of these properties is the clustering which we study through the 2-point correlation function defined as

$$\xi(\mathbf{r}) \equiv \langle \delta(\mathbf{r}_1)\delta(\mathbf{r}_2) \rangle = \langle \delta(\mathbf{r}_1)\delta(\mathbf{r}_1 + \mathbf{r}) \rangle \, , \tag{3.4}$$

where $\mathbf{r}_2 = \mathbf{r}_1 + \mathbf{r}$. Note that $\langle \, \rangle$ denotes an ensemble average, meaning that one should average over an ensemble of universes. Obviously, that is not possible and one has to rely on the assumption of egodicity, i.e. the ensemble averages can be written in terms of a spatial averages. Throughout the thesis, ensemble averages will be evaluated via spatial averages.

Since the real space density fluctuations can be written as a sum over Fourier modes, (see eq. [3.2]), we can write

$$\begin{aligned}
\xi(\mathbf{r}) &= \langle \sum_{\mathbf{k}} \delta_{\mathbf{k}} \exp[-i\mathbf{kr}_1] \sum_{\mathbf{k'}} \delta_{\mathbf{k}'} \exp[-i\mathbf{k}'(\mathbf{r}_1 + \mathbf{r})] \rangle = \\
&= \langle \sum_{\mathbf{k}} \sum_{\mathbf{k'}} \delta_{\mathbf{k}} \, \delta_{-\mathbf{k}'} \exp[-i(\mathbf{k}' - \mathbf{k})\mathbf{r}_1] \exp[-i\mathbf{k}(\mathbf{r})] \rangle = \\
&= \langle \sum_{\mathbf{k}} \sum_{\mathbf{k'}} \delta_{\mathbf{k}} \, \delta^*_{\mathbf{k}'} \exp[-i(\mathbf{k}' - \mathbf{k})\mathbf{r}_1] \exp[-i\mathbf{kr}] \rangle = \\
&= \langle \sum_{\mathbf{k}} |\delta_{\mathbf{k}}|^2 \exp[-i\mathbf{kr}] \rangle = \\
&= \frac{1}{(2\pi)^3} \int \langle |\delta_{\mathbf{k}}|^2 \rangle \exp[-i\mathbf{kr}] \, \mathrm{d}^3\mathbf{k} = \\
&= \frac{1}{(2\pi)^3} \int P(\mathbf{k}) \exp[-i\mathbf{kr}] \, \mathrm{d}^3\mathbf{k} \, ,
\end{aligned} \tag{3.5}$$

where we have used the fact that $\delta_{-\mathbf{k}} = \delta^*_{\mathbf{k}}$ and that cross terms with $\mathbf{k}' \neq \mathbf{k}$ average to zero. The last equality defines the power spectrum $P(\mathbf{k}) = \langle |\delta_{\mathbf{k}}|^2 \rangle$. In an isotropic universe, the density fluctuation correlation (and therefore the power spectrum) cannot have a preferred direction, and we therefore have an isotropic correlation function, $\xi(r)$, and an isotropic power spectrum, $P(k)$ where $r = |\mathbf{r_1} - \mathbf{r_2}|$ and $k = |\mathbf{k_1} - \mathbf{k_2}|$. This allows us to write

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) \exp[-i\mathbf{kr}] \, \mathrm{d}^3\mathbf{k} \, . \tag{3.6}$$

Introducing spherical coordinates with the polar axis along $\mathbf{k}$ and using the fact that $\xi(r)$ is a real quantity, we have

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) \, \frac{\sin(kr)}{kr} \, 4\pi k^2 \, \mathrm{d}k \, . \tag{3.7}$$

It is convenient to define the quantity $\Delta(k)$, as

$$\Delta(k) = \frac{k^3 P(k)}{2\pi^2}\,. \tag{3.8}$$

Eq. (3.7) can then be written as

$$\begin{aligned}\xi(r) &= \frac{1}{(2\pi)^3}\int P(k)\,\frac{\sin(kr)}{k^2 r}\,4\pi k^3\,\mathrm{d}k \\ &= \int \Delta(k)\frac{\sin(x)}{x^2}\mathrm{d}x \qquad \text{with } x = kr\,.\end{aligned} \tag{3.9}$$

Since $\Delta(k)$ is a dimensionless quantity, it is often considered the natural Fourier-space counterpart to the dimensionless correlation $\xi(r)$.

In what follows, we want to link the above definition of correlation function with an intuitive picture of pair counting. Following the definition of correlation function, we have

$$\begin{aligned}\xi(r) &= \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle = \\ &= \left\langle\left(\frac{\rho(\mathbf{x})-\bar\rho}{\bar\rho}\right)\left(\frac{\rho(\mathbf{x}+\mathbf{r})-\bar\rho}{\bar\rho}\right)\right\rangle = \\ &= \frac{1}{\bar\rho^2}\left(\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle - \langle\rho(\mathbf{x})\bar\rho\rangle - \langle\bar\rho\rho(\mathbf{x}+\mathbf{r})\rangle + \bar\rho^2\right) \\ &= \frac{1}{\bar\rho^2}\left(\langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle - \bar\rho^2 - \bar\rho^2 + \bar\rho^2\right)\end{aligned} \tag{3.10}$$

which gives

$$\bar\rho^2\left[1+\xi(r)\right] = \langle\rho(\mathbf{x})\rho(\mathbf{x}+\mathbf{r})\rangle\,. \tag{3.11}$$

Let us now construct a distribution of objects which follows $\rho(r)$. This means that, at each volume element δV, one places an object with probability

$$\delta P = \rho(r)\delta V\,. \tag{3.12}$$

The joint probability that objects are assigned to the volume elements δV_1 and δV_2 placed around r_1 and r_2, or, in other words, the probability to have a pair with one object in δV_1 around r_1 and the other in δV_2 around r_2 can be written as

$$\delta P = \rho(r_1)\delta V_1\,\rho(r_2)\delta V_2\,. \tag{3.13}$$

Averaging this expression over the ensemble, and considering $r = |r_1 - r_2|$, one gets

$$\begin{aligned}\delta P &= \langle\,\rho(r_1)\rho(r_2)\,\rangle\delta V_1\,\delta V_2 = \\ &= \bar\rho^2[1+\xi(r)]\,\delta V_1\,\delta V_2\,.\end{aligned} \tag{3.14}$$

Remembering that in the case of a random distribution the same probability is simply

$$\delta P = \bar{\rho}^2 \, \delta V_1 \, \delta V_2 \,, \tag{3.15}$$

it becomes evident that the correlation function quantifies deviations from random distributions. In particular, $\xi(r) > 0$ means excess over a random distribution, whereas $\xi(r) < 0$ means deficiency over a random distribution.

The above discussion clearly shows that the correlation function can be evaluated by counting the number of pairs as a function of their separation and comparing this with that given by a random sample (see chapter 2 for an application to SDSS data).

3.2 Structure Formation

Fluctuations in the density field are assumed to grow over time due to their self gravity. In particular, overdense regions increase their density contrast. In what follows, we refer to to the regime in which the density contrast is well below unity as linear regime, whereas we refer to the case $\delta \gtrsim 1$ as non-linear regime. The reason behind this name resides in the fact that the former can be described by using linear perturbation theory while this is not possible for the latter. In fact, to describe structure formation in the non-linear regime one needs to rely on numerical simulations. In the following sections, we describe the quantitative results related to the growth of perturbations from the linear regime to the formation of dark matter haloes. Finally, we describe dark matter halo statistical properties such as density profile, abundances and clustering. Note that basic concepts of cosmology (e.g. horizon and matter-radiation equality) are not defined in the scope of this chapter because these concepts are assumed to be known.

3.2.1 The linear regime

The physical properties of the early universe determine the initial power spectrum of the fluctuations in the density field. It is common to use a single power law form

$$P_{\rm init}(k) \propto k^n \,, \tag{3.16}$$

where n is known as *the initial spectral index*.

The choice behind this functional form is not completely arbitrary. First of all, a power law does not introduce any characteristic scales. This is reasonable in absence of a physical justification to introduce such scales. In the literature, the power-law form with $n = 1$ is known as Harrison-Zel'dovich-Peebles spectrum (see e.g. Peebles 1982).

Since the density field evolves with time, the corresponding power spectrum will also evolve. More precisely, since in the matter dominated era, the density contrast grows proportionally to the growth factor, $G(t)$,

$$\delta(\mathbf{x}, t) = G(t)\delta(\mathbf{x}) \,, \tag{3.17}$$

the power spectrum is expected to grow with the second power of $G(t)$:

$$P(k,t) = G^2(t)P_{\rm init}(k) = G^2(t)Ak^n\,, \tag{3.18}$$

where A is the amplitude of the initial power spectrum. Note, however, that the above description is based on the growth of perturbations in the matter dominated era. The growth of perturbation in the radiation-dominated era proceeds differently though, also depending on the scale of the perturbations in comparison to the length of the horizon. The form of the power spectrum thus needs to incorporate such a correction term. As common in the literature, we write

$$P(k,t) = A(t)k^nT^2(k)\,, \tag{3.19}$$

where we have written $A \cdot G(t) = A(t)$ to have a more concise expression, and $T(k)$ is the so-called transfer function. One way to fix the amplitude of the power spectrum at a given time, $A(t)$, is by using its variance smoothed over a scale R

$$\sigma_R(t) = 4\pi \int \frac{k^2{\rm d}k}{(2\pi)^3}\, P(k,t)\,|W(kR)|^2\,, \tag{3.20}$$

where $W(kR)$ is the Fourier transform of a window function. Throughout, we will use a spherical top-hat in real space for which $W(kR) = [3/(kR)^3](\sin kR - (kR)\cos kR)$. For historical reasons, it is common to quantify the variance at a scale of $8h^{-1}$ Mpc. In particular, we will refer to σ_8 as the variance in the density field (smoothed over $8h^{-1}$Mpc) and corresponding to a power spectrum linearly evolved to today.

The transfer function $T(k)$ entering in eq.(3.19) depends on the wavenumber k but also on the energy content of the Universe. It is possible to evaluate the precise shape of this function by solving the linearized Boltzmann equations for the distribution function of photons, baryons, and dark matter (see e.g. Seljak & Zaldarriaga 1996). However, in the literature, fitting functions have been provided that can accurately describe the shape of the transfer function. In this thesis, we will use the functional form suggested by Eisenstein & Hu (1998). in what follows, we describe the qualitative behavior of the transfer function in order to understand the shape of the power spectrum.

In jargon, one says that a mode enters the horizon when the comoving length, $\lambda = a\,2\pi/k$, associated to this mode equals the size of the horizon, $r_{\rm H}$:

$$\lambda = r_{\rm H}\,. \tag{3.21}$$

We indicate with L_0 the comoving size of the horizon at the matter radiation equality:

$$L_0 = a_{\rm eq}\frac{2\pi}{k_0} = r_{\rm H}(a_{\rm eq})\,. \tag{3.22}$$

Accordingly, density fluctuations with $\lambda > L_0$ enter the horizon in the matter dominated era. On the other hand, density fluctuations with $\lambda < L_0$ enter the horizon during the radiation dominated era. Since during this epoch perturbations do not grow. All modes with with $\lambda < L_0$ are *frozen* till

the matter will dominate. Their relative growth up to the present time have then been suppressed compared to the modes with $\lambda > L_0$. The corresponding suppression factor is $f_{\rm suppr} = (a_{\rm enter}/a_{\rm eq})^2$. Since the Hubble radius, $r_{\rm H}$, scales as a^2 during the radiation dominated era, we have

$$a_{\rm enter}\frac{2\pi}{k} \propto a_{\rm enter}^2\,, \tag{3.23}$$

which gives

$$a_{\rm enter} \propto k^{-1}\,. \tag{3.24}$$

We thus obtain

$$f_{\rm suppr} = (a_{\rm enter}/a_{\rm eq})^2 = (k_0/k)^2\,. \tag{3.25}$$

The corresponding power spectrum will be suppressed by a factor $f_{\rm suppr}^2 \propto k^{-4}$, which gives

$$P(k) \quad \propto \quad k^{n-4} \qquad \text{if} \quad k > k_0\,. \tag{3.26}$$

In other words, bearing in mind eq. (3.19), we can see that the transfer function $T(k)$ is equal to 1 for $k < k_0$ and it is proportional to k^{-2} for $k > k_0$. In fig.3.1, one can notice the suppression of modes with $k > k_0$. Note that the same figure shows also the dark matter power spectrum in the non linear regime which we describe in the next section.

3.2.2 From the linear to the non-linear regime: the HKLM procedure

The description of the evolution of the density perturbations presented in the previous section is valid only in the so called linear regime. This means that the density contrast is assumed to be much smaller than unity, $\delta(\mathbf{x}) << 1$. As soon as this condition is not verified, one has to find an appropriate treatment for the description of the evolution of the density field. One can develop an analytical perturbation theory which describes the mildly non linear regime ($\delta(\mathbf{x}) \sim 1$) but as the density contrast becomes much bigger than unity no analytical approach can be used to describe the property of the dark matter density field. In general, studying the non-linear structure evolution requires the use of numerical methods. All the quantitative results presented in the next sections are, indeed, derived via numerical simulations. However, in what follows, we describe an attempt to link the results from the linear to the purely non-linear regime in an analytical way.

Hamilton et al. (1991, hereafter HKLM) developed a method for interpolating between linear theory on large scales and the nonlinear predictions on small scales. Here, we briefly describe it. Let the mass enclosed within a spherical overdensity in the initial stages of evolution be $M_0\,(<\ell)$ and its mass at some later time be $M\,(<x)$. According to the spherical collapse model (see next section), this region will reach a maximum expansion point, turn around and collapse. If there is no shell crossing, then mass is conserved and

$$M_0\,(<\ell) = \frac{4}{3}\pi\rho\,(<\ell)\;\;\ell^3 = \frac{4}{3}\pi\rho\,(<x)\;\;x^3 = M\,(<x))\;\;. \tag{3.27}$$

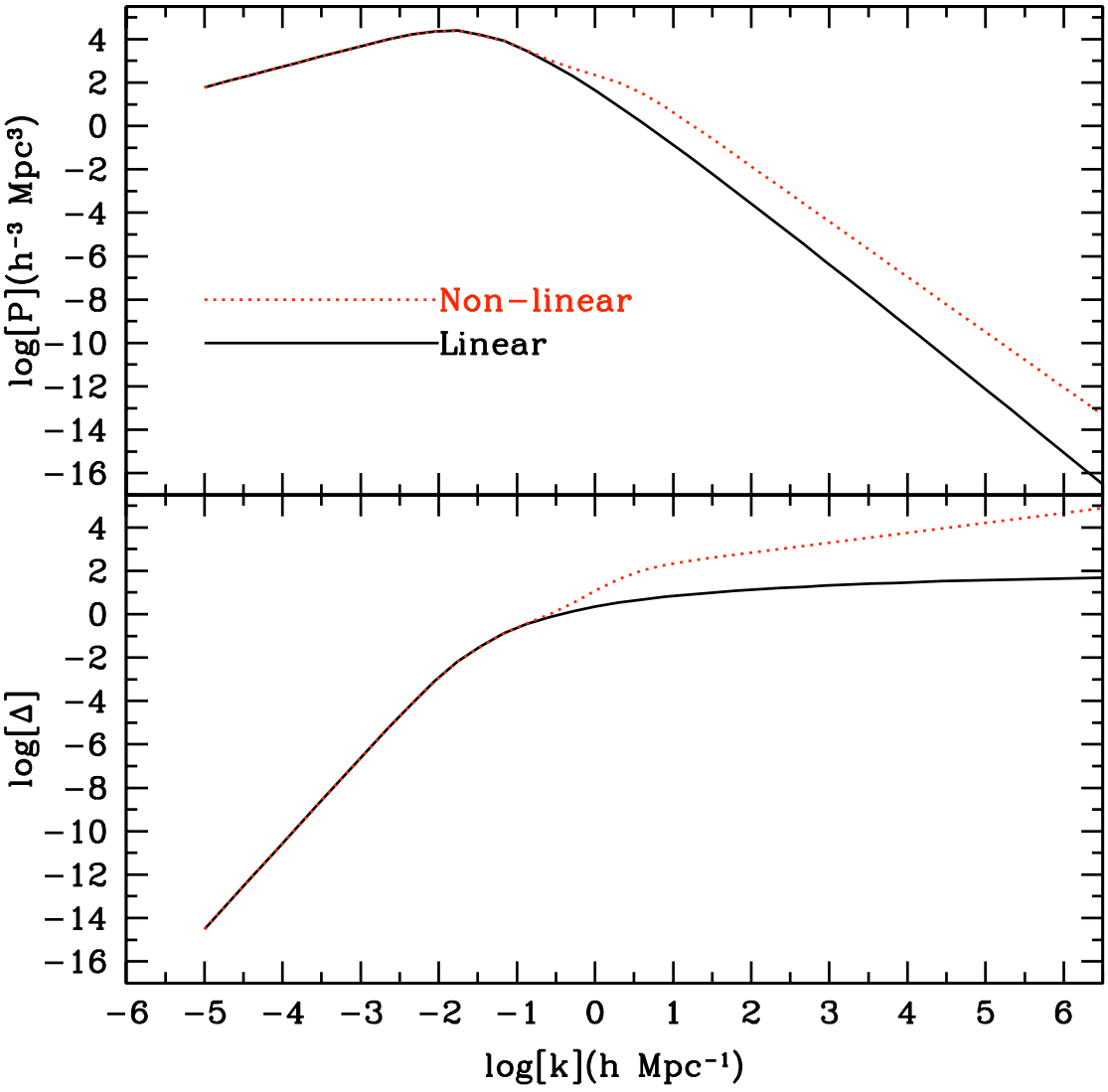

Figure 3.1: The dark matter power spectrum calculated following the fitting function provided by Smith et al (2003). The black solid line represents the linear dark matter power spectrum, whereas the red dotted line represents the non-linear dark matter power spectrum. Note that the lower panel shows the dimensionless power spectrum, Δ.

This implies that

$$x^3 \left[1 + \bar{\xi}_{\rm NL}(x,t)\right] = \ell^3 \ . \tag{3.28}$$

Note that $\bar{\xi}_{\rm NL}(x,t)$ represents the nonlinear volume averaged two-point correlation function:

$$\bar{\xi}_{\rm NL}(x) \equiv \frac{3}{x^3} \int_0^x y^2 \xi_{\rm NL}(y)\, dy \ , \tag{3.29}$$

so $1 + \bar{\xi}_{\rm NL}(x,t)$ is the factor by which the density is enhanced relative to the mean.

After this translation of scales has been performed, the second step is to assume that the nonlinear correlation is a universal function of the linear one:

$$\bar{\xi}_{\rm NL}(x,t) = f_{\rm NL}\left[\bar{\xi}_{\rm L}(\ell,t)\right]. \tag{3.30}$$

The asymptotic behavior of the function $f_{\rm NL}$ is straightforward to calculate. For small arguments, $f_{\rm NL}(x) = x$ because, when fluctuations are not much evolved, the linear correlation is low and the non-linear correlation equals the linear one. Under the assumption of stable clustering, it can be shown that for highly evolved field (corresponding to high values of the correlation) $f_{\rm NL}(x) \propto x^{3/2}$ (with a small dependence on the particular cosmological model). The interpolation between the linear and purely non-linear regime was determined empirically by HKLM, by comparison with numerical simulations.

HKLM's nonlinear scaling argument was further developed by Peacock & Dodds (1994; PD94), who proposed that the scaling ansatz could be used for predicting power spectra by simply replacing $\bar{\xi} \to \Delta$ and letting the linear and nonlinear scales represent linear and nonlinear wavenumbers: $\ell = k_{\rm L}^{-1}$ and $x = k_{\rm NL}^{-1}$. This suggested the formalism

$$\Delta_{\rm NL}(k_{\rm NL}) = f_{\rm NL}\left[\Delta_{\rm L}(k_{\rm L})\right] \ ;$$

$$k_{\rm NL} = \left[1 + \Delta_{\rm NL}(k_{\rm NL})\right]^{1/3} k_{\rm L}. \tag{3.31}$$

The accuracy of the HKLM and PD94 scaling formulae has been intensively tested and further refined over the past years (Jain, Mo & White 1995; Peacock & Dodds 1996; Mo, Jing & Börner 1997; Jain & Bertschinger 1998; Ma & Fry 2000; Scoccimarro et al. 2001, Smith et al. 2003). Fig. 3.1 shows the dark matter power spectrum corresponding to the WMAP3 set of cosmological parameters (see chapter 5) calculated following the fitting function of Smith et al. (2003). Throughout the thesis, we indeed use Smith et al. fitting formula which has been shown to be in excellent agreement with the results from numerical simulations.

3.3 Formation of haloes in the Spherical Collapse Model

In what follows we discuss the spherical collapse model as a specific case for which an analytical description can be used to describe the non-linear evolution of the density field. Rather than a detailed

description of the non-linear regime of structure formation, this simplified model must be considered as a guideline to understand the behavior of perturbations in the non-linear regime.

Let us consider a spherically symmetric perturbation with a top hat profile of radius R and initial overdensity δ. For simplicity, we assume that the background universe has a critical density, $\Omega_{\rm backgr} = 1$. The perturbed region evolves independently of the behavior of the surrounding region, behaving as a universe with critical overdensity $\Omega_{\rm p} = \Omega_{\rm backgr} + \delta$. In other words, the spherical overdensity can be treated as a closed universe which eventually will collapse. For such a universe, one can use parametric solutions to describe the behavior of the size of the perturbation, r, with time, t. Such parametric solutions have the form

$$\begin{aligned} r &= A(1-\cos\theta)\,, \\ t &= B(\theta-\sin\theta)\,, \end{aligned} \tag{3.32}$$

where the development angle θ goes from 0 to 2π and $A^3 = GMB^2$, with M the mass within the spherical perturbation. Note that the value $\theta = \pi$ corresponds to the maximum expansion. We refer to this point as the *turnaround*. If one assume that the overdensity evolves linearly, it turns out that the value of the overdensity at the time of the turnaround is about unity. Basically this tells us that the end of the linear regime, $\delta \sim 1$, corresponds roughly to the moment in which the perturbation breaks away from the background expansion (the turnaround point). Note, however, that if only gravity operates, the spherical perturbation will collapse to a singularity at $\theta = 2\pi$. The value of the linearly evolved overdensity at this time is $\delta \sim 1.69$. This means that the value 1.69 corresponds to the complete gravitational collapse of a spherical perturbation. In reality, collapse to singularity will never occur. Clearly, dissipative physics will eventually intervene. If we calculate the point at which the potential energy is twice the kinetic energy (virial equilibrium), we find that this happens at half the maximum radius. One can show that the non-linear overdensity of the collapsing object at the virial equilibrium, $\Delta_{\rm V}$, is about 178. This kind of analysis was first presented in the late sixties by Peebles and collaborators and later on by Gunn and Gott (1972). Note that in the nineties these results have been confirmed by numerical simulations (Cole and Lacey 1996; Eke et al 1996). Note that, although we used the expression for an homogeneous universe with $\Omega_{\rm backgr} = 1$, the result is more general. However, extending these calculations into more general cosmologies requires the use of numerical solutions. In general, the linear overdensity required for the collapse to happen ($\delta \sim 1.69$) does not depend drastically on the cosmological model. It is indeed common practice to use this value to identify dark matter halo formation. Moreover, the value for $\Delta_{\rm V}$ is often used to define the mass of a dark matter halo. Accordingly, throughout the thesis, we define dark matter haloes as spheres with an average overdensity of 180, with a mass given by

$$M = \frac{4\pi}{3}\,(180\overline{\rho})\,r_{180}^3\,. \tag{3.33}$$

Here r_{180} is the radius of the halo. We assume that dark matter haloes follow the NFW (Navarro, Frenk & White 1997) density distribution

$$\rho(r) = \frac{\overline{\delta}\,\overline{\rho}}{(r/r_*)(1 + r/r_*)^2}\,, \tag{3.34}$$

where r_* is a characteristic radius and $\overline{\delta}$ is a dimensionless amplitude which can be expressed in terms of the halo concentration parameter $c_{\rm dm} \equiv r_{180}/r_*$ as

$$\overline{\delta} = \frac{180}{3}\,\frac{c_{\rm dm}^3}{\ln(1 + c_{\rm dm}) - c_{\rm dm}/(1 + c_{\rm dm})}\,. \tag{3.35}$$

Numerical simulations show that $c_{\rm dm}$ is correlated with halo mass. For example, Macciò et al.(2007) have used a series of cosmological numerical simulations for a flat cold dark matter cosmology to investigate the concentration, $c_{\rm dm}$, in the halo mass range $[\sim 10^9, \sim 10^{13}]h^{-1}M_\odot$ (these boundaries are enforced by the numerical resolution and the volume of the numerical simulation). Fig. 3.2 shows the relation $c_{\rm dm} - M$ for the model proposed by Macciò et al.(2007). Although this dependence has been found at redhift zero, it has been shown that it holds at higher redshifts ($z < 2$) and it scales as $(1 + z)^{-1}$. Throughout this thesis, we use the relation given by Macciò et al.(2007), converted to our definition of halo mass.

If the dark matter follows a NFW density profile, the halo mass within r_{180} can be written as

$$M = 4\pi\overline{\delta}\,\overline{\rho}\,r_*^3\left[\ln(1 + c_{\rm dm}) - \frac{c_{\rm dm}}{1 + c_{\rm dm}}\right]. \tag{3.36}$$

3.4 The Halo Mass Function

The halo mass function, $n(M)$, gives the number of haloes in a comoving volume in the mass interval $[M, M + {\rm d}M]$. In what follows, we summarize the simple idea proposed by Press and Schechter (1974) in which the comoving number density of virialized haloes is analytically predicted on the basis of some statistical arguments. Despite several shortcomings (which we will mention soon), the Press-Schechter formalism is still widely used because it provides an intuitive way to think about the abundance of haloes in the universe.

Press-Schechter theory simply asserts that if we smooth the linear density field on some scale R (and accordingly M), then the fraction of space in which the smoothed density field exceeds a given threshold, $\delta_{\rm c}$, is in collapsed objects of mass greater than M. If one assumes that the linear density field is a Gaussian field, this fraction, $F(M, z)$, can be calculated analytically:

$$F(M, z) = \frac{1}{2}{\rm erfc}\left(\frac{\delta_{\rm c}}{\sqrt{2}\sigma_R(z)}\right), \tag{3.37}$$

where erfc is the complementary error function. This equation represents maybe the major problem of the Press-Schechter theory. In fact, for a Gaussian field, half the volume of the universe is necessarily

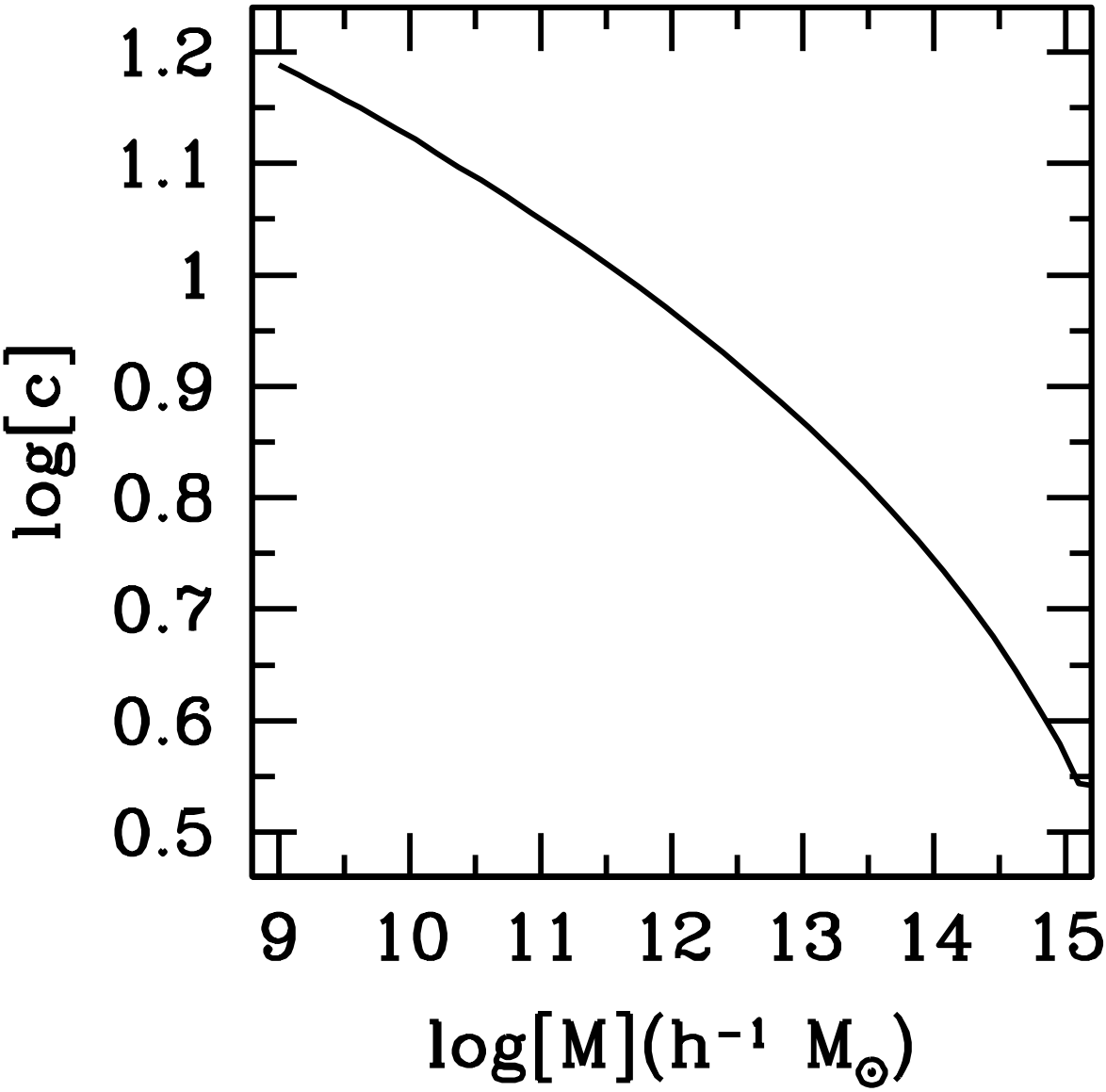

Figure 3.2: The concentration-halo mass relation proposed by Macciò et al. (2007).

underdense, and will never exceed the threshold regardless of how much the density field evolves. Consequently, only half of the mass in the universe can form collapsed haloes. This is, obviously, not the case in a realistic picture of structure formation in the universe. The expression in eq. (3.37) needs to be multiplied by a factor of 2:

$$\tilde{F}(M,z) = 2F(M,z) = \mathrm{erfc}\left(\frac{\delta_\mathrm{c}}{\sqrt{2}\sigma_R(z)}\right), \tag{3.38}$$

Although one could argue that this problem is enough to invalidate the Press-Schechter theory, the "modified" expression was accepted in the astronomical community because in good agreement with results of numerical simulations. More importantly, several studies after the pioneering work of Press and Schechter have been able to justify the factor of two with statistical arguments known as absorbing barrier (Sheth & Tormen 1999). We do not describe here in further detail these solutions. In fact, we are interested in the Press-Schechter formalism only to get insights into the expected number density of haloes of given masses. For all practical purposes, we will always use results from numerical simulations. In this respect, it is worth mentioning that the modified Press-Schecther formalism exhibits small departures from the results of the most recent numerical simulations. Thus, we conclude that this simple analytical model is perhaps able to capture the main essence of the problem but it fails in providing accurate detailed predictions.

To conclude, we provide here the expression for the comoving number density of haloes of mass M, per mass interval $\mathrm{d}M$, at redshift z, as derived from the modified Press and Schechter formalism. By differentiating eq. (3.38) with respect to the filtering mass, we obtain the change in the volume fraction of the universe , brought about by increasing the filter mass by $\mathrm{d}M$:

$$\begin{aligned}
\frac{\partial\tilde{F}(M,z)}{\partial M} &= \frac{\mathrm{d}\sigma_R(z)}{\mathrm{d}M}\frac{\partial\tilde{F}(M,z)}{\partial\sigma_R(z)} = \\
&= \frac{\mathrm{d}\sigma_R(z)}{\mathrm{d}M}\frac{\partial\,\mathrm{erfc}[\delta_\mathrm{c}/(\sqrt{2}\sigma_\mathrm{R}(z)]}{\partial\sigma_R(z)} = \\
&= \frac{\mathrm{d}\sigma_R(z)}{\mathrm{d}M}\sqrt{\frac{2}{\pi}}\frac{\delta_\mathrm{c}}{\sigma_R^2(z)}\exp\left[-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2(z)}\right],
\end{aligned} \tag{3.39}$$

where we have used the fact that

$$\frac{\mathrm{d}\,\mathrm{erfc}(y)}{\mathrm{d}y} = -\frac{2}{\sqrt{\pi}}\exp(y^2)\,. \tag{3.40}$$

Dividing eq. (3.39) by the mean volume occupied by M, i.e. M/ρ_0, we obtain the comoving number density of haloes of mass M:

$$n(M) = \frac{\mathrm{d}\sigma_R(z)}{\mathrm{d}M}\sqrt{\frac{2}{\pi}}\frac{\delta_\mathrm{c}}{\sigma_R^2(z)}\frac{\rho_0}{M}\exp\left[-\frac{\delta_\mathrm{c}^2}{2\sigma_R^2(z)}\right]. \tag{3.41}$$

Note that the term $\mathrm{d}\sigma_R(z)/\mathrm{d}M$ requires to specify the window function (see § 2 of this chapter).

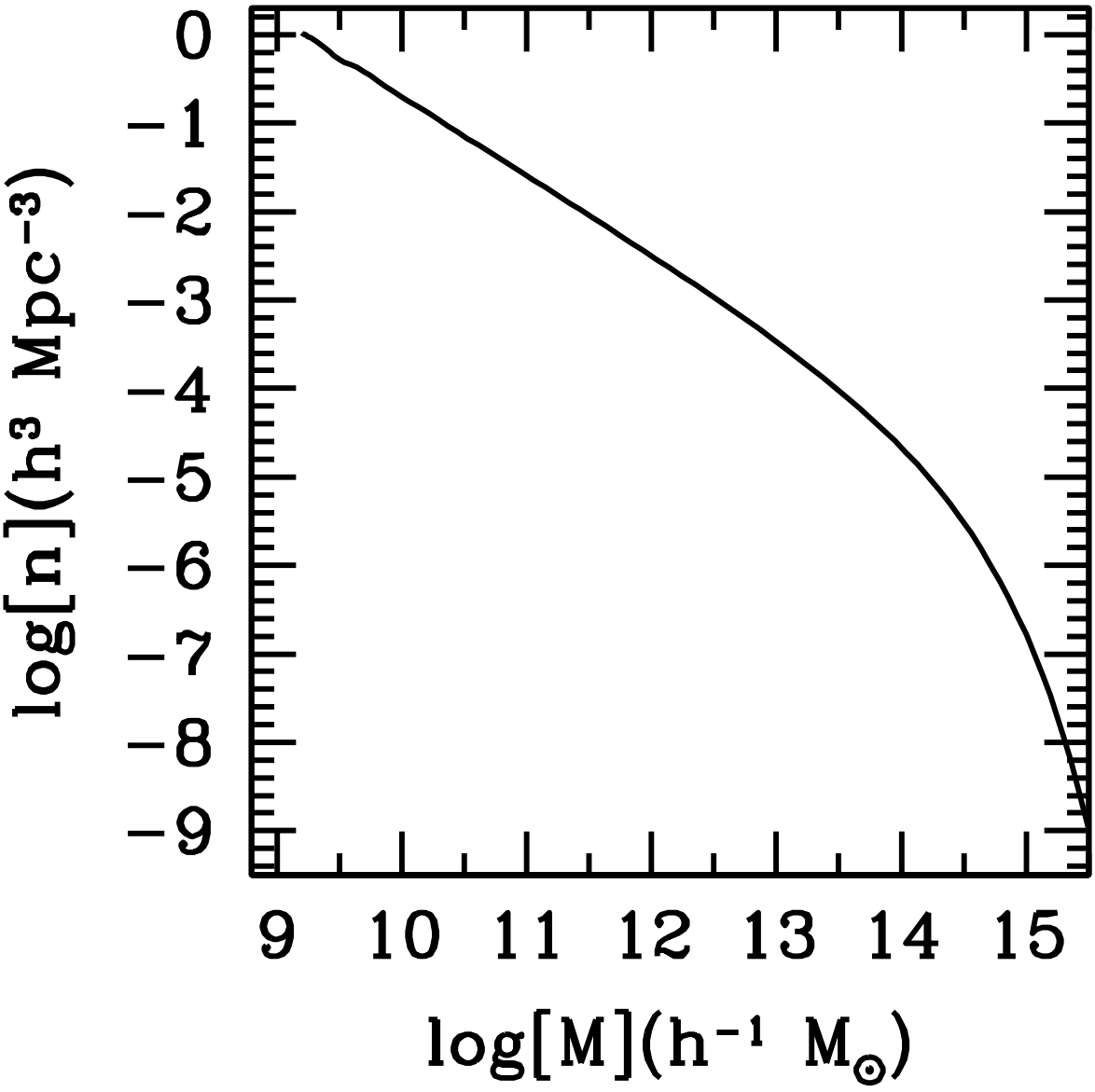

Figure 3.3: The halo mass function proposed by Sheth & Tormen (2002).

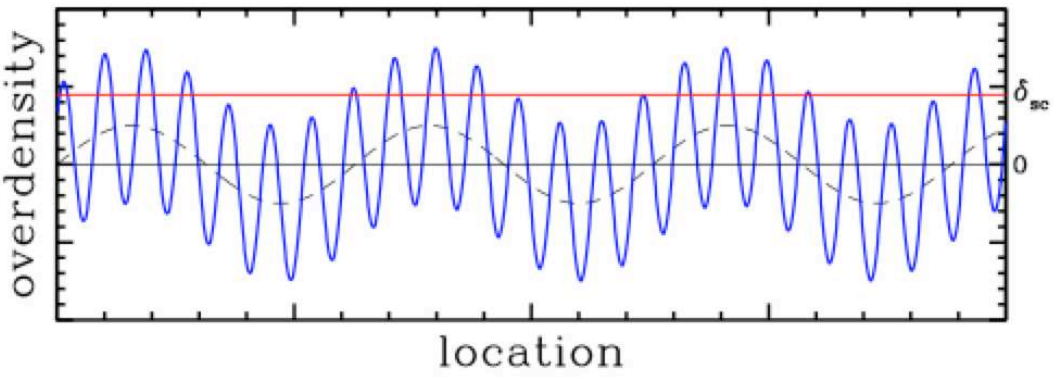

Figure 3.4: Skematic representation of the origin of the halo bias. The blue solid line represents a superposition of a small scale fluctuation with a large scale fluctuation (indicated by the dashed gray line). The threshold for the spherical collapse, δ_{sc}, is indicated by the red straight line.

The Press-Schechter formalism thus allows one to make predictions about the shape of the mass function of dark matter haloes. Two assumptions are necessary: (i) haloes collapse spherically; (ii) initial density fluctuations are Gaussian and small. While the predicted mass function is overall in good agreement with numerical simulations, small differences can be observed. It has been suggested that this discrepancy between theory and simulations can be reduced substantially if haloes are assumed to form from an ellipsoidal, rather than a spherical collapse. Sheth, Mo & Tormen (2001) have incorporated the ellipsoidal collapse into the Press-Schechter formalism and they have shown that with these modifications theoretical predictions are in good agreement with numerical simulations (Sheth & Tormen 2002). Throughout the thesis, we use the expressions introduced by Sheth & Tormen (2002) to describe the halo mass function. Fig. 3.3 shows the halo mass function calculated via the fitting formula of Sheth & Tormen (2002) for the WMAP3 set of cosmological parameters.

3.5 The Clustering of Dark Matter Haloes

In the previous sections, we have explained the formation of dark matter haloes and their spatial abundance. Here, we show that the dark matter halo distribution is biased with respect to the underlying dark matter distribution. Consider, for example, the schematic illustration in fig. 3.4. The one dimensional dark matter overdensity is plotted as a function of the spatial location. The blue solid line represents a superposition of a small scale fluctuation with a large scale fluctuation (indicated by the dashed gray line). Since haloes can form only where the density field exceeds a certain threshold (here indicated with the red straight line), the location of dark matter haloes is highly correlated with the location of the peaks of the large scale fluctuations. In other words, on large scales, the correlation function of haloes and that of dark matter particles are related. Since it has been shown that on scales

much larger than the typical extension of a dark matter halo, the overdensity of haloes is linearly proportional to the overdensity of mass (e.g. Cooray & Sheth 2001) with the constant of proportionality depending on the mass of the haloes and on the redshift at which they virialize:

$$\delta_{\rm h}(m,z) = b(m,z)\delta\,, \tag{3.42}$$

one expects that, on large scales,

$$\xi_{\rm hh}(r) = b^2(m,z)\xi_{\rm dm}(r)\,, \tag{3.43}$$

where $\xi_{\rm hh}$ is the halo-halo correlation and $\xi_{\rm dm}$ is the dark matter correlation.

The function $b(m,z)$ is known in the literature as *asymptotic bias function*. Different fitting functions have been proposed over the last ten years. In this thesis, we use the relation proposed by Sheth & Tormen (2002). Notice that the main features of this function can be predicted analytically on the basis of the Press-Schechter formalism. In particular, one expects that massive haloes have $b(m,z) > 1$ and therefore they are said to be biased with respect to the dark matter distribution. On the contrary, less massive haloes have $b(m,z) < 1$ and they are said anti-biased with respect to the dark matter distribution. Fig. 3.5 shows the halo bias relation calculated with the fitting function proposed by Sheth & Tormen (2002).

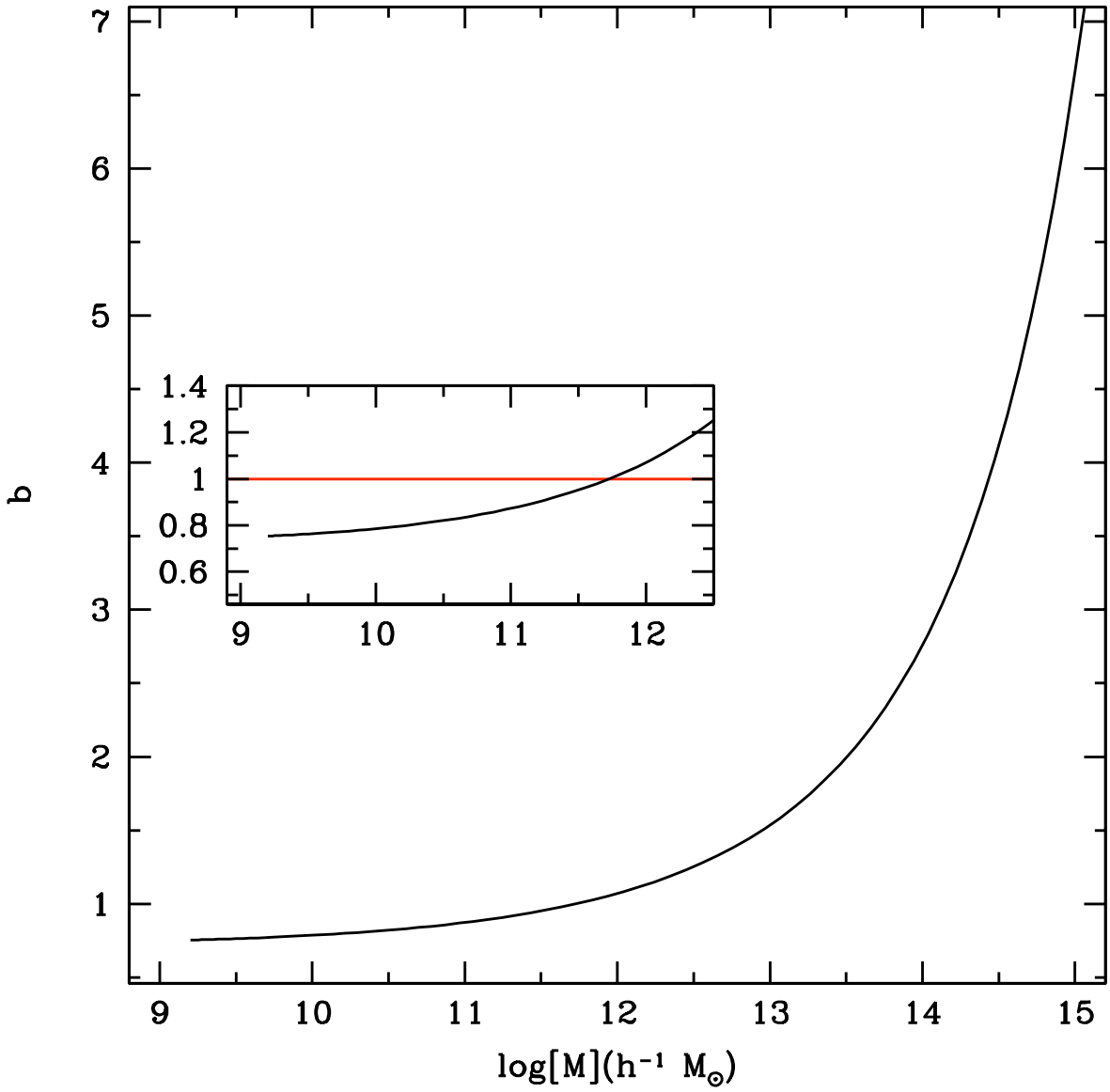

Figure 3.5: The black solid line represents the halo bias function proposed by Sheth & Tormen (2002). In the inner panel, a close-up of the halo bias function is shown. The red straight line represents the unbiased relation. Thus, low mass haloes are anti-biased and high mass haloes are biased with respect to the dark matter distribution.

4

The Analytical Model

The analytical model presented in this chapter is embedded in the framework of the *halo model* (see Cooray & Sheth 2001 for a review). With this term, one refers to a model in which all the dark matter in the universe is partitioned up into distinct units called haloes. The left panel of fig. 4.1 shows the dark matter distribution derived via numerical simulations and color coded according to density. Such a complicated distributions simplifies drastically in the halo model view. In fact, the right panel shows the corresponding density field when only spherical overdense regions (haloes) above a given mass are considered. Basically, in such a simplified configuration, one can describe the statistics of the dark matter distribution in the universe in two steps: on small scales one needs to know the spatial distribution within the haloes, whereas on large scales the important ingredient is the spatial distribution of the haloes . Throughout the thesis, we will refer to these two regimes as one halo (1h) and two halo (2h). The choice for this name will became obvious during the description of the model.

The halo model provides a natural way to link the statistical properties of the dark matter distribution with the statistical properties of the observed galaxies. In fact, baryons can cool and form stars only when *"trapped"* in the gravitational potential of dark matter haloes. As a result, all galaxies are expected to be embedded in dark matter haloes (see White & Rees 1978). More massive haloes may contain many galaxies, whereas low mass haloes may contain fewer or no galaxies. Our interests concern the statistical properties of galaxies and their relation with the underlying dark matter distribution. For this reason, we investigate the clustering properties of galaxies and of dark matter around galaxies. In the next sections, we will describe the clustering of galaxies (§ 4.1), dark matter (§ 4.2), and dark matter around galaxies (§ 4.3) in the framework of the halo model. In particular, we introduce the expressions for the 2-point galaxy correlation function and for the 2-point galaxy-dark matter cross-correlation function. These last two quantities allows us to compute the projected galaxy correlation function w_p and the excess surface density, $\Delta\Sigma$, respectively. The halo occupation statistics required by the correlation functions is derived via the conditional luminosity function in § 4.5, where we also provide the theoretical predictions for the galaxy luminosity function.

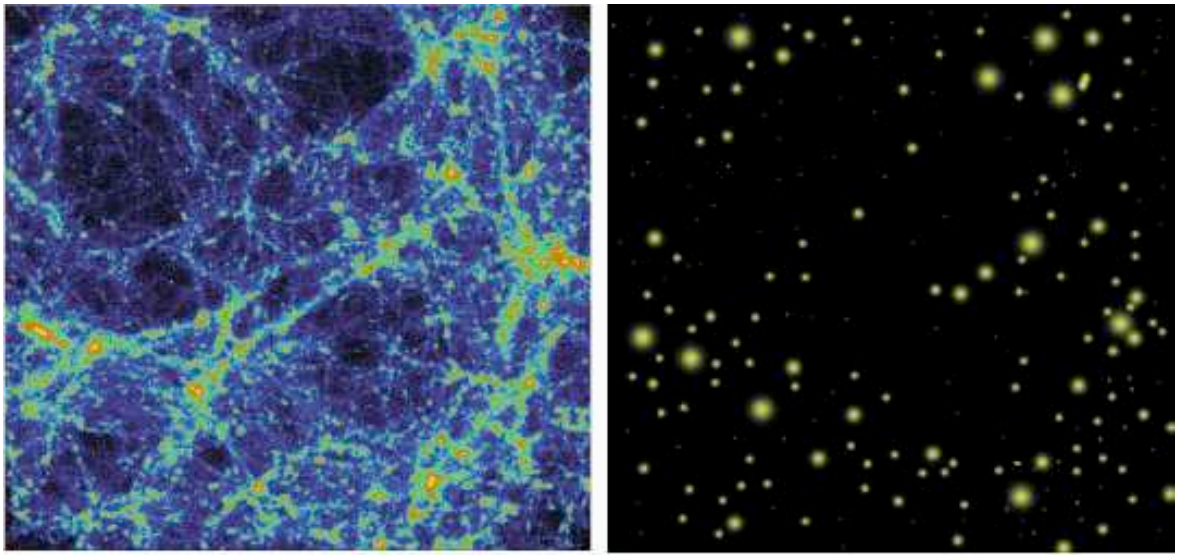

Figure 4.1: The distribution of dark matter in the complex configuration derived by numerical simulations (left) and in the halo model view (right). [Ref. Cooray & Sheth 2001]

4.1 Galaxy Clustering

We begin the discussion about the clustering of galaxies by introducing the galaxy 2-point correlation function in the halo model formalism. We recall that the galaxy correlation function has been defined in chapter 2 from an observational point of view:

$$1 + \xi_{\rm gg}(r, L_1, L_2) = \frac{DD(r, L_1, L_2)}{DR(r, L_1, L_2)} \tag{4.1}$$

where $DD(r, L_1, L_2)$ is the number of galaxy pairs (with luminosity within $[L_1, L_2]$) at distance r as measured from the data and $DR(r, L_1, L_2)$ is the number of galaxy pairs (with luminosity within $[L_1, L_2]$) at distance r as expected in a random distribution with the same number density of galaxies as in the data. Note that the galaxy correlation function is defined as a function of the distance r and the luminosity range $[L_1, L_2]$. This is done because the data described in chapter 2 are obtained using galaxies selected in a given absolute magnitude bin. Since our model will ultimately be compared with these data, theoretical quantities as the correlation function incorporates this information in their definition.

In the halo model, it is common to describe galaxy properties in relation with their host dark matter halo. For example, it is commonly accepted that a galaxy can either reside at the center of the host dark matter halo or it can orbit around the center. One refers to the former as *central galaxies* and to the latter as *satellite galaxies* (see fig. 4.2 for an illustration of this concept). In modelling the galaxy 2-point correlation function, it is recommended to take into account the distinction in central and satellite galaxies. This means that $DD(r, L_1, L_2)$ can be split into different terms each of

which describes different kind of pairs. In fact, inside the same dark matter halo, one can have pairs consisting of a central galaxy with a satellite galaxy, namely one halo central-satellite term (1h, c-s); or one can have pairs using satellite galaxies only, namely one halo satellite-satellite term (1h, s-s). Moreover, in considering pairs between galaxies belonging to different haloes one can connect central galaxies, namely two halo central-central term (2h, c-c); a central galaxy with satellite galaxies, namely two halo central-satellite term (2h, c-s); satellite galaxies between each others, namely two halo satellite-satellite term (2h, s-s). Fig. 4.3 provides a schematic representation of all these terms.

We thus have

$$\begin{aligned} DD(r, L_1, L_2) = N_{\rm pairs}(r, L_1, L_2) &= N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2) + N^{\rm 1h,s-s}_{\rm pairs}(r, L_1, L_2) \\ &+ N^{\rm 2h,c-c}_{\rm pairs}(r, L_1, L_2) + N^{\rm 2h,c-s}_{\rm pairs}(r, L_1, L_2) + N^{\rm 2h,s-s}_{\rm pairs}(r, L_1, L_2)\,, \end{aligned} \tag{4.2}$$

where the notation follows from the discussion above.

As clear from eq (4.1), in order to obtain the galaxy correlation function one needs to specify the number of galaxy pairs in a random distribution, $DR(r, L_1, L_2)$. In doing that, one has

$$DR(r, L_1, L_2) = \left[\int \langle N_{\rm tot}\rangle_M(L_1, L_2)\, n(M)\, {\rm d}M\right]^2 = \bar{n}^2_{\rm tot}(L_1, L_2)\,, \tag{4.3}$$

where $\bar{n}_{\rm tot}(L_1, L_2)$ is the average number of galaxies in the luminosity bin $[L_1, L_2]$. Note that $\bar{n}_{\rm tot}(L_1, L_2)$ does not depend on r due to the fact that we are considering a sample of randomly distributed galaxies and we are applying the ergodic principle (see chapter 3). Moreover, we have not split $DR(r, L_1, L_2)$ into different terms. This is a somewhat technical, and to some extent conceptual, choice. In fact, our aim is to describe the complete 2-point galaxy correlation function. For this purpose, one can write

$$\begin{aligned} 1 + \xi_{\rm gg}(r, L_1, L_2) &= \frac{N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2) + N^{\rm 1h,s-s}_{\rm pairs}(r, L_1, L_2)}{\bar{n}^2_{\rm tot}(L_1, L_2)} + \\ &+ \frac{N^{\rm 2h,c-c}_{\rm pairs}(r, L_1, L_2) + N^{\rm 2h,c-s}_{\rm pairs}(r, L_1, L_2) + N^{\rm 2h,s-s}_{\rm pairs}(r, L_1, L_2)}{\bar{n}^2_{\rm tot}(L_1, L_2)} = \\ &= \tilde{\xi}^{\rm 1h,c-s}_{\rm gg}(r, L_1, L_2) + \tilde{\xi}^{\rm 1h,s-s}_{\rm gg}(r, L_1, L_2) + \\ &+ \tilde{\xi}^{\rm 2h,c-c}_{\rm gg}(r, L_1, L_2) + \tilde{\xi}^{\rm 2h,c-s}_{\rm gg}(r, L_1, L_2) + \tilde{\xi}^{\rm 2h,s-s}_{\rm gg}(r, L_1, L_2)\,. \end{aligned} \tag{4.4}$$

where $\tilde{}$ is used as a reminder that this terms represent only a convenient split of the total correlation function.

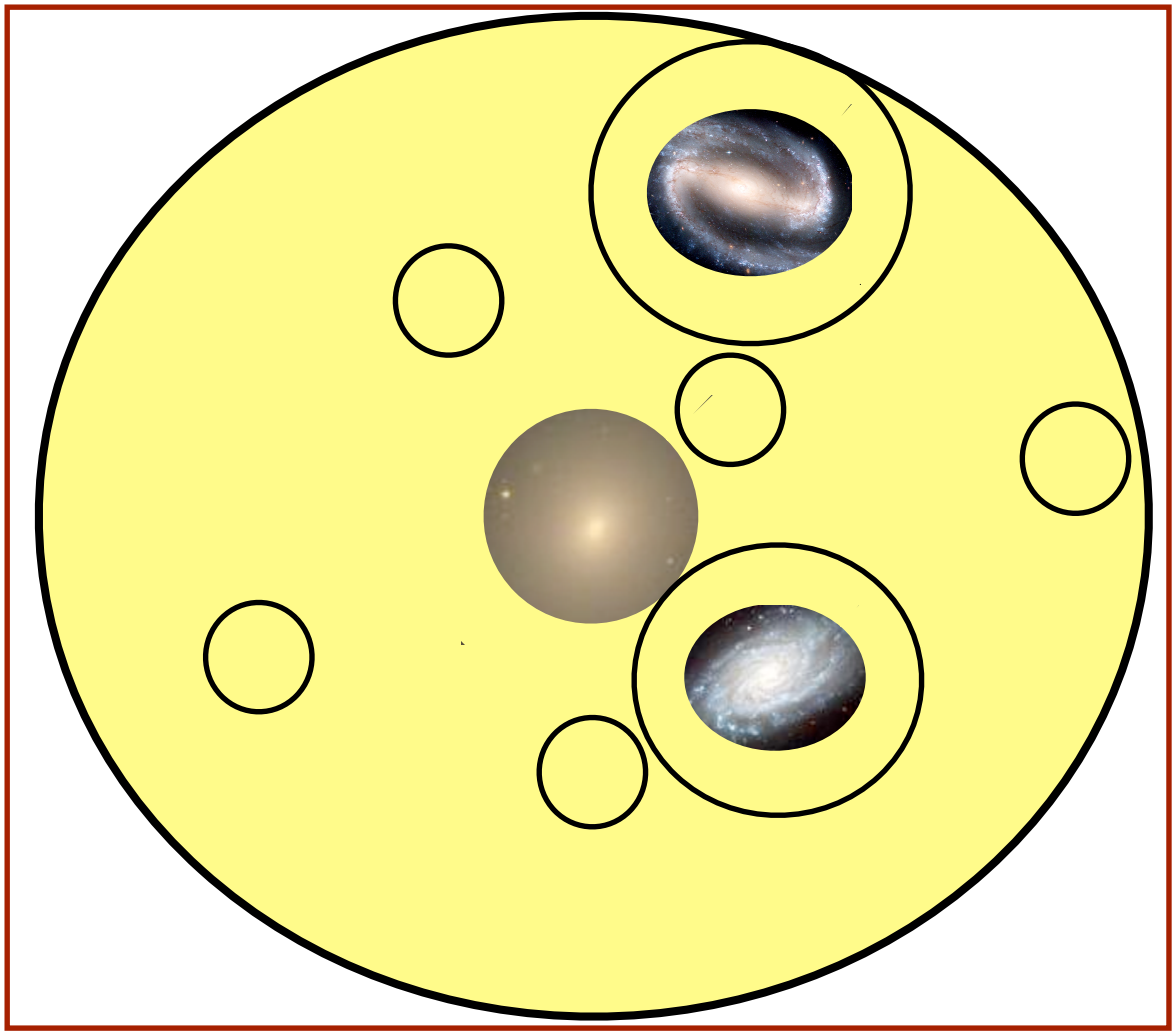

Figure 4.2: *Halo model* representation of a dark matter halo. A central galaxy is sitting at the center of the halo. There are sub-haloes orbiting around the center of the dark matter halo. Some of them may host a (satellite) galaxy and some of them may not because their potential well is not deep enough "to capture" baryons.

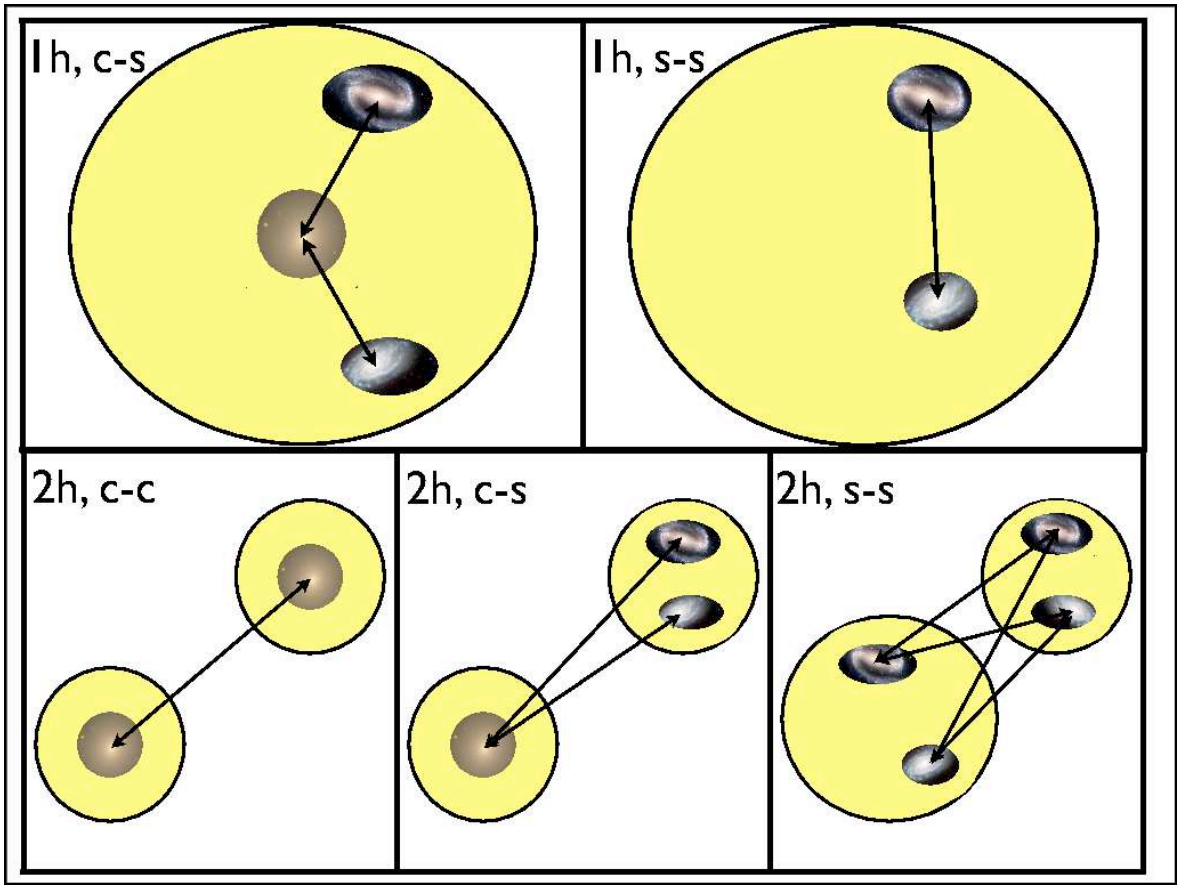

Figure 4.3: Schematic representation of the different terms of the galaxy correlation function.

Note, however, that one could also define

$$\begin{aligned}
&\frac{N^{\rm 1h,c-s}_{\rm pairs}(r,L_1,L_2)}{\bar{n}_{\rm c}(L_1,L_2)\,\bar{n}_{\rm s}(L_1,L_2)} \equiv \breve{\xi}^{\rm 1h,c-s}_{\rm gg}(r,L_1,L_2) \qquad , \qquad \frac{N^{\rm 1h,s-s}_{\rm pairs}(r,L_1,L_2)}{\bar{n}^2_{\rm s}(L_1,L_2)} \equiv \breve{\xi}^{\rm 1h,s-s}_{\rm gg}(r,L_1,L_2) \\
&\frac{N^{\rm 2h,c-c}_{\rm pairs}(r,L_1,L_2)}{\bar{n}^2_{\rm c}(L_1,L_2)} \equiv \breve{\xi}^{\rm 2h,c-c}_{\rm gg}(r,L_1,L_2) \qquad , \qquad \frac{N^{\rm 2h,c-s}_{\rm pairs}(r,L_1,L_2)}{\bar{n}_{\rm c}(L_1,L_2)\,\bar{n}_{\rm s}(L_1,L_2)} \equiv \breve{\xi}^{\rm 2h,c-s}_{\rm gg}(r,L_1,L_2)\,, \\
&\frac{N^{\rm 2h,s-s}_{\rm pairs}(r,L_1,L_2)}{\bar{n}^2_{\rm s}(L_1,L_2)} \equiv \breve{\xi}^{\rm 2h,s-s}_{\rm gg}(r,L_1,L_2)\,,
\end{aligned} \tag{4.5}$$

where $\bar{n}_{\rm c}(L_1,L_2)$ and $\bar{n}_{\rm s}(L_1,L_2)$ are the average number of central and satellite galaxies, respectively. However, in this case, in order to obtain the total galaxy correlation function as a sum of the single terms, one must consider the fraction of central and satellite galaxies with luminosity in the range $[L_1,L_2]$:

$$f_{\rm c}(L_1,L_2) = \frac{\bar{n}_{\rm c}(L_1,L_2)}{\bar{n}_{\rm tot}(L_1,L_2)} \qquad \text{and} \qquad f_{\rm s}(L_1,L_2) = \frac{\bar{n}_{\rm s}(L_1,L_2)}{\bar{n}_{\rm tot}(L_1,L_2)}\,. \tag{4.6}$$

We thus have

$$\begin{aligned}
1+\xi_{\rm gg}(r,L_1,L_2) &= f_{\rm c}(L_1,L_2)\,f_{\rm s}(L_1,L_2)\,\breve{\xi}^{\rm 1h,c-s}_{\rm gg}(r,L_1,L_2) + f^2_{\rm s}(L_1,L_2)\,\breve{\xi}^{\rm 1h,s-s}_{\rm gg}(r,L_1,L_2) + \\
&+ f^2_{\rm c}(L_1,L_2)\,\breve{\xi}^{\rm 2h,c-c}_{\rm gg}(r,L_1,L_2) + f_{\rm c}(L_1,L_2)\,f_{\rm s}(L_1,L_2)\,\breve{\xi}^{\rm 2h,c-s}_{\rm gg}(r,L_1,L_2) + \\
&+ f^2_{\rm s}(L_1,L_2)\,\breve{\xi}^{\rm 2h,s-s}_{\rm gg}(r,L_1,L_2) = \\
&= \tilde{\xi}^{\rm 1h,c-s}_{\rm gg}(r,L_1,L_2) + \tilde{\xi}^{\rm 1h,s-s}_{\rm gg}(r,L_1,L_2) + \\
&+ \tilde{\xi}^{\rm 2h,c-c}_{\rm gg}(r,L_1,L_2) + \tilde{\xi}^{\rm 2h,c-s}_{\rm gg}(r,L_1,L_2) + \tilde{\xi}^{\rm 2h,s-s}_{\rm gg}(r,L_1,L_2)\,.
\end{aligned} \tag{4.7}$$

Throughout the thesis, we use the split presented in eq. (4.4). In the following sections, we shall introduce and describe the single terms in detail.

Before moving to the description of the different terms, we emphasizes that the split into one halo and two halo term has been used by many other authors in this field. Similarly, the division into central and satellite has been already applied in the description of galaxy correlation function. However, all previous studies have assumed that the distinction into central and satellite term is relevant only on small scales (one halo term) and it is negligible on large scales (two halo term). With the quality of the data that we use in this thesis, such an assumption is not accurate and, also on relatively large scales, the split into central and satellite is extremely important for an accurate modelling of the galaxy correlation. From the description of the single terms it will become apparent that it is often convenient to work in Fourier space. Since the 2-point correlation function and the (2-point) power spectrum are

Fourier pairs (see chapter 3), we shall also introduce the galaxy (2-point) power spectrum:

$$\begin{aligned} P_{\rm gg}(k) &= \tilde{P}^{\rm 1h,c-s}_{\rm gg}(k) + \tilde{P}^{\rm 1h,s-s}_{\rm gg}(k) \\ &+ \tilde{P}^{\rm 2h,c-c}_{\rm gg}(k) + \tilde{P}^{\rm 2h,c-s}_{\rm gg}(k) + \tilde{P}^{\rm 2h,s-s}_{\rm gg}(k) = \\ &= f_{\rm c}(L_1,L_2)\, f_{\rm s}(L_1,L_2)\, \check{P}^{\rm 1h,c-s}_{\rm gg}(k) + f^2_{\rm s}(L_1,L_2)\, \check{P}^{\rm 1h,s-s}_{\rm gg}(k) \\ &+ f^2_{\rm c}(L_1,L_2)\, \check{P}^{\rm 2h,c-c}_{\rm gg}(k) + f_{\rm c}(L_1,L_2)\, f_{\rm s}(L_1,L_2)\, \check{P}^{\rm 2h,c-s}_{\rm gg}(k) + \\ &+ f^2_{\rm s}(L_1,L_2)\, \check{P}^{\rm 2h,s-s}_{\rm gg}(k), \end{aligned} \tag{4.8}$$

where $\tilde{.}$ and $\check{.}$ follow the same convention used for the correlation function.

In the next sections, we define all the terms of the galaxy correlation function and the corresponding power spectra.

4.1.1 The one halo central-satellite term

Let us evaluate the number of galaxy pairs, $N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2)$, consisting of a central galaxy and a satellite galaxy at distance r and with luminosity within $[L_1, L_2]$. In the halo model, this quantity can be written as a sum over all haloes of the number of these pairs per halo:

$$N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2) = \sum_{\rm all\ haloes} N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2|M) \tag{4.9}$$

where M is the halo mass and here is used only as a label to identify the halo. Let us begin the derivation of $N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2|M)$ considering simply a halo of mass M for which we indicate the radial distribution of satellite galaxies as $N_{\rm s}(L_1, L_2, M)\, \hat{n}_{\rm s}(r, L_1, L_2|M)$. Let us now average the number of satellite galaxies over all haloes of mass M:

$$\sum_{N_{\rm s}=0}^{\infty} P(N_{\rm s}|M)\, N_{\rm s}(L_1, L_2, M) \equiv \langle N_{\rm s}\rangle_M(L_1, L_2)\,. \tag{4.10}$$

This equation thus defines the number of satellite galaxies with luminosity in $[L_1, L_2]$ which, on average, reside in a halo of mass M.

To evaluate $N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2|M)$, let us first consider the simplistic case in which every halo, regardless of the mass, has a central galaxy in the considered luminosity bin and the satellite galaxies, on average, are distributed according to the description provided above. We then have

$$N^{\rm 1h,c-s}_{\rm pairs}(r, L_1, L_2|M) = \langle N_{\rm s}\rangle_M(L_1, L_2)\, \hat{n}_{\rm s}(r, L_1, L_2|M)\,. \tag{4.11}$$

Note that throughout the thesis, we do not consider the effect of luminosity segregation. This means that we do not allow satellite galaxies of different luminosity to have different radial profile. This means that we can conveniently rewrite $\hat{n}_{\rm s}(r, L_1, L_2|M) = \hat{n}_{\rm s}(r|M)$.

In a realistic model, however, one has to consider the probability that a halo of mass M contains a central galaxy with luminosity within $[L_1, L_2]$. We can write this probability as $\langle N_{\rm c}\rangle_M(L_1, L_2)$, which

is the number of central galaxies that, on average, resides in haloes of mass M. To be more explicit, imagine to have a certain number of haloes of mass M and only a fraction of these haloes actually contains a central galaxy in the considered luminosity bin. Since in our model there is at most one central galaxy in a given halo, we have

$$\sum_{N_\mathrm{c}=0}^{\infty} P(N_\mathrm{c}|M)\, N_\mathrm{c}(L_1, L_2, M) \equiv \langle N_\mathrm{c}\rangle_M(L_1, L_2)\,. \tag{4.12}$$

In this realistic description of the number of central-satellite galaxy pairs, we thus have

$$N_\mathrm{pairs}^\mathrm{1h,c-s}(r, L_1, L_2|M) = \langle N_\mathrm{c}\rangle_M(L_1, L_2)\,\langle N_\mathrm{s}\rangle_M(L_1, L_2)\,\hat{n}_\mathrm{s}(r|M)\,. \tag{4.13}$$

Finally, by summing over all haloes, we take into account the number density of haloes of different mass, i.e. the halo mass function, $n(M)$, introduced in chapter 3. We thus have

$$\begin{aligned} N_\mathrm{pairs}^\mathrm{1h,c-s}(r, L_1, L_2) &= \sum_{\text{all haloes}} N_\mathrm{pairs}^\mathrm{1h,c-s}(r, L_1, L_2|M) = \\ &= \int \langle N_\mathrm{c}\rangle_M(L_1, L_2)\,\langle N_\mathrm{s}\rangle_M(L_1, L_2)\,\hat{n}_\mathrm{s}(r|M)\,n(M)\,\mathrm{d}M\,. \end{aligned} \tag{4.14}$$

The one halo central-satellite term of the galaxy 2-point correlation function then reads

$$\begin{aligned} \tilde{\xi}_\mathrm{gg}^\mathrm{1h,c-s}(r|M, L_1, L_2) &= \frac{N_\mathrm{pairs}^\mathrm{1h,c-s}(r, L_1, L_2)}{\bar{n}_\mathrm{tot}^2} = \\ &= \int \frac{\langle N_\mathrm{c}\rangle_M(L_1, L_2)\,\langle N_\mathrm{s}\rangle_M(L_1, L_2)}{\bar{n}_\mathrm{tot}^2}\,\hat{n}_\mathrm{s}(r|M)\,n(M)\,\mathrm{d}M = \\ &= f_\mathrm{c}(L_1, L_2)\,f_\mathrm{s}(L_1, L_2)\int \frac{\langle N_\mathrm{c}\rangle_M(L_1, L_2)\,\langle N_\mathrm{s}\rangle_M(L_1, L_2)}{\bar{n}_\mathrm{c}(L_1, L_2)\,\bar{n}_\mathrm{s}(L_1, L_2)}\,\hat{n}_\mathrm{s}(r|M)\,n(M)\,\mathrm{d}M = \\ &= f_\mathrm{c}(L_1, L_2)\,f_\mathrm{s}(L_1, L_2)\,\check{\xi}_\mathrm{gg}^\mathrm{1h,c-s}(r, L_1, L_2)\,. \end{aligned} \tag{4.15}$$

This expression is quite straightforward to calculate in real space. However, we define its Fourier space counterpart:

$$\begin{aligned} \tilde{P}_\mathrm{gg}^\mathrm{1h,c-s}(k|L_1, L_2) &= \int \frac{\langle N_\mathrm{c}\rangle_M(L_1, L_2)\,\langle N_\mathrm{s}\rangle_M(L_1, L_2)}{\bar{n}_\mathrm{tot}^2(L_1, L_2)}\,u_\mathrm{s}(k|M, L_1, L_2)\,n(M)\,\mathrm{d}M = \\ &= f_\mathrm{c}(L_1, L_2)\,f_\mathrm{s}(L_1, L_2)\int \frac{\langle N_\mathrm{c}\rangle_M(L_1, L_2)\,\langle N_\mathrm{s}\rangle_M(L_1, L_2)}{\bar{n}_\mathrm{c}(L_1, L_2)\,\bar{n}_\mathrm{s}(L_1, L_2)}\,u_\mathrm{s}(k|M, L_1, L_2)\,n(M)\,\mathrm{d}M = \\ &= f_\mathrm{c}(L_1, L_2)\,f_\mathrm{s}(L_1, L_2)\,\check{P}_\mathrm{gg}^\mathrm{1h,c-s}(k|L_1, L_2) \end{aligned} \tag{4.16}$$

where $u_\mathrm{s}(k|M, L_1, L_2)$ is the Fourier transform of the normalized radial distribution of satellite galaxies, $\hat{n}_\mathrm{s}(r|M, L_1, L_2)$,

$$u_\mathrm{s}(k|M) = 4\pi \int_0^{r_{180}} \hat{n}_\mathrm{s}(r|M)\,\frac{\sin kr}{kr} r^2\,\mathrm{d}r\,. \tag{4.17}$$

Throughout the thesis, for the number density distribution of the satellite galaxies, we assume the following functional form (see e.g., van den Bosch et al.2004):

$$n_{\rm s}(r|M) \propto \left(\frac{r}{\mathcal{R}r_*}\right)^{-\alpha} \left(1+\frac{r}{\mathcal{R}r_*}\right)^{\alpha-3} , \tag{4.18}$$

which scales as $n_{\rm s} \propto r^{-\alpha}$ and $n_{\rm s} \propto r^{-3}$ at small and large radii, respectively. The profile $n_{\rm s}(r|M)$ has an effective scale radius $\mathcal{R}r_*$. Observations of the number density distribution of satellite galaxies in clusters and groups seem to suggest that $n_{\rm s}(r|M)$ is in good agreement with a profile for which $\alpha = 1$ (e.g., Beers & Tonry 1986; Carlberg, Yee & Ellingson 1997a; van der Marel et al.2000; Lin, Mohr & Stanford 2004; van den Bosch et al.2005a). Furthermore, several studies have suggested that satellite galaxies may be distributed according to the profile in eq. (4.18) with $\mathcal{R} > 1$ (e.g., Yang et al.2005; Chen 2007; More et al.2008b). For our fiducial model we assume that satellite galaxies follow the profile in eq. (4.18) with $\alpha = 1$ and $\mathcal{R} = 1$. In §5.2.3.3 we examine how the results depend on α and $\mathcal{R}$.

4.1.2 The one halo satellite-satellite term

In explaining the one halo-satellite term of the galaxy correlation function we will follow exactly the same line of reasoning used for the one halo central-satellite term. The main difference resides in the fact that one needs to correlate satellite galaxies between each others. By definition, satellite galaxies do not reside at the center of their dark matter halo, but follow a number density distribution $n_{\rm s}(r|M)$. It is then intuitive that the number of satellite galaxy pairs at distance r depends on the convolution of $n_{\rm s}(r|M)$ with itself. Let us consider a single halo of mass M whose volume is V_M, we then have

$$N^{\rm 1h,s-s}_{\rm pairs}(r, L_1, L_2|M) = \int_{V_M} {\rm d}^3\mathbf{y}\, n_{\rm s}(\mathbf{y}, L_1, L_2|M)\, n_{\rm s}(\mathbf{y}+\mathbf{r}, L_1, L_2|M)\,. \tag{4.19}$$

Using the same arguments of eq. (4.10), we have

$$\begin{aligned}
N^{\rm 1h,s-s}_{\rm pairs}(r, L_1, L_2|M) &= \int_{V_M} {\rm d}^3\mathbf{y}\, n_{\rm s}(\mathbf{y}, L_1, L_2|M)\, n_{\rm s}(\mathbf{y}+\mathbf{r}, L_1, L_2|M) = \\
&= \int_{V_M} {\rm d}^3\mathbf{y} \sum_{N_{\rm s}=0}^{\infty} P(N_{\rm s}|M)\, N_{\rm s}(L_1, L_2, M)\hat{n}_{\rm s}(\mathbf{y}|M) \\
&\quad N_{\rm s}(L_1, L_2, M)\hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\
&= \int_{V_M} {\rm d}^3\mathbf{y} \sum_{N_{\rm s}=0}^{\infty} P(N_{\rm s}|M)\, N^2_{\rm s}(L_1, L_2, M)\hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\
&= \int_{V_M} {\rm d}^3\mathbf{y}\, \langle N^2_{\rm s}\rangle_M(L_1, L_2)\hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\
&= \langle N^2_{\rm s}\rangle_M(L_1, L_2) \int_{V_M} {\rm d}^3\mathbf{y}\, \hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\
&= \langle N_{\rm s}\rangle^2_M(L_1, L_2) \int_{V_M} {\rm d}^3\mathbf{y}\, \hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M)\,,
\end{aligned} \tag{4.20}$$

where the last equality holds under the assumption that satellite galaxies are Poisson distributed.

In order to quantify the total number of satellite galaxy pairs at a given distance r and in a luminosity bin $[L_1, L_2]$, we simply sum $N_{\rm pairs}^{\rm 1h,s-s}(r, L_1, L_2|M)$ over all haloes,

$$\begin{aligned} N_{\rm pairs}^{\rm 1h,s-s}(r, L_1, L_2) &= \sum_{\text{all haloes}} N_{\rm pairs}^{\rm 1h,s-s}(r, L_1, L_2|M) = \\ &= \sum_{\text{all haloes}} \langle N_{\rm s}\rangle_M^2(L_1, L_2) \int_{V_M} {\rm d}^3\mathbf{y}\, \hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\ &= \int n(M)\, {\rm d}M\, \langle N_{\rm s}\rangle_M^2(L_1, L_2) \int_{V_M} {\rm d}^3\mathbf{y}\, \hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M)\,. \end{aligned} \tag{4.21}$$

This expression allows us to write the one halo satellite-satellite term of the galaxy correlation function as

$$\begin{aligned} \tilde{\xi}_{\rm gg}^{\rm 1h,s-s}(r, L_1, L_2) &= \frac{N_{\rm pairs}^{\rm 1h,s-s}(r, L_1, L_2)}{\bar{n}_{\rm tot}^2(L_1, L_2)} = \\ &= \int n(M)\, {\rm d}M\, \frac{\langle N_{\rm s}\rangle_M^2(L_1, L_2)}{\bar{n}_{\rm tot}^2(L_1, L_2)} \int_{V_M} {\rm d}^3\mathbf{y}\, \hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\ &= f_{\rm s}^2(L_1, L_2) \int n(M)\, {\rm d}M\, \frac{\langle N_{\rm s}\rangle_M^2(L_1, L_2)}{\bar{n}_{\rm s}^2(L_1, L_2)} \int_{V_M} {\rm d}^3\mathbf{y}\, \hat{n}_{\rm s}(\mathbf{y}|M)\, \hat{n}_{\rm s}(\mathbf{y}+\mathbf{r}|M) = \\ &= f_{\rm s}^2(L_1, L_2)\, \check{\xi}_{\rm gg}^{\rm 1h,s-s}(r, L_1, L_2) \end{aligned} \tag{4.22}$$

Since convolution can be extremely tedious to calculate numerically, it is convenient to use the Fourier transform of eq. (4.22). We then obtain

$$\begin{aligned} \tilde{P}_{\rm gg}^{\rm 1h,s-s}(k, L_1, L_2) &= \int \frac{\langle N_{\rm s}\rangle_M^2(L_1, L_2)}{\bar{n}_{\rm tot}^2(L_1, L_2)} |u_{\rm s}(k|M)|^2\, n(M)\, {\rm d}M = \\ &= f_{\rm s}^2(L_1, L_2) \int \frac{\langle N_{\rm s}\rangle_M^2(L_1, L_2)}{\bar{n}_{\rm s}^2(L_1, L_2)} |u_{\rm s}(k|M)|^2\, n(M)\, {\rm d}M = \\ &= f_{\rm s}^2(L_1, L_2)\, \check{P}_{\rm gg}^{\rm 1h,s-s}(k, L_1, L_2) \end{aligned} \tag{4.23}$$

4.1.3 The two halo central-central term

The two halo term of the galaxy correlation function encodes the information about the clustering of the galaxies belonging to different dark matter haloes. See fig. 4.3 for a visualization of the galaxy pairs involved in this term. One can describe the clustering of galaxies as a biased dark matter clustering. This "bias" has to take into account the average number of galaxies in haloes of different masses, as well as their distribution inside the halo. As it will be clear from what follows, also for the two halo term (large scales) the halo occupation statistics is a crucial ingredient to define the 2-point galaxy correlation function.

In the case of the two halo central-central term, one has to correlate central galaxies living in a halo of mass M_1 with central galaxies living in a halo of mass M_2. Roughly speaking, it is expected that the correlation function of central galaxies living in separate haloes is tightly related to the correlation function of the haloes. Basically one can write

$$\begin{aligned}
\check{\xi}_{\rm gg}^{\rm 2h,c-c}(r, L_1, L_2) &= \frac{1}{\bar{n}_{\rm c}^2} \sum_{M_1} \sum_{M_2} \xi_{\rm hh}(r|M_1, M_2) = \\
&= \frac{1}{\bar{n}_{\rm c}^2} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2) n(M_1) {\rm d}M_1 \int \langle N_{\rm c}\rangle_{M_2}(L_1, L_2) n(M_2) {\rm d}M_2 \, \xi_{\rm hh}(r|M_1, M_2) = \\
&= \frac{\xi_{\rm dm}(r)}{\bar{n}_{\rm c}^2} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2) \, b(M_1) n(M_1) {\rm d}M_1 \int \langle N_{\rm c}\rangle_{M_2}(L_1, L_2) \, b(M_2) n(M_2) {\rm d}M_2 = \\
&= \frac{\xi_{\rm dm}(r)}{\bar{n}_{\rm c}^2} \left[\int \langle N_{\rm c}\rangle_{M}(L_1, L_2) \, b(M) n(M) {\rm d}M \right]^2 , \qquad (4.24)
\end{aligned}$$

where we have used the fact that the halo bias is assumed to be scale independent:

$$\xi_{\rm hh}(r|M_1, M_2) = b(M_1) \, b(M_2) \, \xi_{\rm dm}(r) \, . \qquad (4.25)$$

Note also that we have considered the probability that, on average, a halo of mass M hosts a central galaxy in the considered luminosity bin by using the halo occupation statistics $\langle N_{\rm c}\rangle_M(L_1, L_2)$. Ultimately, we are interested in the total galaxy correlation function, $\xi_{\rm gg}(r, L_1, L_2)$. As explained at the beginning of this chapter, the relative contribution due to central galaxies has to take into account the fraction of these galaxies, that is

$$\begin{aligned}
\tilde{\xi}_{\rm gg}^{\rm 2h,c-c}(r, L_1, L_2) &= f_{\rm c}^2(L_1, L_2) \, \check{\xi}_{\rm gg}^{\rm 2h,c-c}(r, L_1, L_2) = \\
&= \frac{\xi_{\rm dm}(r)}{\bar{n}_{\rm tot}^2} \left[\int \langle N_{\rm c}\rangle_{M}(L_1, L_2) \, b(M) n(M) {\rm d}M \right]^2 . \qquad (4.26)
\end{aligned}$$

The corresponding Fourier transform simply reads

$$\begin{aligned}
\tilde{P}_{\rm gg}^{\rm 2h,c-c}(k, L_1, L_2) &= \frac{P_{\rm dm}(k)}{\bar{n}_{\rm tot}^2(L_1, L_2)} \left[\int \langle N_{\rm c}\rangle_{M}(L_1, L_2) \, b(M) n(M) {\rm d}M \right]^2 = \\
&= f_{\rm c}^2(L_1, L_2) \, \frac{P_{\rm dm}(k)}{\bar{n}_{\rm c}^2(L_1, L_2)} \left[\int \langle N_{\rm c}\rangle_{M}(L_1, L_2) \, b(M) n(M) {\rm d}M \right]^2 = \\
&= f_{\rm c}^2(L_1, L_2) \, \check{P}_{\rm gg}^{\rm 2h,c-c}(k, L_1, L_2) \, . \qquad (4.27)
\end{aligned}$$

4.1.4 The two halo central-satellite term

For the two halo central-satellite term, one has to correlate central galaxies with luminosity L within $[L_1, L_2]$ residing in a halo if mass M_1 and satellite galaxies with the same luminosity residing in an other halo of mass M_2 (see fig. 4.3 for a schematic representation).

Following the arguments presented in the previous section, we have

$$\begin{aligned}\check{\xi}^{\rm 2h,c-s}_{\rm gg}(r), L_1, L_2 &= \frac{1}{\bar{n}_{\rm c}(L_1,L_2)\bar{n}_{\rm s}(L_1,L_2)}\sum_{M_1}\sum_{M_2}\langle N_{\rm c}\rangle_{M_1}(L_1,L_2)\,\langle N_{\rm s}\rangle_{M_2}(L_1,L_2)\\ &\int {\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}+\mathbf{x}'-\mathbf{x}|M_2)\,\xi_{\rm hh}(\mathbf{y},M_1,M_2) =\\ &= \frac{1}{\bar{n}_{\rm c}(L_1,L_2)\bar{n}_{\rm s}(L_1,L_2)}\int\langle N_{\rm c}\rangle_{M_1}(L_1,L_2)\,n(M_1){\rm d}M_1\int\langle N_{\rm s}\rangle_{M_2}(L_1,L_2)n(M_2){\rm d}M_2\\ &\int {\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}+\mathbf{x}'-\mathbf{x}|M_2)\,\xi_{\rm hh}(\mathbf{y},M_1,M_2)\,, \end{aligned} \tag{4.28}$$

where $r=|\mathbf{x}'-\mathbf{x}|$ and we have indicated the distance between dark matter haloes of mass M_1 and M_2 with $\mathbf{y}=\mathbf{x_1}-\mathbf{x_2}$. Moreover, we have made explicit use of the fact that the central galaxy resides at the center of the dark matter halo of mass M_1. Note that the expression in eq. (4.28) depends on the convolution of the radial distribution of the satellite galaxies with the correlation function of haloes. We can rewrite eq. (4.28) as

$$\begin{aligned}\tilde{\xi}^{\rm 2h,c-s}_{\rm gg}(r,L_1,L_2) &= f_{\rm c}(L_1,L_2)\,f_{\rm s}(L_1,L_2)\,\check{\xi}^{\rm 2h,c-c}_{\rm gg}(r,L_1,L_2) =\\ &= \frac{1}{\bar{n}^2_{\rm tot}(L_1,L_2)}\int\langle N_{\rm c}\rangle_{M_1}(L_1,L_2)\,n(M_1){\rm d}M_1\int\langle N_{\rm s}\rangle_{M_2}(L_1,L_2)n(M_2){\rm d}M_2\\ &\int {\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}+\mathbf{x}'-\mathbf{x}|M_2)\,\xi_{\rm hh}(\mathbf{y},M_1,M_2)\,. \end{aligned} \tag{4.29}$$

The Fourier counterpart is

$$\begin{aligned}\tilde{P}^{\rm 2h,c-s}_{\rm gg}(k,L_1,L_2) &= \int\langle N_{\rm c}\rangle_{M_1}(L_1,L_2)\,n(M_1){\rm d}M_1\int\langle N_{\rm s}\rangle_{M_2}(L_1,L_2)\,u_{\rm s}(k|M_2)\,n(M_2)\\ &P_{\rm hh}(k,M_1,M_2)\,{\rm d}M_2 =\\ &= \frac{P_{\rm dm}(k)}{\bar{n}^2_{\rm tot}(L_1,L_2)}\int\langle N_{\rm c}\rangle_{M_1}(L_1,L_2)\,b(M_1)\,n(M_1){\rm d}M_1\\ &\int\langle N_{\rm s}\rangle_{M_2}(L_1,L_2)\,u_{\rm s}(k|M_2)\,b(M_2)\,n(M_2){\rm d}M_2 =\\ &= f_{\rm c}(L_1,L_2)\,f_{\rm s}(L_1,L_2)\,\check{P}^{\rm 2h,c-s}_{\rm gg}(k,L_1,L_2) \end{aligned} \tag{4.30}$$

where the first equality holds because a convolution becomes a simple multiplication in Fourier space. Moreover, the second equality is based on the fact that, if the halo bias is scale-independent, the halo-halo power spectrum is

$$P_{\rm hh}(k,M_1,M_2)=b(M_1)b(M_2)P_{\rm dm}(k)\,. \tag{4.31}$$

4.1.5 The two halo satellite-satellite term

For the two halo satellite-satellite term, one has to correlate satellite galaxies with luminosity L within $[L_1,L_2]$ residing in a halo of mass M_1 and satellite galaxies with the same luminosity residing in an other halo of mass M_2 (see fig. 4.3 for a schematic representation).

Following the arguments presented in the previous sections, we have

$$\begin{aligned}\tilde{\xi}_{\rm gg}(r, L_1, L_2) &= f_{\rm s}^2(L_1, L_2)\,\check{\xi}_{\rm gg}(r, L_1, L_2) = \\ &= \frac{1}{\bar{n}^2_{\rm tot}(L_1, L_2)} \int n(M_1)\,\langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\,{\rm d}M_1 \int n(M_2)\,\langle N_{\rm s}\rangle_{M_2}(L_1, L_2)\,{\rm d}M_2 \\ &\quad \int {\rm d}^3\mathbf{x_1}\,\hat{n}_{\rm s}(\mathbf{x}-\mathbf{x_1}|M_1) \int {\rm d}^3\mathbf{x_2}\,\hat{n}_{\rm s}(\mathbf{x}'-\mathbf{x_2}|M_2)\,\xi_{\rm hh}(\mathbf{x_1}-\mathbf{x_2}, M_1, M_2)\,, \end{aligned} \tag{4.32}$$

where, as usual, we have written the sum over all haloes as an integral over halo masses weighted with the halo mass function. Note that the presence of a convolution of the satellite distribution with the halo-halo correlation function. The Fourier counterpart is simpler to calculate due to the fact that convolutions simply become multiplications in Fourier space. The two halo satellite-satellite term of the galaxy power spectrum then reads

$$\begin{aligned}\tilde{P}_{\rm gg}(k, L_1, L_2) &= \frac{1}{\bar{n}^2_{\rm tot}(L_1, L_2)} \int n(M_1)\,\langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\,u_{\rm s}(k|M_1)\,{\rm d}M_1 \\ &\quad \int n(M_2)\,\langle N_{\rm s}\rangle_{M_2}(L_1, L_2)\,u_{\rm s}(k|M_2)\,P_{\rm hh}(k, M_1, M_2){\rm d}M_2 = \\ &= \frac{P_{\rm dm}(k)}{\bar{n}^2_{\rm tot}(L_1, L_2)} \int \langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\,u_{\rm s}(k|M_1)\,b(M_1)\,n(M_1)\,{\rm d}M_1 \\ &\quad \int \langle N_{\rm s}\rangle_{M_2}(L_1, L_2)\,u_{\rm s}(k|M_2)\,b(M_2)\,n(M_2)\,{\rm d}M_2 = \\ &= \frac{P_{\rm dm}(k)}{\bar{n}^2_{\rm tot}(L_1, L_2)} \left[\int \langle N_{\rm s}\rangle_{M}(L_1, L_2)\,u_{\rm s}(k|M)\,b(M)\,n(M)\,{\rm d}M\right]^2 = \\ &= f_{\rm s}^2(L_1, L_2)\,\check{P}^{\rm 2h,s-s}_{\rm gg}(k, L_1, L_2) \end{aligned} \tag{4.33}$$

where we have used the usual approximation for the halo-halo power spectrum

$$P_{\rm hh}(k, M_1, M_2) = b(M_1)b(M_2)P_{\rm dm}(k)\,, \tag{4.34}$$

and we have substituted the variables M_1 and M_2 with M.

4.1.6 The Projected Galaxy Correlation Function

We have shown in chapter 2 that it is convenient to measure the galaxy clustering properties using the 2-point projected galaxy correlation function, w_p. From the theoretical point of view, this quantity can be straightforwardly calculated once the spatial 2-point correlation function, $\xi_{\rm gg}$, is known. In fact we have

$$w_p(R) = \int_R^\infty \xi_{\rm gg}(r)\frac{r\,{\rm d}r}{(r^2-R^2)^{1/2}}\,, \tag{4.35}$$

where R is the comoving separation in the sky (2-D distance) and r is the comoving spatial (3-D) separation. Here, we have implicitly made use of the fact that $\xi_{\rm gg}(r)$ goes to zero in the limit $r \to \infty$

and in practice the angle under which R is observed is small. Note that eq. (4.35) can be rewritten as

$$w_p(R) = 2\,R \int_0^1 \xi_{\rm gg}(R/y) \frac{{\rm d}y}{y^2(1-y^2)^{1/2}}\,. \tag{4.36}$$

4.1.7 The Galaxy Correlation Length

At very large scales, the two halo term of the galaxy correlation function can be further simplified. At distances r much larger than the typical extension r_{180} of dark matter haloes, the location of the galaxies inside the halo is not relevant, $r + r_{180} \sim r$. At these scales thus

$$\begin{aligned} \xi_{\rm gg}(r|L_1,L_2) &\sim \frac{\xi_{\rm dm}(r)}{\bar{n}^2_{\rm tot}(L_1,L_2)} \left[\int \langle N_{\rm tot}\rangle_M(L_1,L_2)\, b(M)\, n(M)\, {\rm d}M\right]^2 = \\ &= b^2_{\rm g}(L_1,L_2)\, \xi_{\rm dm}(r) \end{aligned} \tag{4.37}$$

with

$$b^2_{\rm g}(L_1,L_2) = \left[\int \frac{\langle N_{\rm tot}\rangle_M(L_1,L_2)}{\bar{n}_{\rm tot}(L_1,L_2)}\, b(M)\, n(M)\, {\rm d}M\right]^2\,, \tag{4.38}$$

where $\langle N_{\rm tot}\rangle_M(L_1,L_2) = \langle N_{\rm c}\rangle_M(L_1,L_2) + \langle N_{\rm s}\rangle_M(L_1,L_2)$ comes from the fact that all galaxies are taken into account without the split into central and satellite. In the literature (see e.g. Tinker et al.2005), a correction to this expression has been suggested to take into account the scale dependence of the bias. This correction changes eq. (4.37) in

$$\xi_{\rm gg}(r|L_1,L_2) = b^2_{\rm g}(L_1,L_2)\, \zeta(r)\, \xi_{\rm dm}(r)\,, \tag{4.39}$$

with

$$\zeta(r) = \frac{[1 + 1.17\xi_{\rm dm}(r)]^{1.49}}{[1 + 0.69\xi_{\rm dm}(r)]^{2.09}}\,. \tag{4.40}$$

Eq. (4.39) links the large scale galaxy clustering properties to the underlying dark matter clustering. Since the observed galaxy correlation function is fairly well described by a power law,

$$\xi_{\rm gg} = \left(\frac{r}{r_0(L_1,L_2)}\right)^{\alpha(L_1,L_2)}\,, \tag{4.41}$$

one can use the correlation length, $r_0(L_1,L_2)$ to quantify the galaxy clustering strength as a function of the luminosity. Note that the correlation length is defined such that $\xi_{\rm gg}[r_0(L_1,L_2)] = 1$.

As already illustrated in chapter 2, the correlation length is larger for brighter galaxies. This means that brighter galaxies are, on average, more strongly clustered than fainter galaxies. Since numerical simulations indicates that more massive haloes are more strongly clustered than less massive ones, this suggests that bright galaxies live, on average, in massive haloes.

4.2 Dark Matter Clustering

The aim of this section is to introduce the treatment of the dark matter clustering within the framework of the halo model. In chapter 3, we have already introduced the statistical tools to investigate the properties of the dark matter density field. In particular, we have described the clustering of dark matter via the correlation function and the power spectrum. A basic assumption in that treatment was that the density of dark matter can be described as a continuous field, $\rho_{\rm dm}(\mathbf{x})$. However, in the halo model, we assume that the dark matter distribution can be partitioned up into haloes of different masses. This means that we treat the dark matter density field as a discrete field. In such a discretized model, one can describe the clustering of dark matter as the clustering of objects (in a similar fashion as for the galaxy clustering). It is worth stressing that, in the limit in which also extremely small halo masses can be considered, the discretized description approaches the continuous treatment. Therefore, there is no loss of generality in using the halo model approach to describe the properties of the dark matter density distribution. As in the case of the galaxy clustering, we will use the 2-point correlation function (and the corresponding power spectrum) for quantifying the clustering properties.

Given the above discussion, it is expected that the dark matter distribution will be divided into two parts: the one halo and the two halo regime. Accordingly, the one halo term describes the distribution of dark matter inside haloes of a given mass, whereas the two term relies on the knowledge of the spatial distribution of haloes. Note, however, two obvious difference between galaxies and dark matter distribution: (i) the dark matter does not have the split into central and satellite term, (ii) the dark matter properties do not depend on luminosity.

Following the same line of reasoning used for the galaxy correlation function, we have

$$1 + \xi_{\rm dm}(r) = \frac{DD(r)}{DR(r)}\,, \tag{4.42}$$

where $DD(r)$ is the number of dark matter pairs as *"observed"* in a given realization of dark matter distribution, whereas $DR(r)$ is the number of dark matter pairs in a random distribution. Adopting the ergodic principle (see chapter 3), we write $DR(r) = \bar{\rho}^2$. Dividing the number of pairs into one and two halo, we have

$$1 + \xi_{\rm dm}(r) = \frac{DD^{\rm 1h}(r) + DD^{\rm 2h}(r)}{\bar{\rho}^2} = \tilde{\xi}^{\rm 1h}_{\rm dm}(r) + \tilde{\xi}^{\rm 2h}_{\rm dm}(r)\,. \tag{4.43}$$

The Fourier counterpart, the power spectrum, can be written as

$$P_{\rm dm}(k) = \tilde{P}^{\rm 1h}_{\rm dm}(k) + \tilde{P}^{\rm 2h}_{\rm dm}(k)\,. \tag{4.44}$$

4.2.1 The one halo term

We write the number of dark matter pairs as a sum over all haloes of the number of pairs per halo:

$$DD^{\rm 1h}(r) = \sum_{\rm all\,haloes} DD^{\rm 1h}(r|M)\,. \tag{4.45}$$

If the dark matter inside a halo follows a density distribution, $\rho_{\rm dm}(\mathbf{r}|M)$, then it is clear that $DD^{\rm 1h}(r|M)$ consists of a convolution of the density profile with itself

$$DD^{\rm 1h}(r|M) = \langle \rho_{\rm dm}(\mathbf{y}|M)\, \rho_{\rm dm}(\mathbf{y}+\mathbf{r}|M)\rangle\,. \tag{4.46}$$

By writing the density profile as

$$\rho_{\rm dm}(\mathbf{y}|M) = M\hat{\rho}_{\rm dm}(\mathbf{y})\,, \tag{4.47}$$

we have

$$DD^{\rm 1h}(r|M) = M^2\,\langle \hat{\rho}_{\rm dm}(\mathbf{y}|M)\, \hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M)\rangle\,, \tag{4.48}$$

where, by definition,

$$4\pi \int_0^{r_{180}} {\rm d}r^2 \hat{\rho}_{\rm dm}(r|M) = 1\,. \tag{4.49}$$

Note that we have assumed that the dark matter density profile is actually radially symmetric.

By summing over all haloes, we integrate over all halo masses using the halo mass function $n(M)$:

$$\begin{aligned} DD^{\rm 1h}(r) &= \sum_{\rm all\,haloes} DD^{\rm 1h}(r|M) = \\ &= \int n(M)\, DD^{\rm 1h}(r|M)\, {\rm d}M = \\ &= \int M^2\,\langle \hat{\rho}_{\rm dm}(\mathbf{y}|M)\, \hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M)\rangle\, n(M)\, {\rm d}M = \\ &= \int M^2\, n(M)\, {\rm d}M \int {\rm d}^3\mathbf{y}\, \hat{\rho}_{\rm dm}(\mathbf{y}|M)\, \hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M)\,. \end{aligned} \tag{4.50}$$

The one halo term of the dark matter correlation then reads

$$\begin{aligned} \tilde{\xi}^{\rm 1h}_{\rm dm}(r) &= \frac{DD^{\rm 1h}(r)}{\bar{\rho}^2} = \\ &= \frac{1}{\bar{\rho}^2} \int M^2\, n(M)\, {\rm d}M \int_0^{r_{180}} {\rm d}^3\mathbf{y}\, \hat{\rho}_{\rm dm}(\mathbf{y}|M)\, \hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M)\,. \end{aligned} \tag{4.51}$$

The Fourier counterpart is then

$$\tilde{P}^{\rm 1h}_{\rm dm}(k) = \frac{1}{\bar{\rho}^2} \int M^2\, u^2_{\rm dm}(k|M)\, n(M)\, {\rm d}M\,. \tag{4.52}$$

Here, $u_{\rm dm}(k|M)$ is the Fourier transform of the normalized dark matter density profile within a halo of mass M

$$u_{\rm dm}(k|M) = 4\pi \int_0^{r_{180}} \hat{\rho}_{\rm dm}(r|M)\, \frac{\sin kr}{kr} r^2\, {\rm d}r\,. \tag{4.53}$$

For the NFW profile (see chapter 3),

$$\begin{aligned} u_{\rm dm}(k|M) &= \frac{4\pi\rho_* r_*^3}{M} \Bigg\{ \sin(kr_*) \Big[{\rm Si}([1+c_{\rm dm}]kr_*) - {\rm Si}(kr_*) \Big] - \frac{\sin(c_{\rm dm} kr_*)}{(1+c_{\rm dm})kr_*} \\ &\qquad + \cos(kr_*) \Big[{\rm Ci}([1+c_{\rm dm}]kr_*) - {\rm Ci}(kr_*) \Big] \Bigg\}\,, \end{aligned} \tag{4.54}$$

where the sine and cosine integrals are

$$\mathrm{Ci}(x) = -\int_x^\infty \frac{\cos t}{t}\, dt \quad \text{and} \quad \mathrm{Si}(x) = \int_0^x \frac{\sin t}{t}\, dt\,. \tag{4.55}$$

Figure 4.4 shows $u_{\rm dm}(k|M)$ as a function of k for different values of M (assuming NFW profiles). In general, the shape of $u_{\rm dm}(k|M)$ depends both on the halo concentration parameter, $c_{\rm dm}$, and the mass M. The figure shows an important feature: the small scale power (large k) is dominated by low mass haloes. This will be relevant for performing the actual calculation of the power spectra (see §4.4)

4.2.2 The two halo term

The expression that describes the two halo term of the dark matter correlation function can be derived following the same line of reasoning valid for the one halo term. The only substantial difference is that the correlated dark matter particles must reside in separate haloes. This implies

$$\begin{aligned}\tilde{\xi}_{\rm dm}(r) &= \frac{1}{\bar{\rho}^2}\int n(M_1)\, M_1\, \mathrm{d}M_1 \int n(M_2)\, M_2\, \mathrm{d}M_2 \\ &\int \mathrm{d}^3\mathbf{x}_1\, \hat{\rho}_{\rm dm}(\mathbf{x}-\mathbf{x}_1|M_1) \int \mathrm{d}^3\mathbf{x}_2\, \hat{\rho}_{\rm dm}(\mathbf{x}'-\mathbf{x}_2|M_2)\, \xi_{\rm hh}(\mathbf{x}_1-\mathbf{x}_2|M_1,M_2)\,,\end{aligned} \tag{4.56}$$

The Fourier counterpart of the two halo term of the dark matter correlation function is

$$P^{\rm 2h}_{\rm dm}(k) = \frac{1}{\bar{\rho}^2}\int M_1\, u_{\rm dm}(k|M_1)\, n(M_1)\, \mathrm{d}M_1 \int M_2\, u_{\rm dm}(k|M_2)\, n(M_2)\, \mathrm{d}M_2\, P_{\rm hh}(k|M_1,M_2)\,. \tag{4.57}$$

Under the assumption that the bias between haloes and dark matter is scale independent (see chapter 3), we have that

$$P_{\rm hh}(k|M_1,M_2) = b(M_1)\, b(M_2)\, P_{\rm dm}(k)\,, \tag{4.58}$$

where $b(M)$ is the value of the halo bias function for a halo of mass M. Using eq. (4.58), one can take $P_{\rm dm}(k)$ outside of the integrals over M_1 and M_2, making the two integrals separable. This simplifies the expression of the 2-halo term of the power spectrum as follows:

$$P^{\rm 2h}_{\rm dm}(k) = \frac{P_{\rm dm}(k)}{\bar{\rho}^2}\left[\int M\, u_{\rm dm}(k|M)\, b(M)\, n(M)\, \mathrm{d}M\right]^2. \tag{4.59}$$

Note that we have used the fact that M_1 and M_2 can be substituted by M.

At a first look, eq. (4.59) seems to suggest that the halo model introduces a vicious loop. In fact, the idea is to be able to describe the power spectrum of the dark matter distribution but in order to do it one requires the knowledge of $P_{\rm dm}(k)$ (see eq. 4.59). This is only partially true. In fact, eq. (4.58)

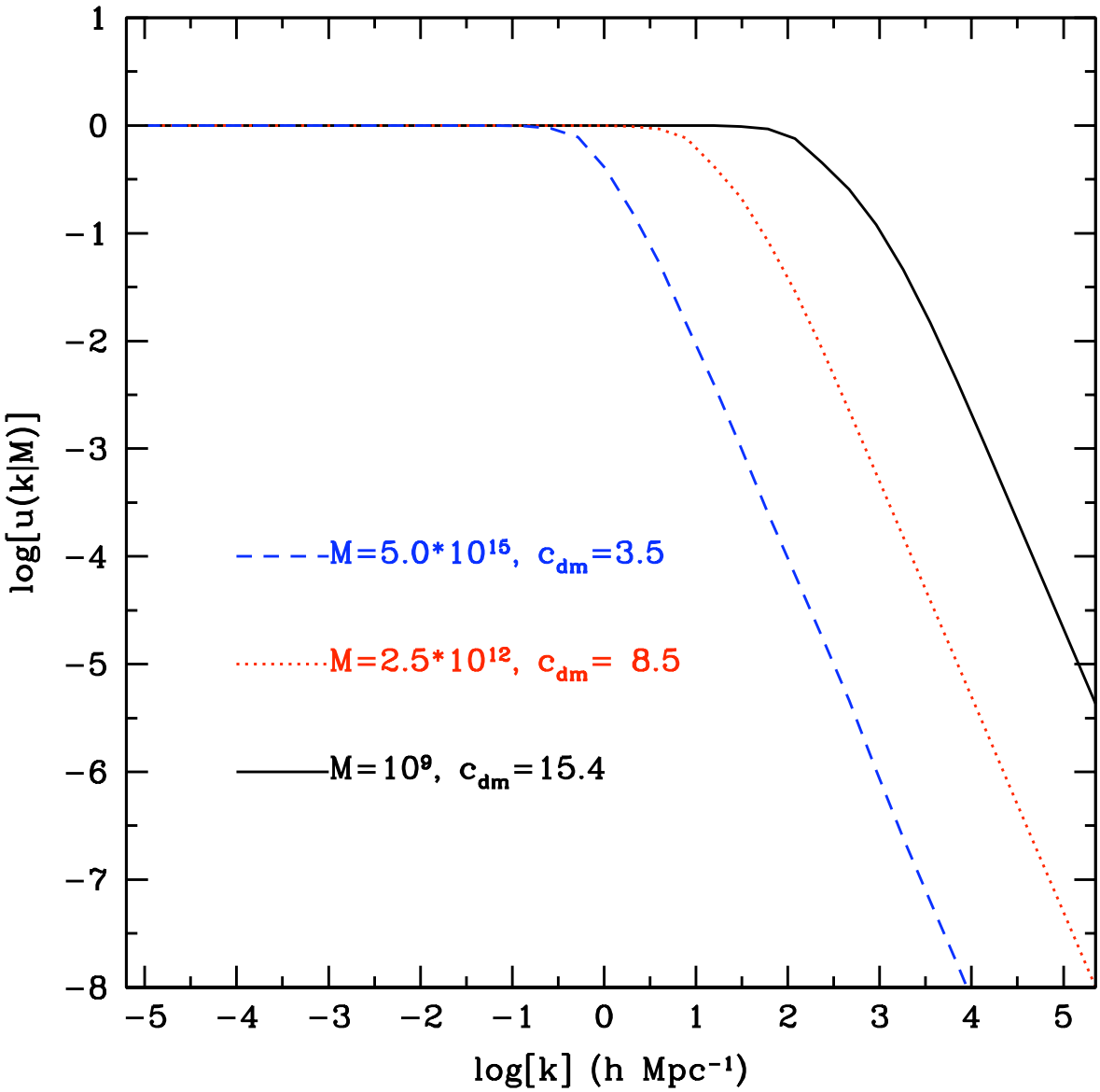

Figure 4.4: The Fourier transform of the dark matter density profile, $u_{\rm dm}(k|M)$ for different halo masses.

is true only in the regime where $b(k)$ is independent of k (the bias must be scale-independent). In this regime, $P_{\rm dm}(k) \sim P^{\rm lin}_{\rm dm}(k)$ which can be evaluated from linear theory (see chapter 3). However, by comparing the power spectrum provided by numerical simulations with the one derived via the halo model approach where $P^{\rm 2h}_{\rm dm}(k) \propto P^{\rm lin}_{\rm dm}(k)$, one can notice that the analytical approach underestimates the power spectrum on scales $0.2 \lesssim k \lesssim 1.2h{\rm Mpc}^{-1}$ (see e.g. Cooray & Sheth 2001). These are indeed the scales at which the scale-dependence of the bias starts to be relevant and eq. (4.58) fails. The only formally correct solution to this problem would be having an analytical estimation of the dark matter power spectrum also in the midly-nonlinear regime (a sort of *quasi-non linear* power spectrum). Although there have been attempts to find such a power spectrum (e.g. one-loop perturbation theory, see Cooray & Sheth 2001 and reference therein) a definitive answer has not yet been found. Some authors have tried to overcome this problem by using a scale-dependent bias, $b(M, r)$ (see e.g. Tinker et al. 2005). Although this attempt improves the quality of the agreement with numerical simulations, it is anyway approximated because it cannot account for the profile of the dark matter inside the haloes (see eq.[4.59]). In this thesis, unless otherwise specified, we don't describe the dark matter power spectrum in the halo model formalism but we use the result from numerical simulations any time that $P_{\rm dm}(k)$ is required. This section is thus mostly an introduction to the description of the dark matter correlation function and power spectrum in the halo model formalism. As explained in the next section, this will be useful to describe the clustering of dark matter around galaxies.

4.3 Dark Matter Clustering around Galaxies

To describe the clustering of dark matter around galaxies, we use the 2-point galaxy-dark matter cross correlation function, $\xi_{\rm g,dm}$. To define this quantity, we will follow the line of reasoning already applied to the clustering of dark matter and that of galaxies. We, therefore, have

$$1 + \xi_{\rm g,dm}(r, L_1, L_2) = \frac{DD(r, L_1L_2)}{DR(r, L_1, L_2)}\,, \tag{4.60}$$

where $DD(r, L_1L_2)$ is the number of galaxy-dark matter pairs in the *"data"*, whereas $DR(r, L_1, L_2)$ is the number of galaxy-dark matter pair in a random realization. Bearing in mind the arguments used for the 2-point correlation function of galaxies and dark matter and the distinction into central and satellite galaxies, it is straightforward to split the number of pairs into four terms:

$$DD(r, L_1L_2) = N^{\rm 1h,c}_{\rm pairs}(r, L_1, L_2) + N^{\rm 1h,s}_{\rm pairs}(r, L_1, L_2) + N^{\rm 2h,c}_{\rm pairs}(r, L_1, L_2) + N^{\rm 2h,s}_{\rm pairs}(r, L_1, L_2)\,. \tag{4.61}$$

See fig. 4.5 for an illustration of these pairs. Moreover, it follows from previous discussion that $DR(r, L_1, L_2) = \bar{n}_{\rm tot}(L_1, L_2)\,\bar{\rho}$.

Accordingly, the cross-correlation function $\xi_{\rm g,dm}(r, L_1, L_2)$ can be split into four terms

$$\begin{aligned}
\xi_{\rm dm}(r, L_1, L_2) &= \tilde{\xi}^{\rm 1h,c}_{\rm dm}(r, L_1, L_2) + \tilde{\xi}^{\rm 1h,s}_{\rm dm}(r, L_1, L_2) + \tilde{\xi}^{\rm 2h,c}_{\rm dm}(r, L_1, L_2) + \tilde{\xi}^{\rm 2h,s}_{\rm dm}(r, L_1, L_2) = \\
&= f_{\rm c}(L_1, L_2)\,\check{\xi}^{\rm 1h,c}_{\rm dm}(r, L_1, L_2) + f_{\rm s}(L_1, L_2)\,\check{\xi}^{\rm 1h,s}_{\rm dm}(r, L_1, L_2) + \\
&+ f_{\rm c}(L_1, L_2)\,\check{\xi}^{\rm 2h,c}_{\rm dm}(r, L_1, L_2) + f_{\rm s}(L_1, L_2)\,\check{\xi}^{\rm 2h,s}_{\rm dm}(r, L_1, L_2)
\end{aligned} \tag{4.62}$$

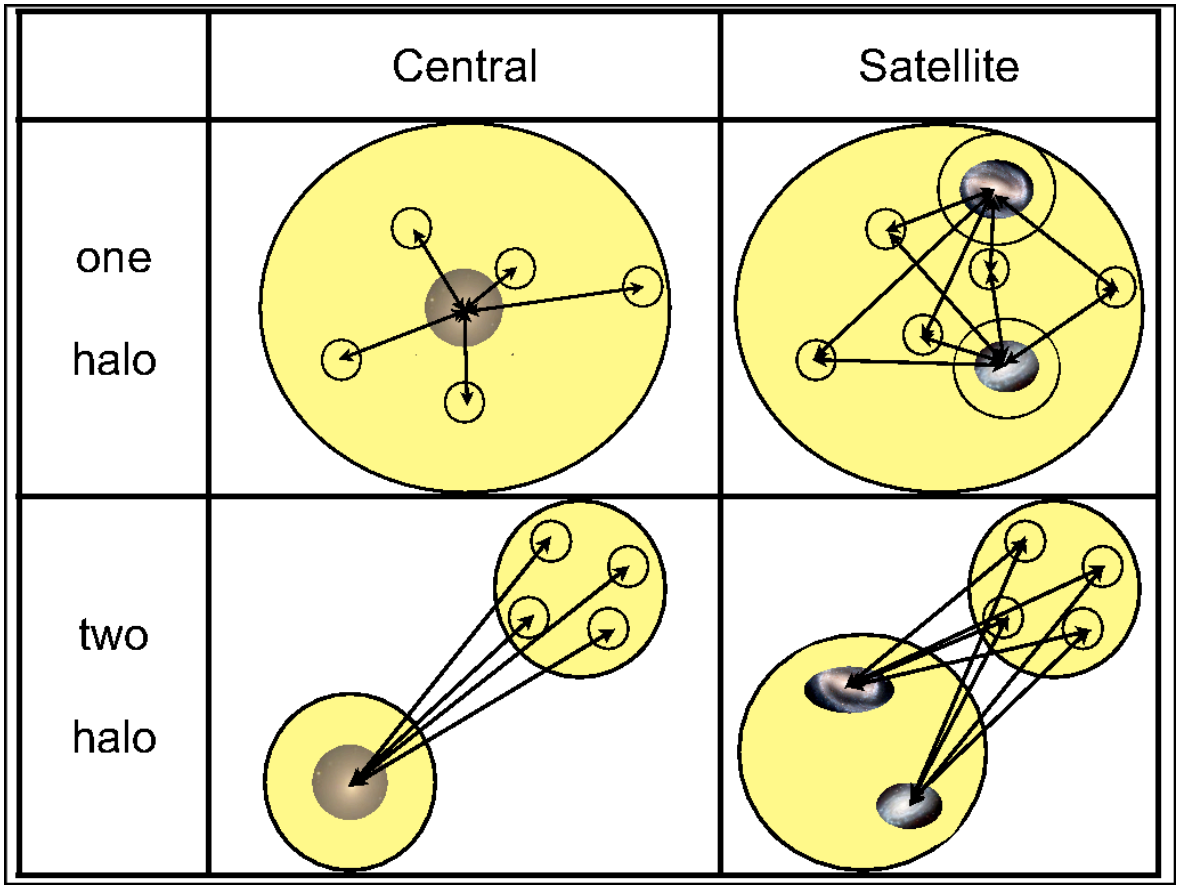

Figure 4.5: Schematic representation of the different terms of the galaxy-dark matter cross correlation function.

and we can define the corresponding power spectra as

$$\begin{aligned} P_{\rm dm}(k,L_1,L_2) &= \tilde{P}^{\rm 1h,c}_{\rm dm}(k,L_1,L_2) + \tilde{P}^{\rm 1h,s}_{\rm dm}(r,L_1,L_2) + \tilde{P}^{\rm 2h,c}_{\rm dm}(k,L_1,L_2) + \tilde{P}^{\rm 2h,s}_{\rm dm}(k,L_1,L_2) = \\ &= f_{\rm c}(L_1,L_2)\,\breve{P}^{\rm 1h,c}_{\rm dm}(k,L_1,L_2) + f_{\rm s}(L_1,L_2)\,\breve{P}^{\rm 1h,s}_{\rm dm}(r,L_1,L_2) + \\ &+ f_{\rm c}(L_1,L_2)\,\breve{P}^{\rm 2h,c}_{\rm dm}(k,L_1,L_2) + f_{\rm s}(L_1,L_2)\,\breve{P}^{\rm 2h,s}_{\rm dm}(k,L_1,L_2)\,. \end{aligned} \tag{4.63}$$

Note that the notation $\breve{\cdot}$ follows from the discussion in the previous sections. In what follows, we describe the four terms of the galaxy-dark matter cross correlation function (and corresponding power spectra) in detail.

4.3.1 The one halo central term

The one halo central term describes the correlation between central galaxies with luminosity in the range $[L_1, L_2]$ and the dark matter in the halo hosting these galaxies. As usual, we write the number

of pairs as sum over the pairs per halo

$$N^{\mathrm{1h,c}}_{\mathrm{pairs}}(r, L_1, L_2) = \sum_{\mathrm{all\,haloes}} N^{\mathrm{1h}}_{\mathrm{pairs}}(r, L_1, L_2|M)\,. \tag{4.64}$$

Since, by definition, the central galaxy sits at the centre of the dark matter halo and the dark matter distribution follows a radial profile, $M\hat{\rho}_{\mathrm{dm}}(r|M)$, we have

$$N^{\mathrm{1h,c}}_{\mathrm{pairs}}(r, L_1, L_2|M) = \langle N_{\mathrm{c}}\rangle_M(L_1, L_2) M\hat{\rho}_{\mathrm{dm}}(r|M)\,, \tag{4.65}$$

where the term $\langle N_{\mathrm{c}}\rangle_M(L_1, L_2)$ describes the probability that a halo of mass M contains, on average, a central galaxy with luminosity in the range $[L_1, L_2]$. By summing over all haloes, we obtain

$$\begin{aligned} N^{\mathrm{1h,c}}_{\mathrm{pairs}}(r, L_1, L_2) &= \sum_{\mathrm{all\,haloes}} N^{\mathrm{1h,c}}_{\mathrm{pairs}}(r, L_1, L_2|M) = \\ &= \int n(M)\,\langle N_{\mathrm{c}}\rangle_M(L_1, L_2)\, M\,\hat{\rho}_{\mathrm{dm}}(r|M)\mathrm{d}M\,. \end{aligned} \tag{4.66}$$

The one halo central term of the galaxy-dark matter cross correlation then reads

$$\begin{aligned} \tilde{\xi}^{\mathrm{1h,c}}_{\mathrm{g,dm}}(r, L_1, L_2) &= \sum_{\mathrm{all\,haloes}} \frac{N^{\mathrm{1h,c}}_{\mathrm{pairs}}(r, L_1, L_2)}{\bar{\rho}\bar{n}_{\mathrm{tot}}(L_1, L_2)} \\ &= \frac{1}{\bar{\rho}\bar{n}_{\mathrm{tot}}(L_1, L_2)} \int n(M)\,\langle N_{\mathrm{c}}\rangle_M(L_1, L_2)\, M\,\hat{\rho}_{\mathrm{dm}}(r|M)\mathrm{d}M\; = \\ &= f_{\mathrm{c}}(L_1, L_2)\,\frac{1}{\bar{\rho}\bar{n}_{\mathrm{c}}(L_1, L_2)} \int n(M)\,\langle N_{\mathrm{c}}\rangle_M(L_1, L_2)\, M\,\hat{\rho}_{\mathrm{dm}}(r|M)\mathrm{d}M\; = \\ &= f_{\mathrm{c}}(L_1, L_2)\,\check{\xi}^{\mathrm{1h,c}}_{\mathrm{g,dm}}(r, L_1, L_2) \end{aligned} \tag{4.67}$$

The corresponding power spectrum then reads

$$\begin{aligned} \tilde{P}^{\mathrm{1h,c}}_{\mathrm{g,dm}}(k, L_1, L_2) &= \frac{1}{\bar{\rho}\bar{n}_{\mathrm{tot}}(L_1, L_2)} \int n(M)\,\langle N_{\mathrm{c}}\rangle_M(L_1, L_2)\, M\, u_{\mathrm{dm}}(k|M)\mathrm{d}M\; = \\ &= f_{\mathrm{c}}(L_1, L_2)\,\frac{1}{\bar{\rho}\bar{n}_{\mathrm{c}}(L_1, L_2)} \int n(M)\,\langle N_{\mathrm{c}}\rangle_M(L_1, L_2)\, M\, u_{\mathrm{dm}}(k|M)\mathrm{d}M \\ &= f_{\mathrm{c}}(L_1, L_2)\,\check{P}^{\mathrm{1h,c}}_{\mathrm{g,dm}}(k, L_1, L_2) \end{aligned} \tag{4.68}$$

4.3.2 The one halo satellite term

Similarly to the one halo central term, we can define satellite galaxy-dark matter pairs as a sum over haloes of the satellite-dark matter pairs per halo

$$N^{\mathrm{1h,s}}_{\mathrm{pairs}}(r, L_1, L_2) = \sum_{\mathrm{all\,haloes}} N^{\mathrm{1h,s}}_{\mathrm{pairs}}(r, L_1, L_2|M)\,. \tag{4.69}$$

Since, by definition, satellite galaxies are distributed according to a radial profile, $\langle N_{\rm s}\rangle_M(L_1,L_2)\hat{n}_{\rm s}(r|M)$, and the dark matter distribution follows a radial profile, $M\hat{\rho}_{\rm dm}(r|M)$, we have

$$N^{\rm 1h,s}_{\rm pairs}(r,L_1,L_2|M) = M\,\langle N_{\rm s}\rangle_M(L_1,L_2)\int_{V_M}{\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}|M)\,\hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M)\,. \tag{4.70}$$

Summing over all the haloes and dividing by $\bar{\rho}\bar{n}_{\rm tot}(L_1,L_2)$ we get the expression for the one halo satellite term of the galaxy-dark matter correlation function:

$$\begin{aligned}
\tilde{\xi}^{\rm 1h,s}_{\rm g,dm}(r,L_1,L_2) &= \frac{N^{\rm 1h,s}_{\rm pairs}(r,L_1,L_2)}{\bar{\rho}\bar{n}_{\rm tot}(L_1,L_2)} = \\
&= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1,L_2)}\sum_{\rm all\ haloes} N^{\rm 1h,s}_{\rm pairs}(r,L_1,L_2|M) = \\
&= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1,L_2)}\sum_{\rm all\ haloes} M\,\langle N_{\rm s}\rangle_M(L_1,L_2)\int_{V_M}{\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}|M)\,\hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M) = \\
&= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1,L_2)}\int M\,\langle N_{\rm s}\rangle_M(L_1,L_2)\,n(M){\rm d}M\int_{V_M}{\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}|M)\,\hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M) = \\
&= f_{\rm s}(L_1,L_2)\,\frac{1}{\bar{\rho}\bar{n}_{\rm s}(L_1,L_2)}\int M\,\langle N_{\rm s}\rangle_M(L_1,L_2)\,n(M){\rm d}M\int_{V_M}{\rm d}^3\mathbf{y}\,\hat{n}_{\rm s}(\mathbf{y}|M)\,\hat{\rho}_{\rm dm}(\mathbf{y}+\mathbf{r}|M) = \\
&= f_{\rm s}(L_1,L_2)\,\check{\xi}^{\rm 1h,s}_{\rm g,dm}(r,L_1,L_2)\,.
\end{aligned} \tag{4.71}$$

Note that this expression involves a convolution of the distribution of galaxies with the dark matter density profile. However, the Fourier counterpart simplifies notably:

$$\begin{aligned}
\tilde{P}^{\rm 1h,s}_{\rm g,dm}(k,L_1,L_2) &= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1,L_2)}\int M\,\langle N_{\rm s}\rangle_M(L_1,L_2)\,n(M){\rm d}M\,u_{\rm s}(k|M)\,u_{\rm dm}(k|M) = \\
&= f_{\rm s}(L_1,L_2)\,\frac{1}{\bar{\rho}\bar{n}_{\rm s}(L_1,L_2)}\int M\,\langle N_{\rm s}\rangle_M(L_1,L_2)\,n(M){\rm d}M\,u_{\rm s}(k|M)\,u_{\rm dm}(k|M) = \\
&= f_{\rm s}(L_1,L_2)\check{P}^{\rm 1h,s}_{\rm g,dm}(k,L_1,L_2)
\end{aligned} \tag{4.72}$$

4.3.3 The two halo central term

Since central galaxies reside at the centre of the dark matter haloes, dark matter is distributed with a density profile, $M\hat{\rho}(\mathbf{x}|M)$, and the spatial distribution of haloes can be described via the halo-halo correlation function, the expression for the two halo central term of the galaxy-dark matter cross

correlation can be written as

$$\begin{aligned}
\tilde{\xi}^{\rm 2h,c}_{\rm g,dm}(r, L_1, L_2) &= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1, L_2)} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2)\, n(M_1)\, {\rm d}M_1 \int M_2\, n(M_2)\, {\rm d}M_2 \\
&\quad \int {\rm d}^3\mathbf{y}\, \hat{\rho}_{\rm dm}(\mathbf{y} + \mathbf{x}' - \mathbf{x}|M_2)\, \xi_{\rm hh}(\mathbf{y}, M_1, M_2) \; = \\
&= f_{\rm c}(L_1, L_2)\, \frac{1}{\bar{\rho}\bar{n}_{\rm c}(L_1, L_2)} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2)\, n(M_1)\, {\rm d}M_1 \int M_2\, n(M_2)\, {\rm d}M_2 \\
&\quad \int {\rm d}^3\mathbf{y}\, \hat{\rho}_{\rm dm}(\mathbf{y} + \mathbf{x}' - \mathbf{x}|M_2)\, \xi_{\rm hh}(\mathbf{y}, M_1, M_2) \; = \\
&= f_{\rm c}(L_1, L_2)\, \check{\xi}^{\rm 2h,c}_{\rm g,dm}(r, L_1, L_2)
\end{aligned} \tag{4.73}$$

The Fourier counterparts reads

$$\begin{aligned}
\tilde{P}^{\rm 2h,c}_{\rm g,dm}(k, L_1, L_2) &= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1, L_2)} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2)\, n(M_1)\, {\rm d}M_1 \\
&\quad \int M_2\, u_{\rm dm}(k|M_2)\, n(M_2)\, P_{\rm hh}(k|M_1, M_2)\, {\rm d}M_2 \; = \\
&= \frac{P_{\rm dm}(k)}{\bar{\rho}\bar{n}_{\rm tot}(L_1, L_2)} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2)\, b(M_1)n(M_1)\, {\rm d}M_1 \\
&\quad \int M_2\, u_{\rm dm}(k|M_2)\, b(M_2)n(M_2)\, {\rm d}M_2 \; = \\
&= f_{\rm c}(L_1, L_2)\, \frac{P_{\rm dm}(k)}{\bar{\rho}\bar{n}_{\rm c}(L_1, L_2)} \int \langle N_{\rm c}\rangle_{M_1}(L_1, L_2)\, b(M_1)n(M_1)\, {\rm d}M_1 \\
&\quad \int M_2\, u_{\rm dm}(k|M_2)\, b(M_2)n(M_2)\, {\rm d}M_2 \; = \\
&= f_{\rm c}(L_1, L_2)\, \check{P}^{\rm 2h,c}_{\rm g,dm}(k, L_1, L_2)
\end{aligned} \tag{4.74}$$

where we have used the assumption that the halo bias is scale independent, i.e.

$$P_{\rm hh}(k|M_1, M_2) = b(M_1)b(M_2)P_{\rm dm}(k)\,. \tag{4.75}$$

4.3.4 The two halo satellite term

Following the same line of reasoning of the previous sections, we can write the two halo satellite term of the galaxy-dark matter cross correlation function as

$$\begin{aligned}
\tilde{\xi}^{\rm 2h,s}_{\rm gdm}(r, L_1, L_2) &= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1, L_2)} \int n(M_1)\, \langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\, {\rm d}M_1 \int n(M_2)\, M_2\, {\rm d}M_2 \\
&\quad \int {\rm d}^3\mathbf{x}_1\, \hat{n}_{\rm s}(\mathbf{x} - \mathbf{x}_1|M_1) \int {\rm d}^3\mathbf{x}_2\, \hat{\rho}_{\rm dm}(\mathbf{x}' - \mathbf{x}_2|M_2)\, \xi_{\rm hh}(\mathbf{x_1} - \mathbf{x_2}, M_1, M_2) \; = \\
&= f_{\rm s}(L_1, L_2)\, \frac{1}{\bar{\rho}\bar{n}_{\rm s}(L_1, L_2)} \int n(M_1)\, \langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\, {\rm d}M_1 \int n(M_2)\, M_2\, {\rm d}M_2 \\
&\quad \int {\rm d}^3\mathbf{x}_1\, \hat{n}_{\rm s}(\mathbf{x} - \mathbf{x}_1|M_1) \int {\rm d}^3\mathbf{x}_2\, \hat{\rho}_{\rm dm}(\mathbf{x}' - \mathbf{x}_2|M_2)\, \xi_{\rm hh}(\mathbf{x_1} - \mathbf{x_2}, M_1, M_2) \; = \\
&= f_{\rm s}(L_1, L_2)\, \check{\xi}^{\rm 2h,s}_{\rm gdm}(r, L_1, L_2)
\end{aligned} \tag{4.76}$$

This expression involves the convolution of the distribution of satellite galaxies inside a halo with the dark matter density with the halo-halo correlation function. The Fourier counterpart simplifies particularly:

$$\begin{aligned}
\tilde{P}^{\rm 2h,s}_{\rm gdm}(k, L_1, L_2) &= \frac{1}{\bar{\rho}\bar{n}_{\rm tot}(L_1, L_2)} \int n(M_1)\, \langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\, {\rm d}M_1 \int n(M_2)\, M_2\, {\rm d}M_2 \\
&\quad u_{\rm s}(k|M_1)\, u_{\rm dm}(k|M_2)\, P_{\rm hh}(k, M_1, M_2) = \\
&= \frac{P_{\rm dm}(k)}{\bar{\rho}\bar{n}_{\rm tot}(L_1, L_2)} \int \langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\, u_{\rm s}(k|M_1)\, b(M_1)\, n(M_1)\, {\rm d}M_1 \\
&\quad \int M_2\, u_{\rm dm}(k|M_2)\, b(M_2)\, n(M_2)\, {\rm d}M_2 \; = \\
&= f_{\rm s}(L_1, L_2)\, \frac{P_{\rm dm}(k)}{\bar{\rho}\bar{n}_{\rm s}(L_1, L_2)} \int \langle N_{\rm s}\rangle_{M_1}(L_1, L_2)\, u_{\rm s}(k|M_1)\, b(M_1)\, n(M_1)\, {\rm d}M_1 \\
&\quad \int M_2\, u_{\rm dm}(k|M_2)\, b(M_2)\, n(M_2)\, {\rm d}M_2 \; = \\
&= f_{\rm s}(L_1, L_2)\, \check{P}^{\rm 2h,s}_{\rm gdm}(k, L_1, L_2) \qquad (4.77)
\end{aligned}$$

where we have used the assumption that the halo bias is scale independent, i.e.

$$P_{\rm hh}(k|M_1, M_2) = b(M_1) b(M_2) P_{\rm dm}(k)\,. \qquad (4.78)$$

4.3.5 The Excess Surface Density

The excess surface density is defined as

$$\Delta\Sigma(R) = \Sigma(< R) - \Sigma(R)\,, \qquad (4.79)$$

where

$$\Sigma(< R) = \frac{2}{R^2} \int \Sigma(R') R' {\rm d}R' \qquad (4.80)$$

and the projected surface density, $\Sigma(R)$, is related to the galaxy-dark matter cross correlation, $\xi_{\rm g,dm}(r)$, according to

$$\Sigma(R) = \overline{\rho} \int_0^{\omega_{\rm s}} \left[1 + \xi_{\rm g,dm}(r)\right] {\rm d}\omega\,, \qquad (4.81)$$

where $\overline{\rho}$ is the average density of matter in the universe and the integral is along the line of sight with ω the comoving distance from the observer. The three-dimensional comoving distance r is related to ω through $r^2 = \omega_{\rm L}^2 + \omega^2 - 2\omega_{\rm L}\omega\cos\theta$ (see Fig. 4.6 for an illustration of the geometry).

Since $\xi_{\rm g,dm}(r)$ goes to zero in the limit $r \to \infty$, and since in practice θ is small, we can approximate Eq. (4.81) using

$$\Sigma(R) = 2\overline{\rho} \int_R^{\infty} \left[1 + \xi_{\rm g,dm}(r)\right] \frac{r{\rm d}r}{\sqrt{r^2 - R^2}}\,, \qquad (4.82)$$

which is the expression we adopt throughout.

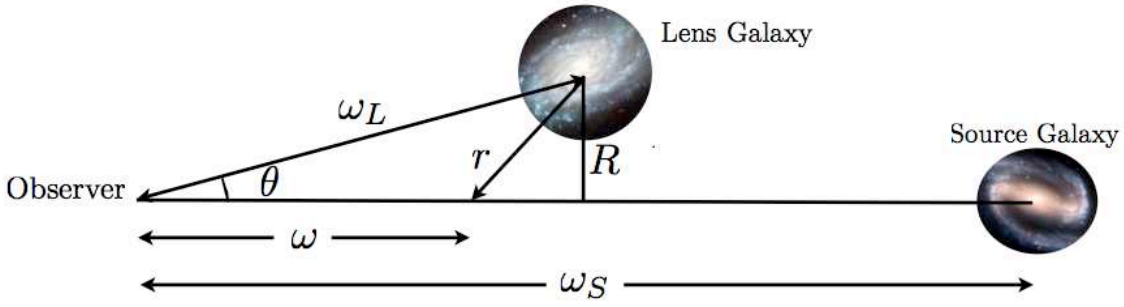

Figure 4.6: Illustration of the geometry between source, lens and observer

4.4 Technical detail: the actual evaluation of the integrals

A technical, as well as conceptual, issue arises in calculating the two halo terms of the galaxy-dark matter cross correlation function. Let us rewrite these two equations in the following compact form:

$$\begin{aligned} \check{P}^{\rm 2h,c}_{\rm g,dm}(k) =& P_{\rm dm}(k)\, \mathcal{I}_{N_{\rm c}}\, \mathcal{I}_M(k) \\ \check{P}^{\rm 2h,s}_{\rm g,dm}(k) =& P_{\rm dm}(k)\, \mathcal{I}_{N_{\rm s}}(k)\, \mathcal{I}_M(k)\,, \end{aligned} \tag{4.83}$$

where $\mathcal{I}_{N_{\rm c}}$, $\mathcal{I}_{N_{\rm s}}(k)$ and $\mathcal{I}_M(k)$ are

$$\begin{aligned} \mathcal{I}_{N_{\rm c}} &= \int \frac{\langle N_{\rm c}\rangle_M(L_1,L_2)}{\overline{n}_{\rm c}}\, b(M)\, n(M)\, {\rm d}M\,, \\ \mathcal{I}_{N_{\rm s}}(k) &= \int \frac{\langle N_{\rm s}\rangle_M(L_1,L_2)}{\overline{n}_{\rm s}}\, u_{\rm s}(k|M)\, b(M)\, n(M)\, {\rm d}M\,, \\ \mathcal{I}_M(k) &= \int \frac{M}{\overline{\rho}} u_{\rm dm}(k|M)\, b(M)\, n(M)\, {\rm d}M\,. \end{aligned} \tag{4.84}$$

The evaluation of these integrals is somewhat tedious numerically, as it requires knowledge of the halo mass function and the halo bias function over the entire mass range $[0,\infty[$. Since these have only been tested against numerical simulations over a limited range of halo masses ($10^9 h^{-1} M_\odot \lesssim M \lesssim 10^{15} h^{-1} M_\odot$), it is also unclear how accurate they are. In practice, though, these problems can be circumvented as follows. First of all, because of the exponential cut-off in the halo mass function, it is sufficiently accurate to perform the integrations of Eq. (4.84) only up to $M = 10^{16} h^{-1} M_\odot$. Secondly, $\mathcal{I}_{N_{\rm c}}$ and $\mathcal{I}_{N_{\rm s}}(k)$ contain the halo occupation statistics, $\langle N_{\rm c}\rangle_M(L_1,L_2)$ and $\langle N_{\rm s}\rangle_M(L_1,L_2)$, respectively, which, for all luminosities of interest in this thesis, are equal to zero for $M \lesssim 10^9 h^{-1} M_\odot$. Therefore, $\mathcal{I}_{N_{\rm c}}$ and $\mathcal{I}_{N_{\rm s}}(k)$ can be computed accurately by only integrating over the mass range $[10^9 - 10^{16}] h^{-1} M_\odot$. Unfortunately, the integrand of $\mathcal{I}_M(k)$ does not become negligibly small below a given halo mass. However, in this case we can use the approach introduced by Yoo et al.(2006): we write $\mathcal{I}_M(k)$ as the

sum of two terms, $\mathcal{I}_M(k) = \mathcal{I}_{M_1}(k) + \mathcal{I}_{M_2}(k)$, where:

$$\begin{aligned} \mathcal{I}_{M_1}(k) &= \int_0^{M_{\rm min}} \frac{M}{\bar{\rho}}\, u_{\rm dm}(k|M)\, b(M)\, n(M)\, {\rm d}M\,, \\ \mathcal{I}_{M_2}(k) &= \int_{M_{\rm min}}^{\infty} \frac{M}{\bar{\rho}}\, u_{\rm dm}(k|M)\, b(M)\, n(M)\, {\rm d}M\,. \end{aligned} \tag{4.85}$$

Following Yoo et al.(2006), we use the fact that $u_{\rm dm}(k|M) = 1$ over the relevant range of k as long as M is sufficiently small (see also fig. 4.4). This allows us to write

$$\begin{aligned} \mathcal{I}_{M_1}(k) &\simeq \int_0^{M_{\rm min}} \frac{M}{\bar{\rho}}\, b(M)\, n(M)\, {\rm d}M \\ &= 1 - \int_{M_{\rm min}}^{\infty} \frac{M}{\bar{\rho}}\, b(M)\, n(M)\, {\rm d}M\,. \end{aligned} \tag{4.86}$$

where the last equality follows from the fact that the distribution of matter is by definition unbiased with respect to itself. Detailed tests have shown that this procedure yields results that are sufficiently accurate as long as $M_{\rm min} \lesssim 10^{10} h^{-1} M_\odot$. Throughout we adopt $M_{\rm min} = 10^9 h^{-1} M_\odot$.

As a final remark, we would like to emphasize the phenomenological character of the halo model. Although the model is fully analytical, it is tightly linked to numerical simulations. In fact, the halo mass function, $n(M)$, the halo density profile, $\rho_{\rm dm}(r|M)$, the halo bias function, $b(M)$, must be provided by numerical simulations.

4.5 The Conditional Luminosity Function

The analytical definitions of the correlation functions and power spectra defined above require the halo occupation statistics, $\langle N_{\rm c}\rangle_M(L_1, L_2)$ and $\langle N_{\rm s}\rangle_M(L_1, L_2)$. In order to obtain it, we use the conditional luminosity function (CLF), $\Phi(L|M){\rm d}L$, which specifies the average number of galaxies with luminosities in the range $L \pm {\rm d}L/2$ that reside in a halo of mass M (Yang, Mo & van den Bosch 2003; van den Bosch, Yang, Mo 2003). Following Cooray & Milosavljević (2005) and Cooray (2006), we write the CLF as

$$\Phi(L|M) = \Phi_{\rm c}(L|M) + \Phi_{\rm s}(L|M)\,, \tag{4.87}$$

where $\Phi_{\rm c}(L|M)$ and $\Phi_{\rm s}(L|M)$ represent central and satellite galaxies, respectively. The occupation numbers required for the computation of the galaxy-dark matter cross correlation then simply follow from

$$\langle N_{\rm x}\rangle_M(L_1, L_2) = \int_{L_1}^{L_2} \Phi_{\rm x}(L|M){\rm d}L\,. \tag{4.88}$$

where 'x' refers to either 'c' (centrals) or 's' (satellites).

4.5.1 The Central Galaxy Term

Motivated by the results of Yang, Mo & van den Bosch (2008; hereafter YMB08), who analyzed the CLF obtained from the SDSS galaxy group catalogue of Yang et al.(2007, hereafter Y07), we assume the contribution from the central galaxies to be a log-normal:

$$\Phi_{\rm c}(L|M) = \frac{1}{\sqrt{2\pi}\ln(10)\,\sigma_{\rm c}\,L}\exp\left[-\frac{(\log L - \log L_{\rm c})^2}{2\sigma_{\rm c}^2}\right]. \tag{4.89}$$

Note that $\sigma_{\rm c}$ is the scatter in $\log L$ (of central galaxies) at a fixed halo mass. Moreover, $\log L_{\rm c}$ is, by definition, the expectation value for the (10-based) logarithm of the luminosity of the central galaxy, i.e.,

$$\log L_{\rm c} = \int_0^\infty \Phi_{\rm c}(L|M)\,\log L\,{\rm d}L\,. \tag{4.90}$$

4.5.2 The Satellite Galaxy Term

For the contribution from the satellite galaxies we adopt a modified Schechter function:

$$\Phi_{\rm s}(L|M) = \frac{\phi_{\rm s}^*}{L_{\rm s}^*}\left(\frac{L}{L_{\rm s}^*}\right)^{\alpha_{\rm s}}\exp\left[-\left(\frac{L}{L_{\rm s}^*}\right)^2\right]. \tag{4.91}$$

which decreases faster than a Schechter function at the bright end. Note that $L_{\rm c}$, $\sigma_{\rm c}$, $\phi_{\rm s}^*$, $\alpha_{\rm s}$ and $L_{\rm s}^*$ are all functions of the halo mass M. In the parametrization of these mass dependencies, we again are guided by the results of YMB08. In particular, for the luminosity of the central galaxies we adopt

$$L_{\rm c}(M) = L_0\frac{(M/M_1)^{\gamma_1}}{[1+(M/M_1)]^{\gamma_1-\gamma_2}}\,, \tag{4.92}$$

so that $L_{\rm c}\propto M^{\gamma_1}$ for $M\ll M_1$ and $L_c\propto M^{\gamma_2}$ for $M\gg M_1$. Here M_1 is a characteristic mass scale, and $L_0 = 2^{\gamma_1-\gamma_2}L_c(M_1)$ is a normalization. Using the SDSS galaxy group catalogue, YMB08 found that to good approximation

$$L_{\rm s}^*(M) = 0.562 L_{\rm c}(M) \tag{4.93}$$

and we adopt this parameterization throughout. For the faint-end slope and normalization of $\Phi_{\rm s}(L|M)$ we adopt

$$\alpha_{\rm s}(M) = -2.0 + a_1\left(1-\frac{2}{\pi}\arctan[a_2\log(M/M_2)]\right), \tag{4.94}$$

and

$$\log[\phi_{\rm s}^*(M)] = b_0 + b_1(\log M_{12}) + b_2(\log M_{12})^2\,, \tag{4.95}$$

with $M_{12} = M/(10^{12}h^{-1}\ {\rm M}_\odot)$. This adds a total of six free parameters: a_1, a_2, b_0, b_1, b_2 and the characteristic halo mass M_2. Neither of these functional forms has a physical motivation; they merely were found to adequately describe the results obtained by YMB08. Finally, for simplicity, and to limit the number of free parameters, we assume that $\sigma_{\rm c}(M) = \sigma_{\rm c}$ is a constant. As shown in More

et al.(2009b), this assumption is supported by the kinematics of satellite galaxies in the SDSS. Thus, altogether the CLF has a total of eleven free parameters.

Note that, with the parametrization of the CLF introduced above, the halo occupation statistics can be rewritten as:

$$\langle N_{\rm c}\rangle_M(L_1,L_2) = \int_{L_1}^{L_2} \Phi_{\rm c}(L|M){\rm d}L = \frac{1}{2}\Big[{\rm erf}(x_2) - {\rm erf}(x_1)\Big] \tag{4.96}$$

and

$$\langle N_{\rm s}\rangle_M(L_1,L_2) = \int_{L_1}^{L_2} \Phi_{\rm s}(L|M){\rm d}L = \frac{\phi_{\rm s}^*}{2}\left\{\Gamma\left[\frac{\alpha_{\rm s}}{2}+\frac{1}{2},\left(\frac{L_1}{L_{\rm s}^*}\right)^2\right] - \Gamma\left[\frac{\alpha_{\rm s}}{2}+\frac{1}{2},\left(\frac{L_2}{L_{\rm s}^*}\right)^2\right]\right\}, \tag{4.97}$$

where ${\rm erf}(x_i)$ is the error function calculated at $x_i = \log(L_i/L_{\rm c})/(\sqrt{2}\sigma_{\rm c})$ with $i = 1,2$ and Γ is the incomplete gamma function. Moreover, we write the galaxy luminosity function, $\Phi(L)$, as

$$\Phi(L) = \int \Phi(L|M)n(M){\rm d}M\,. \tag{4.98}$$

4.6 The setting of the reference model

This section is basically a list of all the assumptions that enter the model which we shall use throughout the thesis. In practice, the correlation functions introduced in this chapter rely on the knowledge of certain relations quantifying the properties of dark matter haloes and of galaxies within them. Moreover, specific choices have been made in order to simplify the analytical approach. In what follows, we list all the assumptions characterizing our reference model.

- Dark matter haloes are spherical overdense region with mass $M = (4\pi/3)180\bar{\rho}r_{180}^3$.
- The dark matter density profile is assumed to follow the NFW functional form.
- For the concentration-halo mass relation, which characterizes the NFW profile, we use the relation suggested by Macciò et al. 2007.
- The number density of dark matter haloes, $n(M)$, follows the EPS formalism and in particular the fitting function provided by Sheth & Tormen (2002) is used.
- A central galaxy is sitting exactly at the center of the dark matter halo (no off-set is taken into account).
- Satellite galaxies are assumed to follow a Poisson distribution.
- The radial distribution of galaxies within a halo is assumed to be independent of the luminosity, $\hat{n}_{\rm s}(r|M,L_1,L_2) = \hat{n}_{\rm s}(r|M)$. This means that no luminosity segregation is taken into account.

- The radial distribution of satellite galaxies within the halo is assumed to be the same as the one for dark matter, namely $\hat{n}_{\rm s}(r|M) = \hat{\rho}_{\rm dm}(r|M)$.
- The halo occupation statistics, $\langle N_{\rm c}\rangle_M(L_1, L_2)$ and $\langle N_{\rm s}\rangle_M(L_1, L_2)$, is provided by the conditional luminosity function.
- The halo bias is assumed to be scale-independent. Only in defining the correlation length, the scale dependence is taken into account following the prescription from Tinker et al. (2005).
- The halo bias-halo mass relation is the one suggested by Sheth & Tormen (2002)
- The two halo terms of the power spectra are written as a function of the non-linear dark matter power spectrum, $P_{\rm dm}(k) = P^{\rm nl}_{\rm dm}(k)$.
- The non-linear dark matter power spectrum, $P^{\rm NL}_{\rm dm}(k)$, is calculated via the relation suggested by Smith et al. (2003)

5

Galaxy Clustering & Galaxy-Galaxy Lensing: An Astrophysical Perspective

In this chapter, we establish in a quantitative way the connection between galaxies and dark matter haloes by constraining the parameters which identify the conditional luminosity function (CLF). Furthermore, we predict the galaxy-galaxy lensing signal using the halo occupation statistics derived from the CLF. We extensively investigate the impact of the assumptions which enter our analytical treatment. We then conclude that the CLF is ideal for an accurate description of the g-g lensing signal and the uncertainties due to the model assumptions do not substantially affect the predictions.

5.1 Establishing the Galaxy-Dark Matter Connection

As shown in Yang et al.(2003) and van den Bosch et al.(2003), the CLF can be constrained using the observed luminosity function, $\Phi(L)$, and the galaxy-galaxy correlation lengths as a function of luminosity, $r_0(L)$. Here we use the luminosity function (hereafter LF) of Blanton et al.(2003a) uniformly sampled at 41 magnitudes covering the range $-23.0 \leq {}^{0.1}M_r - 5\log h \leq -16.4$. Here ${}^{0.1}M_r$ indicates the r-band magnitude K+E corrected to $z = 0.1$ following the procedure of Blanton et al.(2003b). For the correlation lengths as function of luminosity we use the results obtained by Wang et al.(2007) for six volume limited samples selected from the SDSS DR4. For a thorough description of the data see chapter 2. Finally, to strengthen our constraints, and to assure agreement with the CLF obtained from a SDSS group catalogue, we use the constraints on $L_c(M)$, $\alpha_s(M)$ and $\phi_s^*(M)$ obtained by YMB08. We present the main results of this catalogue in the next section, whereas the method to constrain the CLF is explained in § 5.1.2.

5.1.1 The SDSS Galaxy Group Catalogue

A direct way of studying the galaxy-dark matter halo connection is by using sets of galaxies that reside in the same dark matter halo. In our jargon, we refer to such systems as *galaxy groups* regardless of the number of galaxies in them (i.e. from groups with a single member to rich clusters of galaxies). During the past two decades, numerous galaxy group catalogues have been constructed from various galaxy redshift surveys, e.g. the CfA redshift survey (e.g. Geller & Huchra 1983), the 2-degree Field Galaxy Redshift Survey (hereafter 2dFGRS; Merchán & Zandivarez 2002; Eke et al.2004, Yang et al.2005; Tago et al.2006; Einasto et al.2007), the high-redshift DEEP2 survey (Gerke et al.2005), and the Two Micron All Sky Redshift Survey (Crook et al.2007). Various group catalogues have also been constructed from the SDSS redshift samples (e.g. Berlind et al.2006; Abazajian et al.2004; Merchán & Zandivarez 2005) and from the SDSS photometric data (e.g. Koester et al.2007). All these catalogues differ either for the selected galaxies or for the criteria used to identify the groups.

Yang et al.(2005) have developed a halo-based group finder that is optimized for grouping galaxies that reside in the same dark matter halo. Using mock galaxy redshift surveys constructed from the conditional luminosity function model (see Yang et al. 2004), they found that this group finder is very successful in associating galaxies according to their common dark matter haloes. In particular, the group finder performs also reliably for poor systems, including isolated galaxies in small mass haloes. This makes this halo-based group finder ideally suited to study the relation between galaxies and dark matter haloes over a wide dynamic range in halo masses. In this section, we briefly describe a version applied to the NYU-VAGC based on the SDSS DR4 (see Yang et al. 2007 for a thorough description of the methodology used to build this galaxy group catalogue). The strength of the group finder adopted here, hereafter referred to as the halo-based group finder, is that it is iterative and based on an adaptive filter modeled after the general properties of dark matter haloes. In addition, unlike the traditional FOF method, this group finder can also identify groups with only a single member. Furthermore, we stress that an important aspect of each galaxy group catalogue is the determination of the masses of the groups. Most studies infer the (dynamical) group mass from the velocity dispersion of their member galaxies. However, the vast majority of the groups in our sample contain only a few members making a dynamical mass estimate based on its members extremely unreliable. Mass estimates based on gravitational lensing (either strong or weak) or on X-ray emission, also can only be applied to the most massive systems. Furthermore, these latter two methods require high-quality data in excess to the information directly available from the redshift survey used to construct the group catalogue, rendering them impractical.

In the galaxy group catalogue used in this thesis, group masses are estimated from the group luminosities. We define the group's *characteristic* luminosity, $L_{19.5}$, as the combined luminosity of all group members with ${}^{0.1}\mathrm{M}_r - 5\log h \leq -19.5$ (here again, all absolute magnitudes are $K+E$ corrected to $z = 0.1$). In order to convert the characteristic luminosities to halo masses, we make the assumption that there is a one-to-one relation between $L_{19.5}$ and M. For a given (comoving) volume and a given halo mass function, $n(M)$, one can then link the characteristic luminosity to a halo mass by matching their rank orders. The use of the group's characteristic luminosity has the advantage

that (i) it is equally applicable to groups spanning the entire range in richness, and (ii) it does not require any additional data. However, a shortcoming of this method is that it requires the halo mass function, which is cosmology dependent (e.g. Sheth, Mo & Tormen 2001; Warren et al. 2006). However, as explained in Yang et al. (2007) , it is extremely easy to convert the group masses to another cosmology, without having to rerun the group finder. Clearly, the assumption of a one-to-one relation between the characteristic luminosity and the halo mass is oversimplified. In reality, this relation contains some scatter, which results in errors in the inferred group masses. However, detailed tests with mock galaxy redshift surveys have shown that this method nevertheless allows for a very accurate recovery of *average* halo occupation statistics. In particular, the group finder yields average halo occupation numbers and average mass-to-light ratios that are in excellent agreement with the input values (Yang et al.2005; Weinmann et al.2006a).

For the purposes of this thesis, the relevance of the group catalogue resides indeed in providing information about statistical properties of the halo occupation statistics. Firstly, the average luminosity of central galaxies is provided as a function of the host halo mass. Secondly, the group catalogue provides quantitative constraints on the average number of satellite galaxies living in a halo of a certain mass. More precisely, if the number of satellite galaxies of a given luminosity residing in a halo of a certain mass is assumed to follow a given functional form, the parameters which define this functional form may be constrained by the group catalogue. In chapter 4, we have introduced the conditional luminosity function, $\Phi(L|M)$, as the sum of two contributions: the one which describes the central galaxies and the one which describes the satellite galaxies. The analytical form assumed in that chapter has 11 free parameters. In the next section, we show how to constrain these parameters with observational data and with the indications from the group catalogue.

5.1.2 Constraining the CLF

For a given set of model parameters, we compute the LF using

$$\Phi(L) = \int_0^\infty \Phi(L|M)\, n(M)\, \mathrm{d}M\,. \tag{5.1}$$

The galaxy-galaxy correlation function for galaxies with luminosities in the interval $[L_1, L_2]$ is computed using

$$\xi_{\rm gg}(r) = b^2_{\rm gal}(L_1, L_2)\, \zeta(r)\, \xi^{\rm NL}_{\rm dm}(r)\,. \tag{5.2}$$

Here $\xi^{\rm NL}_{\rm dm}(r)$ is the non-linear correlation function of the dark matter, which is the Fourier transform of $P^{\rm NL}_{\rm dm}(k)$. Moreover

$$\zeta(r) = \frac{[1 + 1.17\xi^{\rm NL}_{\rm dm}(r)]^{1.49}}{[1 + 0.69\xi^{\rm NL}_{\rm dm}(r)]^{2.09}} \tag{5.3}$$

is the radial scale dependence of the bias as obtained by Tinker et al.(2005), and $b_{\rm gal}(L_1, L_2)$ is the bias of the galaxies, which is related to the CLF according to

$$b_{\rm gal}(L_1, L_2) = \frac{\int_0^\infty \langle N\rangle_M\, b(M)\, n(M)\, \mathrm{d}M}{\int_0^\infty \langle N\rangle_M\, n(M)\, \mathrm{d}M}\,, \tag{5.4}$$

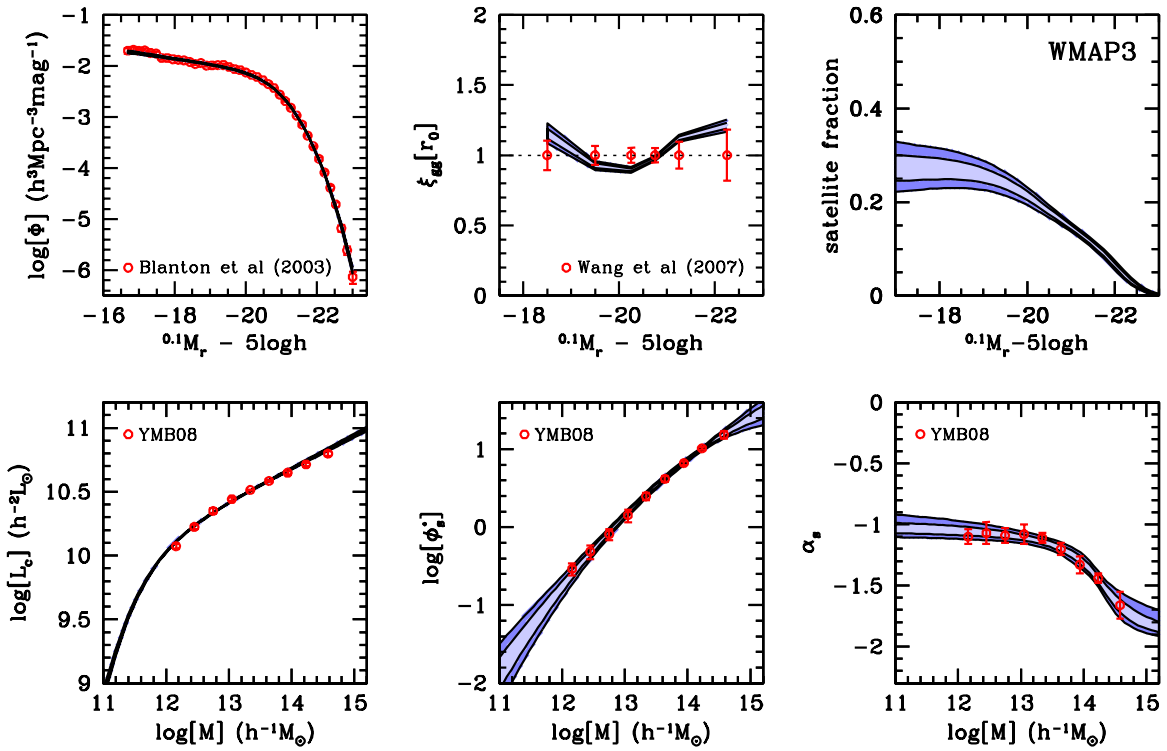

Figure 5.1: *Upper row, left and central panels.* The luminosity function of galaxies and the luminosity dependence of the galaxy correlation length are plotted. Data come from the analysis of Blanton et al.(2003a) and Wang et al.2007. The blue contours indicate the 68 and 95 percent confidence level obtained from the MCMC. The agreement is extremely accurate for the luminosity function whereas is modest for the correlation length. *Lower row, three panels.* The additional information coming from the group catalogue of YMB08 is plotted together with the corresponding 68 and 95 percent confidence level derived with the MCMC. In particular, the halo mass dependence of the central galaxy luminosity, the satellite conditional luminosity function normalization ϕ_s^* and the the exponent α_s are shown in the left, central and middle panel, respectively. *Upper row, right panel.* The 68 and 95 percent confidence levels of the satellite fraction, f_s, obtained from the CLF (see eq. [5.10]).

with

$$\langle N\rangle_M = \int_{L_1}^{L_2} \Phi(L|M)\mathrm{d}L = \langle N_\mathrm{c}\rangle_M(L_1,L_2) + \langle N_\mathrm{s}\rangle_M(L_1,L_2)\,. \tag{5.5}$$

the average number of galaxies with luminosities in the range $[L_1, L_2]$ that reside in a halo of mass M.

To determine the likelihood function of our 11 free parameters we follow van den Bosch et al.(2007) and use the Monte-Carlo Markov Chain (hereafter MCMC) technique. We construct a chain of 5 million models. Each element of the chain is a model consisting of 11 parameters. At any point in the chain we generate a new trial model by drawing the shifts in the 11 free parameters from 11 independent Gaussian distributions centered on the current value of the corresponding model parameter. The chain is thinned by a factor of 2500 to remove the correlations between neighboring models. The end result is a MCMC of 2000 independent models properly sampling the posterior distributions of the 11 free parameters. The probability of accepting the trial model is

$$P_\mathrm{accept} = \begin{cases} 1.0 & \text{if } \chi^2_\mathrm{new} < \chi^2_\mathrm{old} \\ \exp[-(\chi^2_\mathrm{new} - \chi^2_\mathrm{old})] & \text{if } \chi^2_\mathrm{new} \geqslant \chi^2_\mathrm{old} \end{cases} \tag{5.6}$$

Here $\chi^2 = \chi^2_\Phi + \chi^2_{r_0} + \chi^2_\mathrm{GC}$ with

$$\chi^2_\Phi = \sum_{i=1}^{41} \left[\frac{\Phi(L_i) - \hat{\Phi}(L_i)}{\Delta\hat{\Phi}(L_i)}\right]^2 , \tag{5.7}$$

$$\chi^2_{r_0} = \sum_{i=1}^{6} \left[\frac{\xi_\mathrm{gg}(r_{0,i}) - 1}{\Delta\hat{\xi}_\mathrm{gg}(r_{0,i})}\right]^2 , \tag{5.8}$$

and

$$\begin{aligned} \chi^2_\mathrm{GC} &= \sum_{i=1}^{9} \left[\frac{\log L_c(M_i) - \log \hat{L}_c(M_i)}{\Delta \log \hat{L}_c(M_i)}\right]^2 \\ &+ \sum_{i=1}^{9} \left[\frac{\alpha_s(M_i) - \hat{\alpha}_s(M_i)}{\Delta\hat{\alpha}_s(M_i)}\right]^2 \\ &+ \sum_{i=1}^{9} \left[\frac{\phi^*_s(M_i) - \hat{\phi}^*_s(M_i)}{\Delta\hat{\phi}^*_s(M_i)}\right]^2 . \end{aligned} \tag{5.9}$$

Here $\hat{\cdot}$ indicates an observed quantity and the subscripts 'Φ', 'r_0' and 'GC' refer to the luminosity function, the galaxy-galaxy correlation length and the group catalogue, respectively. Note that, by definition, $\hat{\xi}_\mathrm{gg}(r_{0,i}) = 1$. Table 5.1 lists the best-fit parameters with the corresponding 95% confidence levels obtained with the MCMC technique, as well as the corresponding value of $\chi^2_\mathrm{red} = \chi^2/N_\mathrm{dof}$. Here $N_\mathrm{dof} = 74 - 11 = 63$ is the number of degrees of freedom.

Table 5.1: Best-fit CLF parameters obtained from SDSS clustering analysis

	WMAP3
$\chi^2_{\rm red}$	1.42
$\log L_0$	9.935 ±0.043
$\log M_1$	11.07 ±0.11
γ_1	3.273 ±0.575
γ_2	0.255 ±0.008
a_1	0.501 ±0.069
a_2	2.106 ±1.437
$\log M_2$	14.28 ±0.16
b_0	-0.766 ±0.146
b_1	1.008 ±0.197
b_2	-0.094 ±0.065
$\sigma_{\rm c}$	0.143 ±0.005

The best-fit CLF parameters and the corresponding 95% confidence levels obtained from the MCMC analysis and the value of the corresponding reduced χ^2. Masses and luminosities are in h^{-1} $M_\odot$ and h^{-2} $L_\odot$, respectively. Note that all the parameters are tightly constrained except γ_1, a_2 and b_2.

5.1.2.1 Results

Fig. 5.1 shows the results of the MCMC which explores the 11 parameter space defining the CLF. In each panel the blue contours indicate the 68 and 95 percent confidence levels obtained from the MCMC. The upper left-hand panel shows that the CLF model accurately fits the galaxy LF of Blanton et al.(2003a). The fit to the correlation lengths as function of luminosity, shown in the upper middle panel, is less accurate, although data and model typically agree at the 1σ level. We have been unable to obtain a better fit, despite various attempts, including small changes in the CLF model. The lower panels of Fig. 5.1 show the 68 and 95 percent confidence levels on $L_{\rm c}(M)$, $\phi^*_{\rm s}(M)$ and $\alpha_{\rm s}(M)$, compared with the results obtained by YMB08 from the SDSS group catalogue of Y07. Since these data have been used as additional constraints, it should not come as a big surprise that the CLF is in good agreement with these data. We emphasise, though, that it is not trivial that a single halo occupation model can be found that can simultaneously fit the LF, the luminosity dependence of the galaxy-galaxy correlation functions, and the results obtained from a galaxy group catalogue. Note that the lower left panel of Fig. 5.1 plots the CLF predictions down to $\log L_{\rm c} = 9$. Although there are no data from the group catalogue at this luminosity scale, the central galaxy luminosity-halo mass relation at the faint end is tightly constrained by the clustering data and the luminosity function.

Finally, the upper right-hand panel of Fig. 5.1 shows the satellite fraction,

$$f_{\rm s}(L) = \frac{1}{\Phi(L)} \int_0^\infty \Phi_{\rm s}(L|M)\, n(M)\, {\rm d}M\,. \tag{5.10}$$

as function of luminosity. This is found to decrease from $\sim 0.27 \pm 0.03$ at ${}^{0.1}M_r - 5\log h = -17$ to virtually zero at ${}^{0.1}M_r - 5\log h = -23$. The fact that the satellite fraction decreases with increasing luminosity is in qualitative agreement with previous studies (Tinker et al.2006; Mandelbaum et al.2006; van den Bosch et al.2007).

5.2 Predicting the Excess Surface Density

Although the advent of wide and deep surveys has resulted in clear detections of galaxy-galaxy lensing, a proper interpretation of these data in terms of the link between galaxies and dark matter haloes has been hampered by the fact that the lensing signal can only be detected when stacking the signal of many lenses. Since not all lenses reside in haloes of the same mass, the resulting signal is a non-trivial average of the lensing signal due to haloes of different masses. In addition, central galaxies (those residing at the center of a dark matter halo) and satellite galaxies (those orbiting around a central galaxy) contribute very different lensing signals, even when they reside in haloes of the same mass (e.g., Yang et al.2006). Fig. 5.2 shows a schematic representation of this concept. All these complications have to be properly accounted for, and require knowledge of both the satellite fractions and of the spatial number density distribution of satellite galaxies within their dark matter haloes.

The CLF presented in this thesis is ideally suited to model galaxy-galaxy lensing. In particular, it is straightforward to account for the fact that there is scatter in the relation between the luminosity of

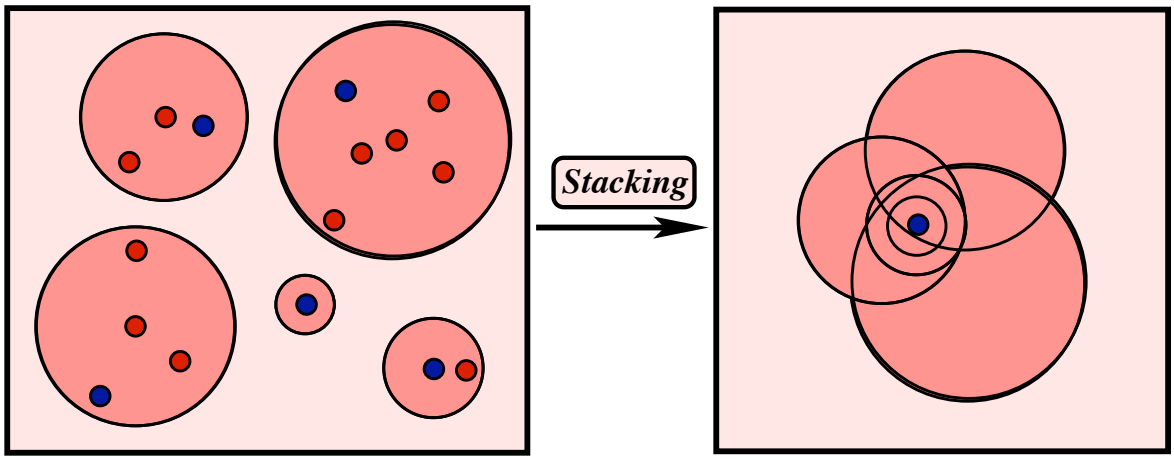

Figure 5.2: Schematic representation of the stacking of galaxies. Pink circles represents dark matter haloes, whereas red and blue circles represent galaxies of different luminosity. By stacking all galaxies with the same luminosity (all those in blue), haloes of different masses are stacked together. Note also that satellite galaxies contribute to the stack in a different way than central galaxies.

a central galaxy and the mass of its dark matter halo. This represents an improvement with respect to previous attempts to model the g-g lensing signal obtained from the SDSS, which typically ignored this scatter (e.g. Seljak et al.2005; Mandelbaum et al.2006). However, in agreement with Tasitsiomi et al.(2004), we show that the scatter in this relation has an important impact on the g-g lensing signal and cannot be ignored. Given this dependence on the scatter it may seem advantageous to use stellar mass as a stacking property rather than luminosity (r-band). However, as shown in Yang, Mo & van den Bosch (2008), although the scatter in stellar mass-halo mass is significantly smaller than for the luminosity-halo mass relation, it is still substantial. Consequently, it is important that scatter is properly accounted in the model when stellar mass is used to stack lens galaxies. We also improved upon previous studies by modelling the two halo term (the contribution to the lensing signal due to the mass distribution outside of the halo hosting the lens galaxy), including an approximate treatment for halo exclusion.

5.2.1 Model Predictions

In order to compute the ESD, $\Delta\Sigma$, as a function of the comoving separation on the sky, R, we proceed as follows. We start by calculating the four different terms of the galaxy-dark matter cross power spectrum defined in chapter 4. Next, we inverse Fourier transform each of these terms using

$$\xi^{\mu,\mathrm{x}}_{\mathrm{g,dm}}(r) = \frac{1}{2\pi^2}\int f_{\mathrm{x}}\,\check{P}^{\mu,\mathrm{x}}_{\mathrm{g,dm}}(k)\,\frac{\sin(kr)}{kr}\,k^2\,\mathrm{d}k = \frac{1}{2\pi^2}\int \tilde{P}^{\mu,\mathrm{x}}_{\mathrm{g,dm}}(k)\,\frac{\sin(kr)}{kr}\,k^2\,\mathrm{d}k\,, \qquad (5.11)$$

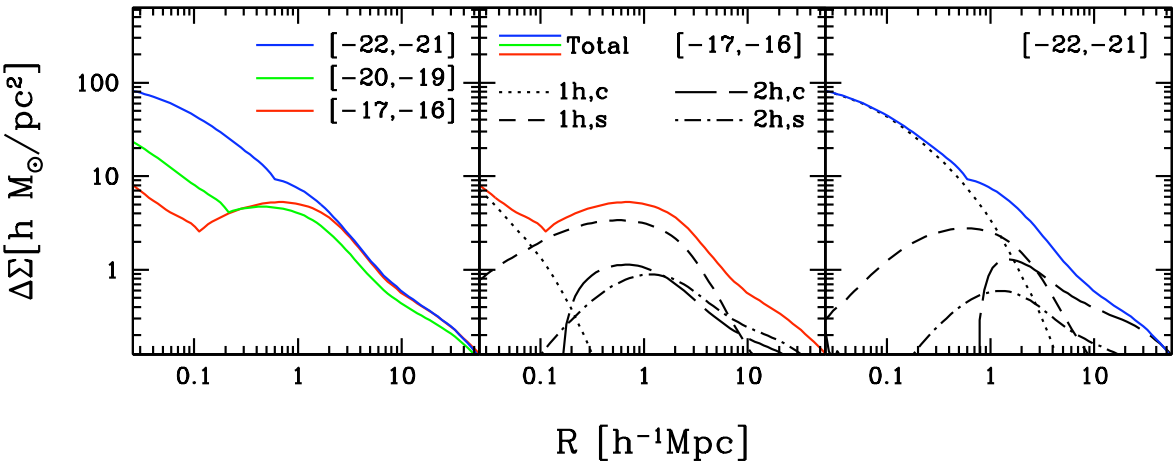

Figure 5.3: The predicted ESD up to large scales ($R \sim 30h^{-1}$Mpc) for three luminosity bins, as indicated. The solid lines refer to the total signal as predicted according to our model. The dotted lines refer to the one halo central term, whereas the dashed lines refer to the one halo satellite term. Note that they dominate the signal on different scales (see text). The long dashed lines refer to the two halo central term. It rises steeply at relatively large scales due to our halo exclusion treatment (see § 5.2.3.4). The two halo satellite term is indicated with the dotted-dashed line.

where 'μ' stands for 1h or 2h, and 'x' refers to either 'c' (centrals) or 's' (satellites). These are used to compute the corresponding four terms of the surface density, $\Sigma^{\mu,\mathrm{x}}(R)$,

$$\Sigma^{\mu,\mathrm{x}}(R) = 2\overline{\rho} \int_R^\infty \xi^{\mu,\mathrm{x}}_{\mathrm{g,dm}}(r) \frac{r\mathrm{d}r}{\sqrt{r^2 - R^2}} \,. \tag{5.12}$$

Note that we are allowed to neglect the contribution coming from the constant background density, $\bar{\rho}$, (cf. equation [4.82]) because it will cancel in defining the ESD (this is known in gravitational lensing theory as the mass-sheet degeneracy). The final ESD then simply follows from

$$\begin{aligned} \Delta\Sigma(R) &= \Delta\Sigma^{\mathrm{1h,c}}(R) + \Delta\Sigma^{\mathrm{1h,s}}(R) \\ &+ \Delta\Sigma^{\mathrm{2h,c}}(R) + \Delta\Sigma^{\mathrm{2h,s}}(R) \,. \end{aligned} \tag{5.13}$$

in which the relative weight of each term is already included via the central and satellite fractions in the definitions of the corresponding power spectra.

Before comparing the g-g lensing predictions from our CLF models to actual data, we first demonstrate how the four different terms contribute to the total ESD. The left-hand panel of Fig. 5.3 shows the $\Delta\Sigma(R)$ obtained from our best-fit CLF model for three different luminosity bins, as indicated[1]. Note that the fainter luminosity bins reveal a more 'structured' excess surface density profile, with

[1]Here, for simplicity, we have used the halo mass function and halo bias function computed at $z = 0$. When we compare our models to data, we will use the halo mass function and halo bias function at the average redshift of the lenses instead.

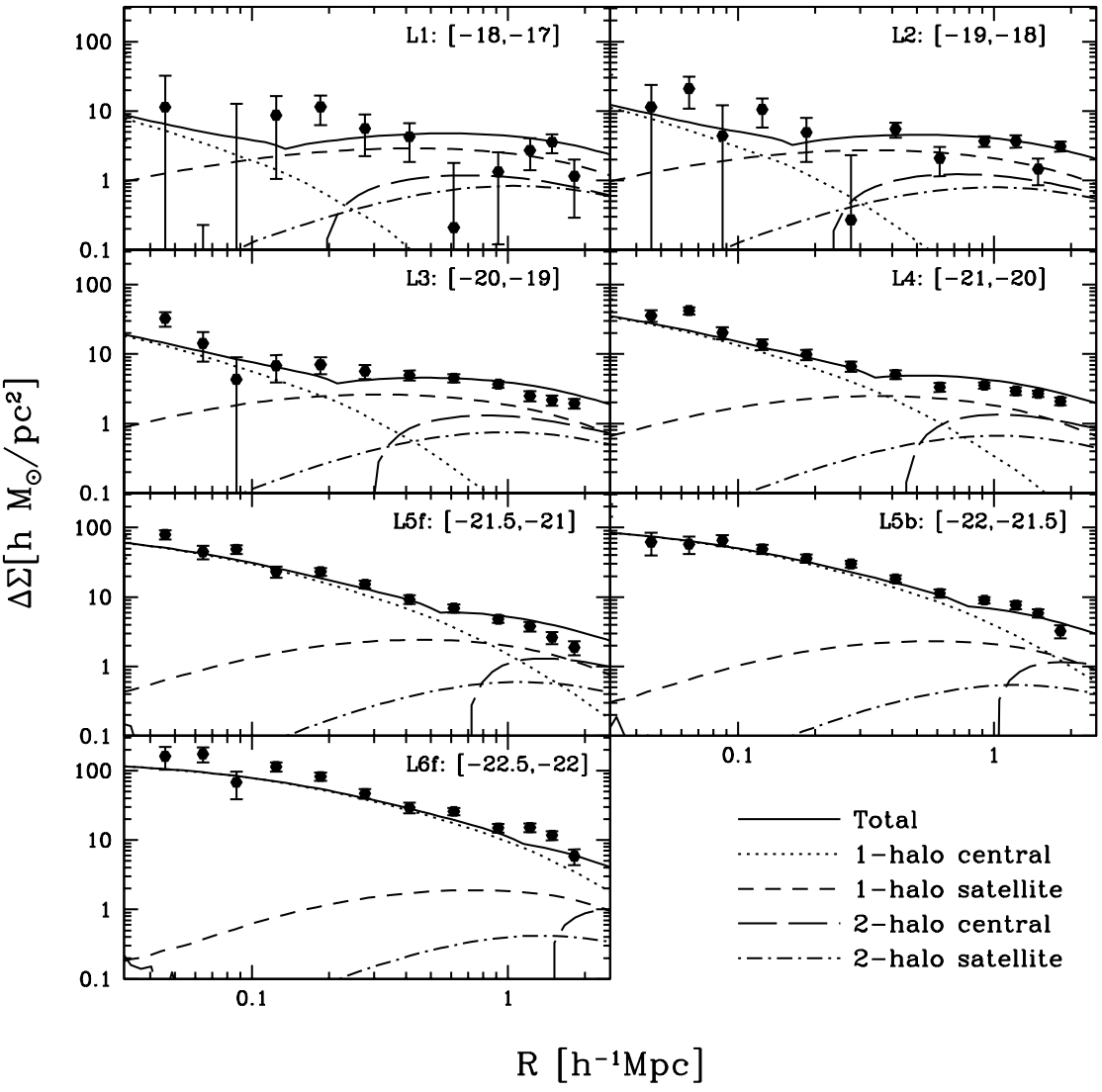

Figure 5.4: The excess surface density $\Delta\Sigma$ as a function of the comoving transverse separation R is plotted for different bins in luminosity of the lens galaxy (see Table 2.2). The solid line represents the total signal as predicted by the model, data points and error bars come from Seljak et al.(2005), see text. The different contributions to the signal are also plotted. The dotted line represents the one halo central term which obviously dominates at the smallest scales in all cases. Note that this term dominates on larger and larger scales when brighter galaxies are considered, reflecting the idea that brighter galaxies live on average in more massive haloes. The dashed line represents the one halo satellite term which is dominant only for faint galaxies and only on intermediate scales (0.1-1 h^{-1}Mpc). The two halo central is plotted with a long dashed line and it becomes relevant on large scales ($R > 1h^{-1}$Mpc). Note that the strong truncation of this term at small scales is due to our implementation of halo exclusion. The two halo satellite term (dotted-dashed line) never dominates but it can contribute up to 20% of the total signal.

a pronouced 'bump' at $R \sim 1h^{-1}$ Mpc, which is absent in the $\Delta\Sigma(R)$ of the brighter galaxies. The reason for this is evident from the middle and right-hand panels of Fig. 5.3, which show the contributions to $\Delta\Sigma(R)$ from the four different terms for the faint ($-16 \geq {}^{0.1}M_r - 5\log h \geq -17$) and bright ($-21 \geq {}^{0.1}M_r - 5\log h \geq -22$) luminosity bins, respectively. In both cases, the one halo central term dominates on small scales. In the case of the faint galaxies, the one halo satellite term dominates over the radial range $0.1h^{-1}$ Mpc $\lesssim R \lesssim 5h^{-1}$ Mpc, and is responsible for the pronounced bump on intermediate scales. In the case of the bright bin, however, the one halo central term dominates all the way out to $R \sim 2h^{-1}$ Mpc. This owes to the fact that bright centrals reside in more massive haloes, which are larger and cause a stronger lensing signal, and due to the fact that the satellite fraction of brighter galaxies is smaller. The fact that the one halo satellite term peaks at intermediate scales, rather than at $R = 0$, owes to the fact that $\Delta\Sigma^{\rm 1h,s}(R)$ reflects a convolution of the host halo mass density profile with the number density distribution of satellite galaxies. On large scales ($R \gtrsim 3h^{-1}$ Mpc), which roughly reflects two times the virial radius of the most massive dark matter haloes, the ESD is dominated by the two halo terms. Note that the faint galaxies, with $-16 \geq {}^{0.1}M_r - 5\log h \geq -17$, have the same large scale ESD as the bright galaxies with $-21 \geq {}^{0.1}M_r - 5\log h \geq -22$, indicating that they have similar values for their bias. This owes to the fact that many of the faint galaxies are satellites which reside in massive haloes. Note also that the two halo central term reveals a fairly abrupt truncation at small radii, which owes to halo-exclusion (see § 5.2.3.4). This truncation also leaves a signature in the total lensing signal, which is more pronounced for the fainter lenses. We caution, however, that the *sharpness* of this feature is partially an artefact due to our approximate implementation of halo-exclusion. Nevertheless, it is clear from Fig. 5.3 that the excess surface densities obtained from g-g lensing measurements contain a wealth of information regarding the galaxy-dark matter connection.

5.2.2 Results for the WMAP3 Cosmology

Using the methodology outlined in chapter 4 and in §5.2.1, we now apply the CLF for the WMAP3 cosmology obtained in §5.1.2.1 to predict the g-g lensing signal for the 7 luminosity bins listed in Table 2.2. For each luminosity bin we compute the ESD profile, $\Delta\Sigma(R)$, at the mean redshift of the sample, i.e., we use the halo mass function, $n(M)$, the halo bias function, $b(M)$, and the non-linear power spectrum, $P^{\rm NL}_{\rm dm}(k)$ that correspond to the mean redshift listed in the third column of Table 2.2. We have verified, though, that computing $\Delta\Sigma(R)$ simply at $z = 0$ instead has a negligible impact on the results.

The results are shown in Fig. 5.4, where the solid dots with errorbars correspond to the SDSS data and the solid lines are the predictions of our best-fit CLF model (whose parameters are listed in Table 5.1). Note that this model fits the data remarkably well, which is quantified by the fact that the reduced χ^2 is 3.1. We emphasize that there are no free parameters here: the ESD has been computed using the CLF that has been constrained using the LF and the clustering data. The good agreement between model and lensing data thus provides independent support for the halo occupation statistics described by our WMAP3 CLF model, in particular for the mass-to-light ratios and satellite fractions, which have an important impact on the lensing signal.

The different curves in each of the panels in Fig. 5.4 show the contribution to the lensing signal due to the four separate terms: $\Delta\Sigma^{1h,c}$ (dotted lines), $\Delta\Sigma^{1h,s}$ (short-dashed lines), $\Delta\Sigma^{2h,c}$ (long-dashed lines), and $\Delta\Sigma^{2h,s}$ (dot-dashed lines). In agreement with the examples shown in Fig. 5.3, the one halo central term becomes increasingly more dominant for more luminous lenses. In fact, in the brightest luminosity bin (L6f) it dominates over the entire radial range probed. In the low-luminosity bins, most of the observed lensing signal at $R \gtrsim 200h^{-1}$ kpc is dominated by the one halo satellite term. The fact that our model accurately fits the data, thus supports the satellite fractions inferred from our CLF model, and shown in the upper right-hand panel of Fig. 5.1.

Both Seljak et al.(2005) and Mandelbaum et al.(2006) did not account for the contributions of the two halo terms in their analyses of the galaxy-galaxy lensing signal. Our model indicates that, although the two halo terms never dominate the total signal, they can contribute as much as 50 percent at large radii ($R \simeq 1h^{-1}$ Mpc). We thus conclude that the two halo terms cannot simply be ignored.

5.2.3 The Impact of the Model Assumptions

Although our computation of the g-g lensing signal does not involve any free parameters (these are already constrained by the clustering data), a number of assumptions are made. In particular, haloes are assumed to be spherical and to follow a NFW density distribution, central galaxies are assumed to reside exactly at the center of their dark matter haloes, and satellite galaxies are assumed to follow a radial number density distribution that has the same shape as the dark matter mass distribution. In addition, we made assumptions regarding the functional form of the CLF. Although most of these simplifications are reasonable, and have support from independent studies, they may have a non-negligible impact on the predictions of the g-g lensing signal. In this section we therefore investigate how strongly our model predictions depend on these oversimplified assumptions.

Some of these dependencies have been investigated together with Li and collaborators and they have been published in Li et al.(2009). Fig. 5.5 summarizes the results. We have examined the impact of our assumption of halo sphericity. For this purpose, we have considered a tri-axial density distribution as an alternative to the fiducial NFW density profile. We have labeled this model as 'TRI' (see Li et al. 2009 for detail). The impact that the triaxial density profile has on the g-g lensing predictions is shown in fig. 5.5 where it can be seen that the fact that realistic dark matter haloes are ellipsoidal, rather than spherical, can be safely ignored (i.e., its impact on the ESD profiles is completely negligible). Moreover, if central galaxies are not located exactly at the center of their dark matter haloes, this may have a non-negligible impact on the lensing signal on small scales ($R \lesssim 0.1h^{-1}$ Mpc). Fortunately, for realistic amplitudes of this offset (van den Bosch et al.2005b), the effect is fairly small and only restricted to the most luminous galaxies. More quantitatively, in haloes with masses $M_h > 10^{13}h^{-1}\mathrm{M}_\odot$ there is evidence to suggest that central galaxies are offset from their halo centers by ~ 3 percent of the virial radius (e.g. van den Bosch et al. 2005 and Berlind et al. 2003 who obtained a similar result by using SPH simulations of galaxy formation). To examine how such an offset impacts on the ESDs, we have considered two models (see Li et al. 2009 for detail). Following van den Bosch et al. (2005) and Yang et al. (2006), we have assumed that central galaxies in haloes

with $M_h > 10^{13}h^{-1}\mathrm{M}_\odot$ are offset from their halo centers by an amount that is drawn from a Gaussian distribution with zero mean. We have used two different values for the dispersion of the distribution: 3% of the virial radius (model 'Dev1') and 6% of the virial radius (model 'Dev2'). The corresponding lensing predictions are shown in the middle column of fig. 5.5 as the dotted (Dev1) and dashed (Dev2) lines. Note that the offsets only affect the lensing signal for the brightest sample (L6f), where a larger offset results in a stronger suppression of the ESD on small scales ($R < 0.1\,h^{-1}$Mpc). For fainter samples, no (significant) differences with respect to the fiducial model are apparent, which owes to the fact that fainter centrals typically reside in haloes with $M_h < 10^{13}h^{-1}\mathrm{M}_\odot$ which do not have an offset (at least in our models).

Furthermore, we have made two additional assumptions: (i) we have neglected the contribution to the ESD due to sub-haloes hosting satellite galaxies, and (ii) we have assumed that luminosity segregation in satellite galaxies can be neglected. As shown in Li et al. (2009), the contribution due to sub-haloes hosting satellite galaxies is confined on the smallest scales of interest here and it accounts for few percent of the signal. Moreover, van den Bosch et al. (2009) have shown that, although present, luminosity segregation in satellite galaxies is negligible.

Below we investigate four additional model dependencies: the scatter in the relation between light and mass, the concentration of dark matter haloes, the radial number density distribution of satellite galaxies, and the approximate treatment of halo exclusion. To that extent we compare our fiducial model (the best-fit CLF model for the WMAP3 cosmology presented above), to models in which we change only one parameter.

5.2.3.1 Scatter in the $L_\mathrm{c} - M$ relation

An important improvement of our halo occupation model over that used by Seljak et al.(2005) and Mandelbaum et al.(2006) is that we allow for scatter in the relation between light and mass. In particular, we model the probability function $\mathcal{P}_c(L|M) = \Phi_c(L|M)$ as a log-normal with a scatter, σ_c, that is assumed to be independent of halo mass. As demonstrated in More et al.(2009b), this assumption is consistent with the kinematics of satellite galaxies, and it is supported by semi-analytical models for galaxy formation. Note, though, that this does not imply that the scatter in $\Phi_c(M|L)$, which is the probability function which actually enters in the computation of the g-g lensing signal, is also constant. In fact, it is not. This is illustrated in Fig. 5.6, which shows, for four luminosity bins, the halo occupation statistics of central galaxies which reflects $\mathcal{P}_c(M|L_1, L_2)$ of our fiducial model. Two trends are evident: more luminous centrals reside, on average, in more massive haloes, and they have a larger scatter in halo masses. As discussed in More, van den Bosch & Cacciato (2009), the fact that the scatter in $\mathcal{P}_\mathrm{c}(M|L)$ increases with luminosity simply owes to the fact that the slope of $L_\mathrm{c}(M)$ becomes shallower with increasing M (see the lower right-hand panels of Figs. 5.1 and 6.3). As is extremely evident from Fig. 5.6, this is a strong effect, with the scatter in $\mathcal{P}_\mathrm{c}(M|L)$ becoming very large at the bright end. Note that this scatter is not dominated by the width of the luminosity bin. Hence, even if one were able to use infinitesimally narrow luminosity bins when stacking lenses, the scatter in $\mathcal{P}_\mathrm{c}(M|L)$ will still be very appreciable.

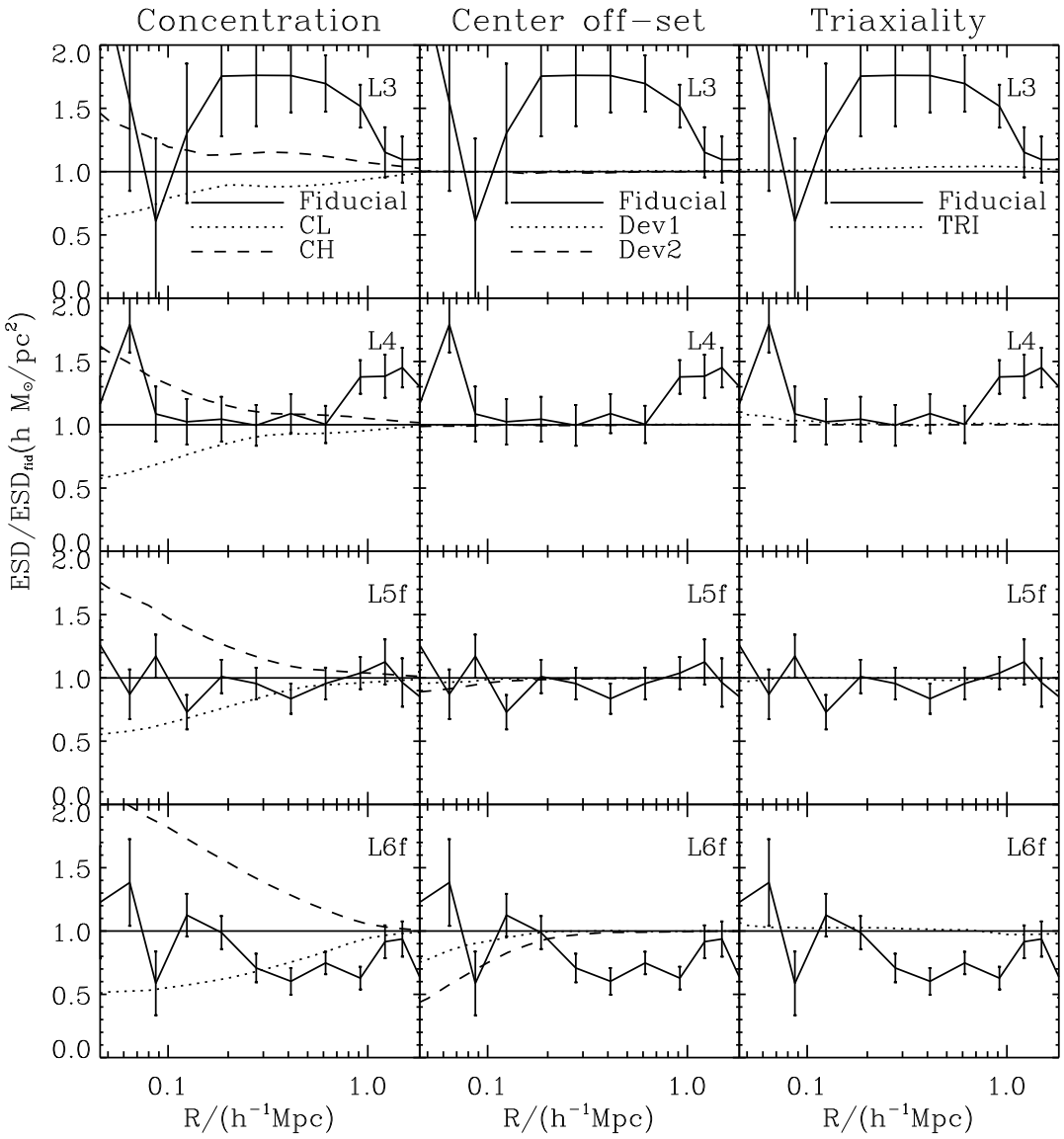

Figure 5.5: Impacts of some model assumptions on the predicted ESD. The effect is shown only for four luminosity bins. For comparison, the observational data are included as data points with error-bars. The left column shows the dependence on halo concentration: the halo concentrations in CL (dotted line) and CH (dashed line) are assumed to be 1/2 and 2 times those in the fiducial model (solid line), respectively. The middle column shows the effects of halo center offset. The dotted line and the dashed line show the results of models Dev1 and Dev2 model, respectively, while the solid line is the fiducial model. The right column shows the effect of assuming triaxial halo density profile. The dotted line shows the result of model TRI, while the solid line again shows the fiducial model. Both the observation and model predicted $\Delta\Sigma(R)$ are normalized by the fiducial prediction.

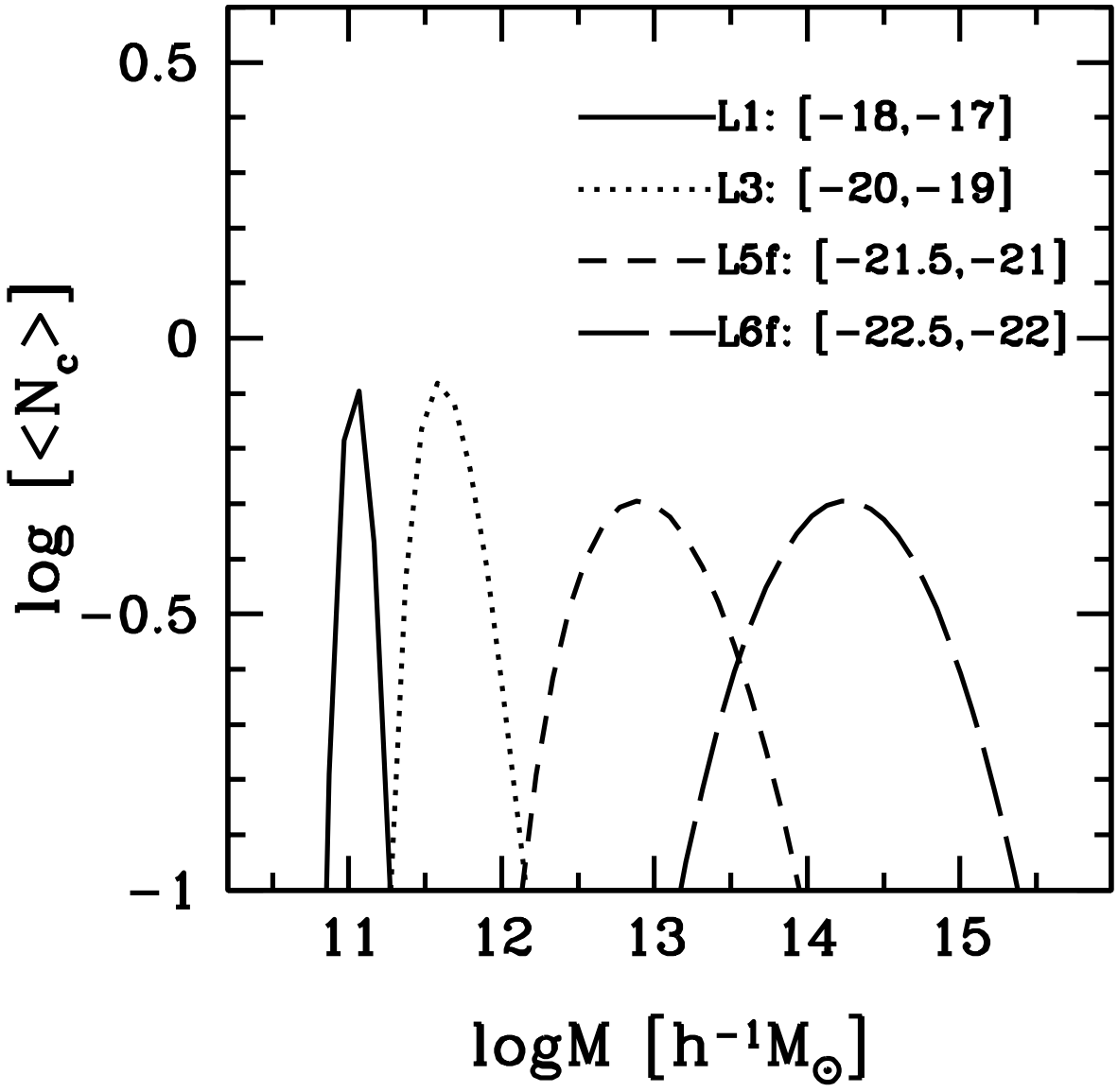

Figure 5.6: Halo occupation distribution for central galaxies. Different lines refer to different luminosity buns as explained in the legend. Note that brighter central galaxies reside, on average, in more massive haloes. Note also that the scatter increases as luminosity increases (see text).

As first shown by Tasitsiomi et al.(2004), scatter in the relation between light and mass can have a very significant impact on the ESDs. This is demonstrated in the upper panels of Fig. 5.7, which show the impact on $\Delta\Sigma(R)$ of changing σ_c by 0.05 compared to our best-fit CLF value of $\sigma_c = 0.14$; all other parameters are kept fixed at their fiducial values (see Table 5.1). Note that these changes in σ_c have a negligible impact on $\Delta\Sigma(R)$ for the low luminosity bins. At the bright end, however, relatively small changes in σ_c have a very significant impact on $\Delta\Sigma(R)$. In particular, increasing the amount of scatter *reduces* the ESD. This behavior owes to the shape of the halo mass function. Increasing the scatter adds both low mass and high mass haloes to the distribution, and the overall change in the average halo mass depends on the slope of the halo mass function. Brighter galaxies live on average in more massive haloes where the halo mass function is steeper. In particular, when the average halo mass is located at the exponential tail of the halo mass function, an increase in the scatter adds many more low mass haloes than massive haloes, causing a drastic shift in the average halo mass towards lower values. On the other hand, fainter galaxies live in less massive haloes, where the slope of the halo mass function is much shallower. Consequently, a change in the scatter does not cause a significant change in the average mass. Note that from the clustering analysis presented in §5.1.2.1, we obtain $\sigma_c = 0.14 \pm 0.01$. This is in good agreement with previous studies: Cooray (2006), using a CLF to model the SDSS r-band LF, obtained $\sigma_c = 0.17^{+0.02}_{-0.01}$. YMB08, using a SDSS galaxy group catalogue, obtained $\sigma_c = 0.13 \pm 0.03$, and More et al.(2009b), using the kinematics of satellite galaxies in the SDSS, find $\sigma_c = 0.16 \pm 0.04$ (all errors are 68% confidence levels).

5.2.3.2 The dark matter halo concentration

The g-g lensing signal on small scales reflects the projected mass distribution of the haloes hosting the lensing galaxies. Therefore, the detailed shape of $\Delta\Sigma(R)$ on small scales is sensitive to the mass distribution of dark matter haloes. In our model, we have assumed that dark matter haloes follow NFW profiles, which are characterized by their concentration parameters, $c_{\rm dm}$. Halo concentrations are known to depend on both halo mass and cosmology, and various analytical models have been developed to describe these dependencies (Navarro et al.1997; Bullock et al.2001; Eke, Navarro & Steinmetz 2001; Neto et al.2007; Macciò et al.2007, 2008). Unfortunately, these models make slightly different predictions for the mass dependence of $c_{\rm dm}$ (mainly due to the fact that the numerical simulations used to calibrate the models covered different limited mass ranges). In Li et al.(2009), we have shown that changing $c_{\rm dm}$ by a factor of two has a very large impact on the ESD profiles (see fig.5.5). However, this is much larger than the typical discrepancies between the different models for $c_{\rm dm}(M)$. The second row of panels in Fig. 5.7 shows $\Delta\Sigma(R)$ obtained for three of these models: the solid lines (labelled MAC) corresponds to our fiducial model for which we have used the $c_{\rm dm}(M)$ relation of Macciò et al.(2007). The dotted lines (labelled BUL) and dashed lines (labelled ENS) correspond to the $c_{\rm dm}(M)$ relations of Bullock et al.(2001) and Eke et al.(2001), respectively. The BUL model predicts halo concentrations that are about 15 percent higher than for the MAC model. The ENS model predicts a $c_{\rm dm}(M)$ that is somewhat shallower than the BUL and MAC models. As evident from Fig. 5.7, though, the results based on these three different models are very similar. We

thus conclude that our results are robust to uncertainties in the relation between halo mass and halo concentration.

5.2.3.3 Number density of satellite galaxies

In our modelling of the g-g lensing signal, we have assumed that the number density distribution of satellite galaxies can be described by a generalised NFW profile, which is parameterized by two free parameters: α and $\mathcal{R}$. In the models presented above, we have assumed that $\alpha = \mathcal{R} = 1$, so that the number density distribution of satellite galaxies has exactly the same shape as the dark matter distribution. Note that there is observational evidence which suggests that satellite galaxies are spatially anti-biased with respect to the dark matter (i.e., their radial distribution is less concentrated than that of the dark matter). This is also supported by numerical simulations, which show that dark matter subhaloes (which are believed to host satellite galaxies) are also spatially anti-biased with respect to the dark matter (e.g., Moore et al.1999, De Lucia 2004).

The panels in the third row of Fig. 5.7 show the impact of changing the concentration of the radial number density distribution of satellite galaxies. In particular, we compare the ESD profiles obtained for our fiducial model ($\mathcal{R} = 1.0$, solid lines) with models in which $\mathcal{R} = 0.5$ (dotted lines) and $\mathcal{R} = 2.0$ (dashed lines). Recall that $\mathcal{R} = c_{\rm dm}/c_{\rm s}$, so that $\mathcal{R} > 1$ ($\mathcal{R} < 1$) corresponds to satellite galaxies being less (more) centrally concentrated than the dark matter. Note that changes in $\mathcal{R}$ have a negligible effect on $\Delta\Sigma(R)$ for the bright luminosity bins. This simply owes to the fact that the ESD of bright lenses is completely dominated by the one halo central term (i.e., the satellite fraction of bright galaxies is very small). For the fainter luminosity bins, however, an increase (decrease) in $\mathcal{R}$ causes a decrease (increase) in $\Delta\Sigma(R)$ on intermediate scales ($0.1h^{-1}$ Mpc $\lesssim R \lesssim 1h^{-1}$ Mpc), which is the scale on which the one halo satellite term dominates. The effect, though, is fairly small (typically smaller than the errorbars on the data points).

The last row of Fig. 5.7 shows the impact of changing the central slope, α, of $n_{\rm s}(r)$. If the number density distribution of satellite galaxies has a central core ($\alpha = 0$), rather than a NFW-like cusp ($\alpha = 1$), it has a similar impact on the lensing signal as assuming a less centrally concentrated $n_{\rm s}(r)$. In fact, the ESD profiles for $(\alpha, \mathcal{R}) = (0.0, 1.0)$ are very similar to those for $(\alpha, \mathcal{R}) = (1.0, 2.0)$. The main conclusion is that our results are not very sensitive to the exact form of $n_{\rm s}(r)$ (see also Yoo et al.2006).

5.2.3.4 Halo exclusion

In the halo model, haloes can not overlap. Note that our analytical model presented in chapter 4 does not account for that explicitely. In this section, we present an approximate method to account for this effect. As many other authors in this field, we will refer to it as *halo exclusion*. We discuss the impact of halo exclusion only for the case of $\Delta\Sigma(R)$. Qualitatively, the same considerations hold for the case of the projected correlation function.

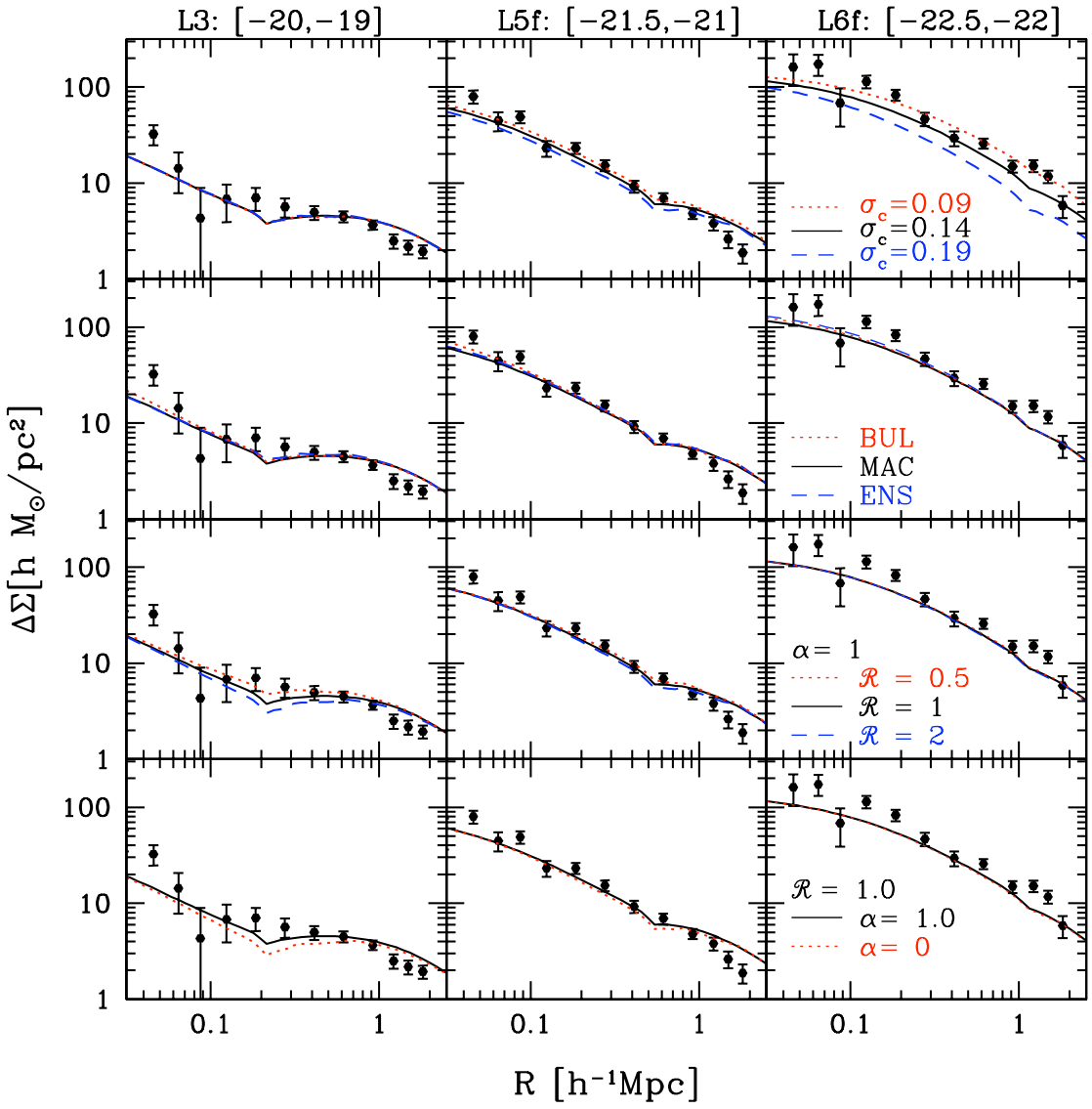

Figure 5.7: The impact of various model parameters on $\Delta\Sigma(R)$. Results are shown for three luminosity bins, as indicated at the top of each column. In each panel the solid line corresponds to our fiducial model (the best-fit CLF model for the WMAP3 cosmology presented in Fig. 5.4), while the dotted and dashed lines correspond to models in which we have only changed one parameter or model ingredient. *Upper panels:* the impact of changes in the parameter σ_c, which describes the amount of scatter in $\Phi_c(L|M)$ (see equation [4.89]). *Second row from the top:* the impact of changes in the halo concentration, $c_{\rm dm}(M)$. In particular, we compare three models for the mass dependence of $c_{\rm dm}$: Macciò et al.(2007; MAC), Bullock et al.(2001; BUL), and Eke et al.(2001; ENS). *Third row from the top:* the impact of changes in $\Re = c_{\rm dm}/c_{\rm s}$, which controls the concentration of the radial number density distribution of satellite galaxies relative to that of the dark matter. *Lower panels:* The impact of changes in α, which specifies the central slope of the radial number density distribution of satellite galaxies. See text for a detailed discussion.

Far from being a detailed treatment, the suggested procedure accounts only for the most relevant effect, i.e. the exclusion of dark matter particles residing in the same host halo of central galaxies. By definition, the two halo terms of the galaxy-dark matter cross correlation, $\xi_{\rm g,dm}(r)$, only considers pairs of galaxies and dark matter particles that reside in different haloes. Since two haloes can not overlap spatially, this implies that the two halo terms given in chapter 4 need to be modified to take account of halo exclusion. The concept of halo exclusion is illustrated in Fig. 5.8 for the two halo central and two halo satellite terms separately. Consider a spherical halo of mass M and radius r_{180} that hosts a central galaxy. It is clear that this central galaxy cannot contribute any signal to the two halo term on scales smaller than r_{180}. Hence, if all central galaxies lived in haloes of a fixed mass M, then there would be no pairs contributing to the 2h,c term for $r < r_{180}$. In reality, though, one needs to account for the fact that centrals occupy haloes that span a range in halo masses, even if the centrals all have the same luminosity. In the case of the satellite galaxies the situation is even more complicated. In particular, since satellite galaxies can reside at the outskirts of dark matter haloes, the two halo satellite term can still have non-zero power at $r \ll r_{180}$. Thus, halo exclusion has less impact on the two halo satellite term than on the two halo central term.

Although the concept of halo exclusion is quite simple, a proper implementation of it in the halo model is extremely tedious numerically. We therefore use only an approximate treatment, which has the advantage that it is straightforward to implement numerically. First of all, we ignore halo exclusion for the two halo satellite term. Since this term is always smaller than the two halo central term, and since halo exclusion is less important for satellites than for centrals, this should not have a significant impact on the results. For the two halo central term we proceed as follows: for each luminosity bin, $[L_1, L_2]$, we simply set to zero the contribution from pairs in the 2h,c term at distances $r < r_{180}(\bar{M})$. Here $\bar{M}$ is the *average* halo mass of the central galaxies

$$\bar{M} = \int_0^\infty \mathcal{P}_{\rm c}(M|L_1, L_2)\, M\, {\rm d}M\,, \tag{5.14}$$

where $\mathcal{P}_{\rm c}(M|L_1, L_2)$ is the probability that a central galaxy with luminosity $L_1 \leq L \leq L_2$ resides in a halo of mass M, and is given by

$$\mathcal{P}_{\rm c}(M|L_1, L_2) = \frac{\langle N_{\rm c}\rangle_M(L_1, L_2) n(M)}{\bar{n}_{\rm c}(L_1, L_2)}\,. \tag{5.15}$$

The corresponding radius, $r_{180}(\bar{M})$, follows from Eq. (3.33).

Although this treatment of halo exclusion is clearly oversimplified, we emphasize that previous attempts to include halo exclusion in the halo model are also approximations (e.g. Magliocchetti & Porciani 2003; Tinker et al.2005; Yoo et al.2006). In a numerical approach, Li et al. (2009) have taken halo exclusion into account. Our oversimplified approach shows a qualitative agreement with the numerical method presented in Li et al. 2009.

In Fig. 5.9 we show the model predictions for the excess surface density in three different luminosity bins. As is evident from Fig. 5.9, halo exclusion only has a mild impact on the overall results.The black lines, labelled HE, show the ESDs obtained from our fiducial model in which halo exclusion

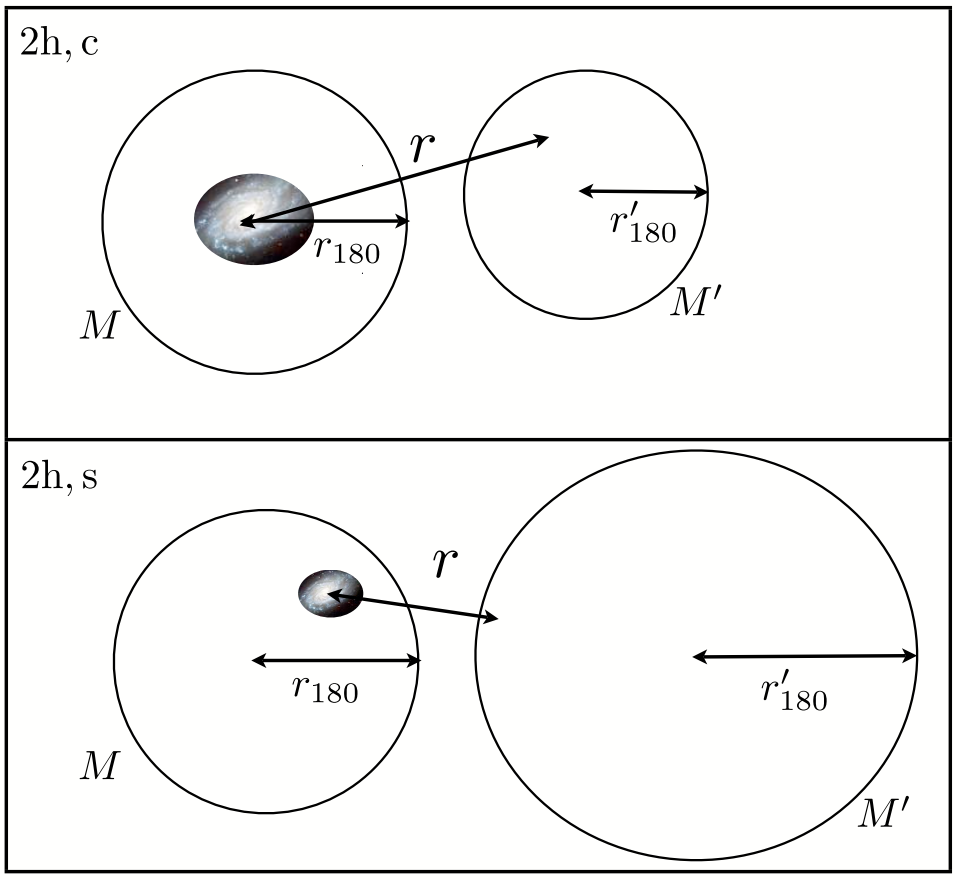

Figure 5.8: Illustration of halo exclusion. The upper panel shows two haloes, of masses M and M', and corresponding radii r_{180} and r'_{180}, respectively. The halo of mass M hosts a central galaxy. Since two haloes cannot overlap, this central galaxy does not contribute any signal to the two halo central term of the galaxy-dark matter cross correlation function on scales $r < r_{180}$. In the case of the two halo satellite term, illustrated in the lower panel, there is still a contribution even on very small scales ($r \ll r_{180}$), simply because satellite galaxies can reside near the edge of the halo.

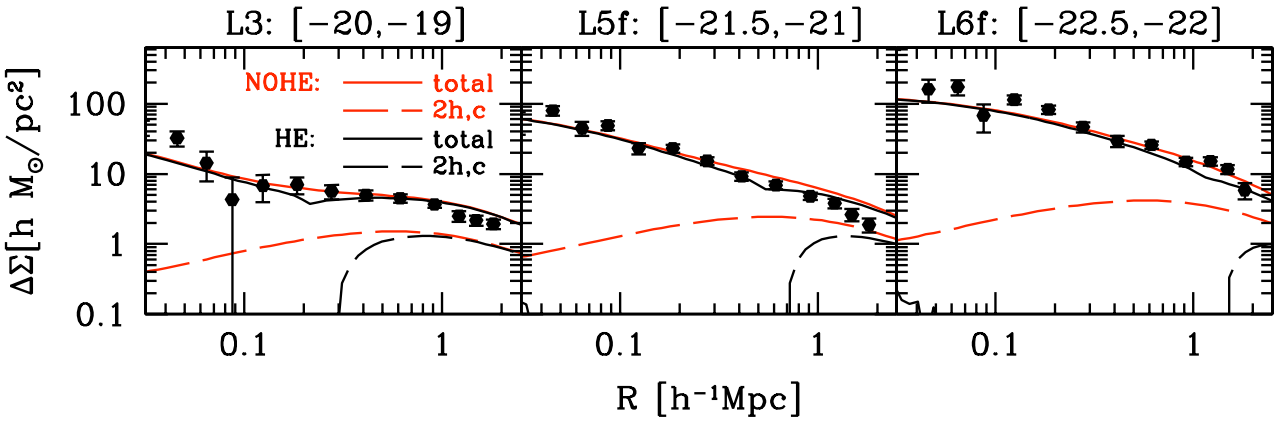

Figure 5.9: The ESD is shown for three luminosity bins. The black lines refer to the fiducial model (HE) and the red lines to the model without halo exclusion (NOHE). The solid lines indicate the total signal, whereas the long dashed lines show the two halo central terms (note that the we ignore halo exclusion for the two halo satellite term). Although the two halo central term is strongly affected by halo exclusion, the impact on the total ESD is only mild. Note that the sharpness of the dip in the black solid lines is (at least partially) an artefact of our oversimplified treatment of halo exclusion, as discussed in the text..

is implemented using the method outlined above. For comparison, the red lines, labelled NOHE, show the results in which we ignore halo exclusion altogether (i.e. in which the two halo terms are simply computed using the equations in chapter 4). The dashed lines show the corresponding two halo central terms, which are clearly suppressed on small scales in the HE model. Since brighter central galaxies are hosted by more massive (and therefore more extended) haloes, the effect of halo exclusion is apparent out to larger radii for brighter galaxies. Note also that the truncation is fairly sharp; this, however, is partially an artefact due to our approximate treatment in which we have only considered the *average* halo mass $\bar{M}(L_1, L_2)$. In reality, the central galaxies live in haloes that span a range in halo masses, and thus a range in sizes. If this were to be taken into account, the truncation would still occur at roughly the same radius, but would be less sharp.

Although halo exclusion clearly has a strong impact on the two halo central term, the impact on the *total* ESD is only modest. This mainly owes to the fact that the total signal on small scales is completely dominated by the one halo terms. Overall, halo exclusion only results in a small reduction of the total ESD on intermediate scales. Due to the arteficial sharpness of the break in the two halo central term, halo exclusion introduces a sharp feature in the total ESD at the radius corresponding to this break. Although the sharpness of this feature is an artefact of our oversimplified treatment of halo exclusion, it does not influence our overall results.

5. GALAXY CLUSTERING & GALAXY-GALAXY LENSING: AN ASTROPHYSICAL PERSPECTIVE

6

Galaxy Clustering & Galaxy-Galaxy Lensing: Constraining Cosmological Parameters

In chapter 5, we have shown that the g-g lensing signal holds a great amount of information about the galaxy-dark matter connection. However, a proper interpretation requires a detailed modeling which accounts for the main complications affecting the g-g lensing measurements. The model based on the CLF is able to incorporate these complications in a straightforward manner. As a result, the g-g lensing signal can be described in great detail with the approach based on the CLF formalism. However, the CLF approach presented in chapter 5 relies on the assumption that the underlying cosmology is known a priori. In the first part of this chapter, we shall explain in detail what is the impact of the underlying cosmology on the properties of the dark matter distribution and on the halo occupation statistics. We thus investigate how a change in the underlying cosmology affects theoretical predictions of galaxy clustering and g-g lensing. In the second part of this chapter, we address the same topic under a different perspective: we use the information encodes in the observed galaxy clustering and g-g lensing to infer the underlying cosmology. In particular, we introduce a novel method to constrain the value of cosmological parameters.

6.1 The Relevance of the Underlying Cosmology

In chapter 3, we have introduced the cold dark matter paradigm and the basic quantities for the description of the dark matter density field. Moreover, in chapter 4, we have described the analytical model which provides theoretical predictions for the clustering of galaxies and the g-g lensing. The analytical expressions provided in that chapter rely on two kinds of information: (i) the properties of the dark matter distribution; (ii) the halo occupation statistics. More specifically, from the

Table 6.1: Cosmological Parameters

	Ω_m	Ω_Λ	Ω_b	h	n	σ_8
WMAP3	0.238	0.762	0.041	0.734	0.951	0.744
WMAP1	0.3	0.7	0.040	0.7	1.0	0.9

first category one requires the knowledge of halo mass function, $n(M)$, halo bias function, $b(M)$, concentration-halo mass relation, $c_{dm}(M)$, and dark matter power spectrum, $P_{dm}(k)$. From the second category, all that is required is the average number of central and satellite galaxies of a given luminosity residing in haloes of a given mass.

In our analytical approach, we describe the dark matter quantities listed above by using fitting functions proposed in the literature and derived by numerical simulations. These fitting functions are clearly cosmology dependent in the sense that the values of the cosmological parameters shape these relations. The way galaxies populate dark matter haloes is also, to some extent, cosmology dependent. In fact, we will see that, in order to reproduce the clustering of galaxies, the halo occupation statistics varies according to the underlying dark matter distribution.

In the first part of this chapter, we consider two flat ΛCDM cosmologies. The first has a matter density $\Omega_m = 0.238$, a baryonic matter density $\Omega_b = 0.041$, a Hubble parameter $h = H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$=0.734, a power-law initial power spectrum with spectral index $n = 0.951$ and a normalization $\sigma_8 = 0.744$. These are the parameters that best-fit the 3-year data release of the Wilkinson Microwave Anisotropy Probe (WMAP, see Spergel et al.2007), and we will refer to this set of cosmological parameters as the WMAP3 cosmology. The second cosmology has $\Omega_m = 0.3$, $\Omega_b = 0.04$, $h = 0.7$, $n = 1.0$ and $\sigma_8 = 0.9$. With strong support from the first year data release of the WMAP mission (see Spergel et al.2003), this choice of parameters has been considered in many previous studies. In what follows we will refer to this second set of parameters as the WMAP1 cosmology. For clarity, the parameters of both cosmologies are listed in Table 6.1.

6.1.1 Comparison with WMAP1 cosmology

As clear from Table 6.1, we are interested in small differences in cosmologies. In fact, WMAP1 and WMAP3 cosmologies differ only in the values of few parameters. In particular, the values of Ω_m and σ_8 differ of $\sim 20\%$, the value of n differ of $\sim 5\%$ and both h and Ω_b are basically unchanged. In what follows, we investigate how these changes affect the relations we are interested in.

Fig. 6.1 shows the halo mass function, $n(M)$, the halo bias function, $b(M)$, and the concentration-mass relation, $c(M)$, for WMAP1 and WMAP3 cosmology. From the uppermost panel, one can note that WMAP1 cosmology predicts that, at the low mass end, the number of haloes in a comoving volume is basically the same of that predicted by WMAP3. However, a significant difference can be noted

at the high mass end: in a WMAP1 universe there are many more massive haloes than in a WMAP3 universe. Qualitatively, this can be understood by noticing that in WMAP1 cosmology, the amount of dark matter is larger (the value of $\Omega_{\rm m}$ is higher) and the dark matter is also more strongly clustered (the value of σ_8 is higher). The middle panel shows the halo bias-halo mass relation. Again, differences between the two cosmologies are relevant only at the high mass end ($M > 10^{12} h^{-1} M_\odot$). Note, however, that the halo bias is plotted on a linear scale and the difference between the two cosmologies are at most of the order of 50%. It is interesting to note that WMAP1 cosmology predicts overall a halo bias which is lower than the one predicted by WMAP3. This can be understood by considering that in WMAP1 the dark matter is highly clustered ($\sigma_8 = 0.9$) and therefore the clustering of massive haloes (which, by definition, are formed in strongly clustered regions) does not differ drastically from the clustering of dark matter. This implies that the value of the halo bias for WMAP1 is lower than in WMAP3. The lowermost panel shows the concentration-halo mass relation. Interestingly, all haloes are more concentrated in a WMAP1 cosmology than in a WMAP3 one. Note, however, that the concentration-halo mass relation has a relatively large scatter (not considered in our model and therefore not plotted). The scatter is approximately of the order of 0.25 dex at all masses. This means that the difference between the relations in the two different cosmologies is less than the intrinsic scatter in the relation. In addition, it is worth noticing that the concentration-halo mass relation enters in the definition of the NFW radial density profile used throughout the thesis. In fig. 6.2, the effect is shown on the Fourier counterpart, $u_{\rm dm}(k|M)$, for three different masses ($M = 5 \cdot 10^{15}, 2.5 \cdot 10^{12}, 10^9 h^{-1} M_\odot$). Fig. 6.2 also shows the dark matter power spectrum as provided by the fitting function of Smith et al. (2003) for the two different cosmologies. Note that the differences between the two cosmologies are quite substantial and they look small only because the plot spans fourteen orders of magnitude.

Once the underlying cosmology is fixed (for example with the WMAP1 parameters), one can look for the CLF parameters which best fit the luminosity function of galaxies, $\Phi(L)$, and the luminosity dependence of the clustering length, $r_0(L)$. In practice, we repeat the same exercise presented in chapter 5. We run a MCMC which explore the 11 CLF parameter space and we look for the set of parameters which provides the best fit. As for the case of WMAP3 (in chapter 5), we use additional constraints from the galaxy group catalogue to relate the average central galaxy luminosity and the number of satellite galaxies to the mass of dark matter haloes. As evident from fig. 6.3, for the WMAP1 cosmology we can obtain a CLF that fits the data almost as well as for WMAP3 cosmology (the reduced χ^2 values are only slightly different; see Table 6.2). Note that the group data (shown in the lower panels) differ from that in Fig. 5.1, even though the group catalogue is the same. This owes to the fact that the halo mass assignments of the groups are cosmology dependent (see Y07 for details). The satellite fractions inferred for this cosmology, shown in the upper right-hand panel of Fig. 6.3, are similar, though slightly higher, than for the WMAP3 cosmology, in excellent agreement with van den Bosch et al.(2007).

Once the best-fit CLF parameters are obtained, one can straightforwardly calculate the corresponding halo occupation statistics, i.e. the average number of central and satellite galaxies in a luminosity bin as a function of the halo mass. Fig. 6.4 shows the halo occupation statistics derived

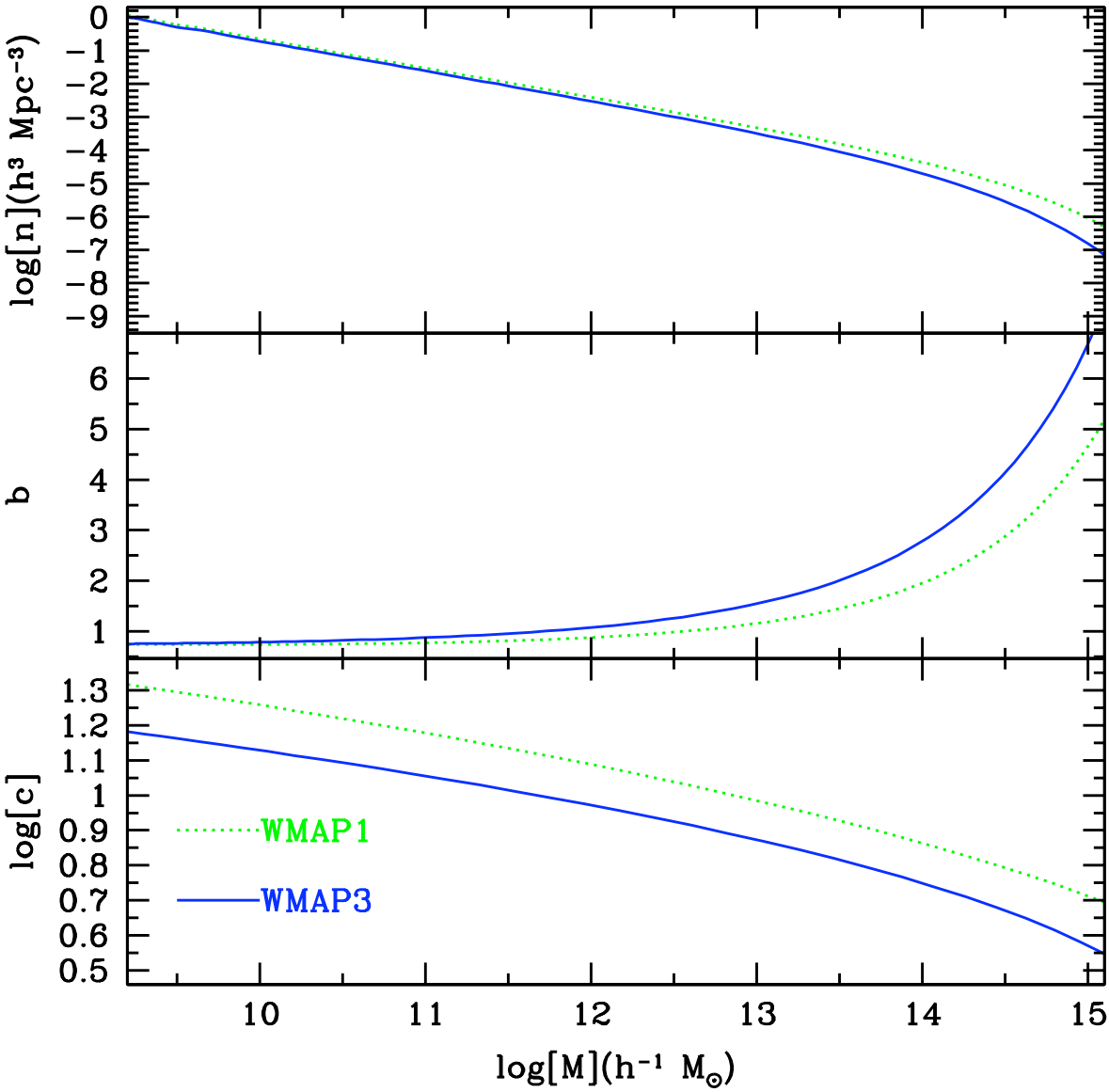

Figure 6.1: *Upper panel.* The halo mass functions for WMAP3 (blue solid line) and WMAP1 (green dotted line) are plotted. Both the relations are calculated using the fitting function of Sheth & Tormen (2002). Note that significant differences can be found only at the massive end. *Middle panel.* The halo bias functions for WMAP3 (blue solid line) and WMAP1 (green dotted line) are plotted. Both the relations are calculated using the fitting function of Sheth & Tormen (2002). Note that the predictions for the two cosmologies may differ up to $\sim$ 40% at the high mass end. *Lower panel.* The halo concentration-halo mass relation for WMAP3 (blue solid line) and WMAP1 (green dotted line) is plotted. Both the relations are calculated using the fitting function of Macciò et al. (2007). Note that the WMAP1 concentrations are systematically above the ones derived for WMAP3. However, the difference is within the intrinsic scatter (not plotted here).

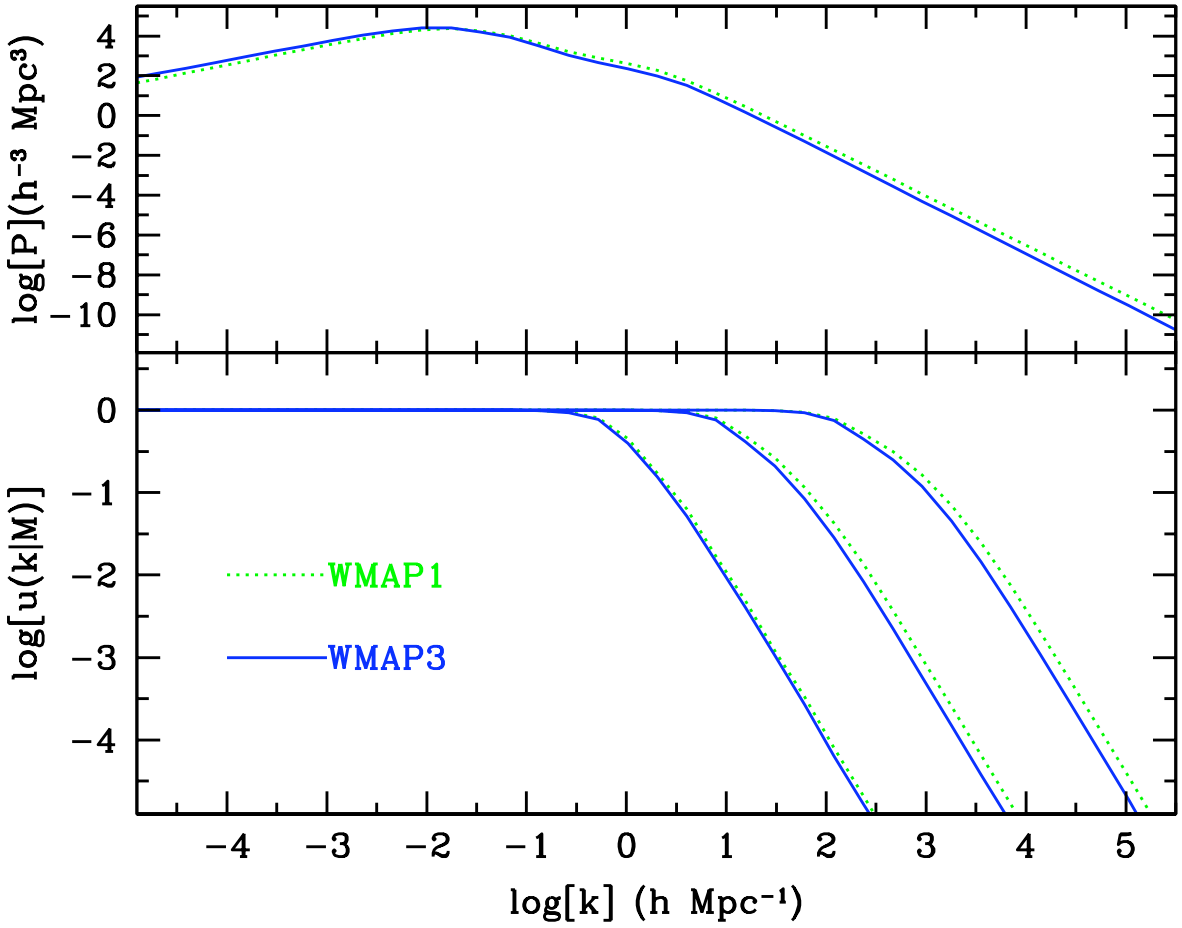

Figure 6.2: *Upper panel.* The dark matter power spectrum for both WMAP3 (blue solid line) and WMAP1 (green dotted line) is plotted. Both relations are calculated using the fitting function proposed by Smith et al. (2003). Note that the differences between the two cosmologies appear small only because the plot covers fourteen orders of magnitude. *Lower panel.* The normalized Fourier transform of the dark matter halo profile is plotted for WMAP3 (blue solid line) and WMAP1 (green dotted line). The different pairs of $u(k|M)$ refers to different halo masses (see text). Note that the halo concentration enters in the definition of $u(k|M)$ and it is indeed responsible for the differences between the predictions based on the two different cosmologies.

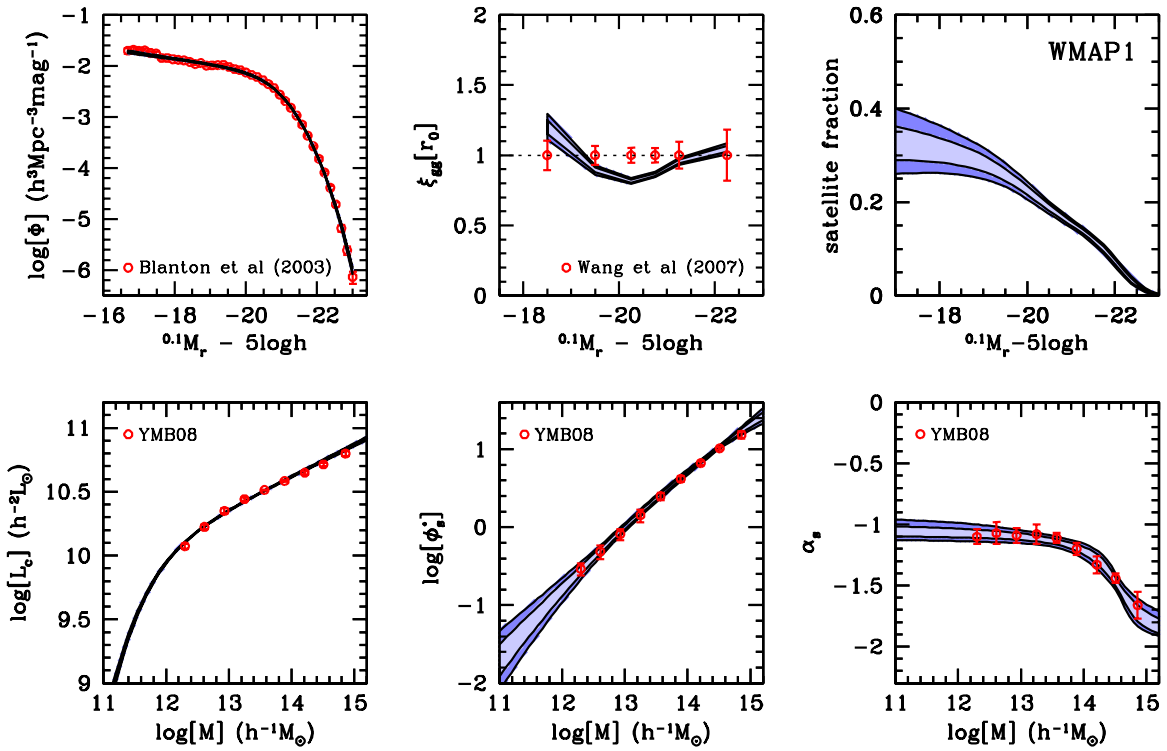

Figure 6.3: *Upper row, left and central panels*. The luminosity function of galaxies and the luminosity dependence of the galaxy correlation length are plotted. Data come from the analysis of Blanton et al.(2003a) and Wang et al.2007. The blue contours indicate the 68 and 95 percent confidence level obtained from the MCMC based on WMAP1. The agreement is extremely accurate for the luminosity function whereas is modest for the correlation length. *Lower row, three panels*. The additional information coming from the group catalogue of YMB08 is plotted together with the corresponding 68 and 95 percent confidence level derived with the MCMC based on WMAP1. In particular, the halo mass dependence of the central galaxy luminosity, the satellite conditional luminosity function normalization ϕ_s^* and the the exponent α_s are shown in the left, central and middle panel, respectively. *Upper row, right panel*. The 68 and 95 percent confidence levels of the satellite fraction, f_s, obtained from the CLF (see eq. [5.10]).

Table 6.2: Best-fit CLF parameters obtained from SDSS clustering analysis

	WMAP3	WMAP1
$\chi^2_{\rm red}$	1.42	1.70
$\log L_0$	9.935 ±0.043	9.915 ±0.052
$\log M_1$	11.07 ±0.11	11.16 ±0.16
γ_1	3.273 ±0.575	3.336 ±1.012
γ_2	0.255 ±0.008	0.248 ±0.007
a_1	0.501 ±0.069	0.484 ±0.057
a_2	2.106 ±1.437	2.888 ±1.943
$\log M_2$	14.28 ±0.16	14.54 ±0.14
b_0	-0.766 ±0.146	-0.854 ±0.159
b_1	1.008 ±0.197	0.906 ±0.187
b_2	-0.094 ±0.065	-0.062 ±0.054
$\sigma_{\rm c}$	0.143 ±0.005	0.140 ±0.005

The best-fit CLF parameters and the corresponding 95% confidence levels obtained from the MCMC analysis for the WMAP3 and WMAP1 cosmologies and the value of the corresponding reduced χ^2. Masses and luminosities are in h^{-1} $M_\odot$ and h^{-2} $L_\odot$, respectively. Note that, for both cosmologies, all the parameters are tightly constrained except γ_1, a_2 and b_2.

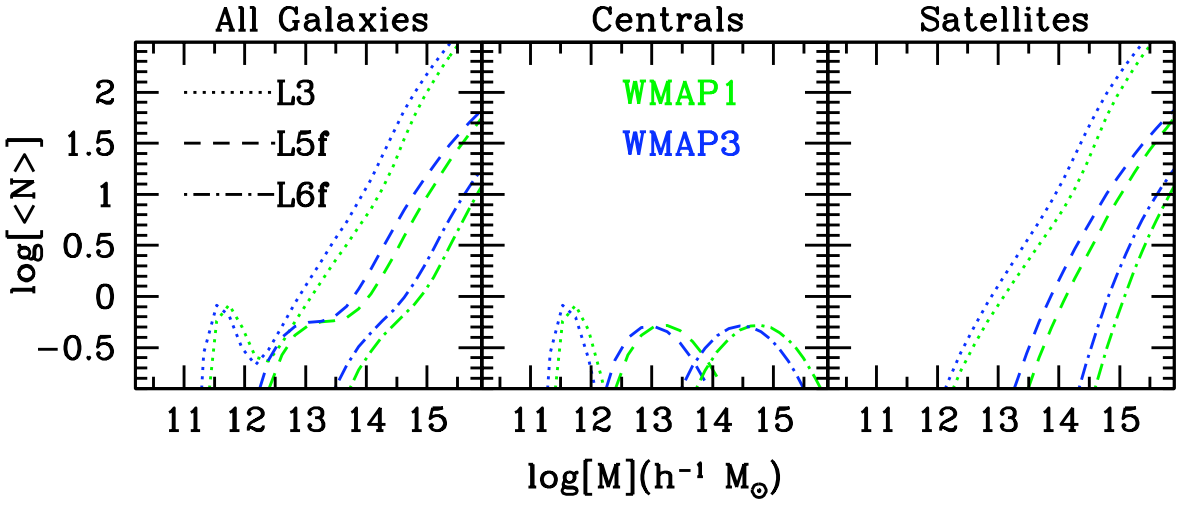

Figure 6.4: The halo occupation statistics (HOS) of the best fit model for WMAP1 and WMAP3. The blue lines refer to WMAP3 cosmology, whereas the green line refers to WMAP1. Only the HOS for three luminosity bins is displayed. The left panel shows the total number of galaxies in given luminosity bins as a function of the halo mass. The middle and right panel show respectively the number of central and satellite galaxies as a function of the halo mass.

from the best fit CLF parameters based on WMAP3 cosmology (blue lines) and based on WMAP1 cosmology (green lines). Halo occupation statistics for different luminosity bins are indicated with different line types (see Table 2.2 for definition of the labels). Recall that $L3 < L5\text{f} < L6\text{f}$. Regardless of cosmology, the average number of central galaxies in a luminosity bin has a clear peak at a given halo mass and a scatter around it. This is the result of the parametrization of the CLF for central galaxies. Note that the scatter is bigger for brighter luminosity bins. This simply owes to the fact that the slope of $\langle \log(L_c) \rangle - \log(M)$ becomes shallower with increasing L_c, as qualitatively illustrated in fig. 6.5. In addition, brighter central galaxies have higher average halo mass. This is in concordance with the idea that brighter galaxies live on average in more massive haloes. Moreover, note that at a given luminosity bin, the results from WMAP1 cosmology peak at higher halo masses. In other words, central galaxies of the same luminosity are hosted in more massive haloes in WMAP1 than in WMAP3. For the halo occupation of satellite galaxies, one can notice that, independently of cosmology, relatively bright galaxies are distributed as a power law, whereas faint galaxies have a more structured behavior as a function of the halo mass. For what concerns the cosmology dependence, we can notice the same trend as for the central galaxies. At a given luminosity and at a given number of satellite galaxies, the halo masses for WMAP1 results are higher than for WMAP3.

Despite all the differences described above (and in particular the relatively large differences in halo mass function and halo bias), both WMAP1 and WMAP3 cosmologies allow an (almost) equally good fit to the data. This clearly illustrates that the abundance and clustering properties of galaxies allow a fair amount of freedom in cosmological parameters. However, as demonstrated in van den

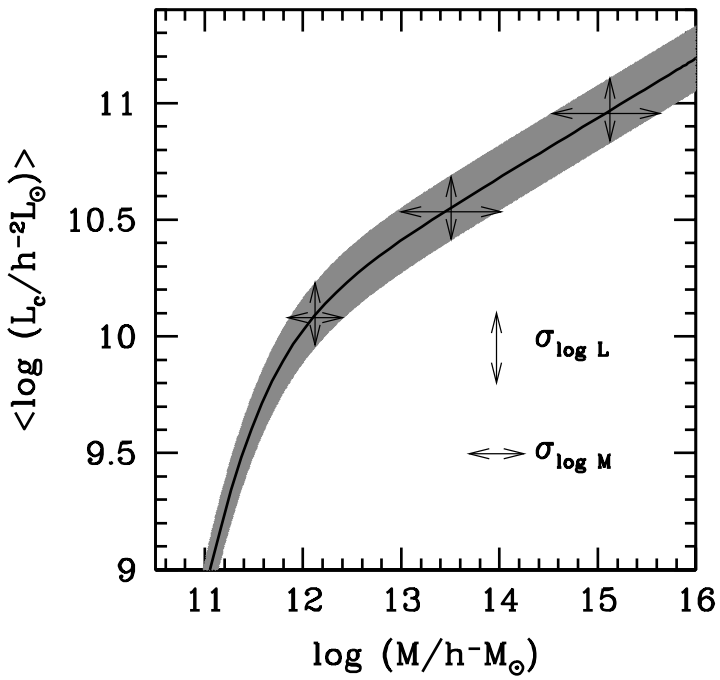

Figure 6.5: The solid black line indicates the mean of the $\langle \log(L_c) \rangle - \log(M)$ relation, while the gray scale region reflects the scatter. Note that the scatter, $\sigma_{\log L} = \sigma_c$ (indicated by vertical arrows) is taken to be constant with halo mass. Note, though, that the scatter in the horizontal direction (indicated by horizontal arrows) increases with increasing L_c; this is simply due to the fact that the slope of the $\langle \log(L_c) \rangle - \log(M)$ relation becomes shallower with increasing halo mass.

Bosch, Mo & Yang (2003), the best-fit CLFs for different cosmologies predict different mass-to-light ratios as function of halo mass. This is evident from Fig. 6.6, which shows the mass-to-light ratios $M/\langle L_{19.5}\rangle_M$ as function of halo mass inferred from our CLF MCMCs for the WMAP1 and WMAP3 cosmologies. Here $\langle L_{19.5}\rangle_M$ is the average, total luminosity of all galaxies with ${}^{0.1}M_r - 5\log h \leq -19.5$ that reside in a halo of mass M, which follows from the CLF according to

$$\langle L_{19.5}\rangle_M = \int_{L_{\rm min}}^{\infty} \Phi(L|M)\, L\, {\rm d}L\,, \tag{6.1}$$

with $L_{\rm min}$ the luminosity corresponding to a magnitude ${}^{0.1}M_r - 5\log h = -19.5$. Clearly, the mass-to-light ratios inferred for the WMAP1 cosmology are significantly higher than for the WMAP3 cosmology (see also van den Bosch et al.2007, where a similar result was obtained using data from the 2dFGRS). Hence, the abundance and clustering properties of galaxies can be used to constrain cosmological parameters, as long as one has independent constraints on the mass-to-light ratios of dark matter haloes. This is exactly what is provided by g-g lensing. It is to be expected that our WMAP3 and WMAP1 CLFs will predict significantly different ESD profiles, thus allowing us to discriminate between these two cosmologies.

Fig. 6.7 shows the 95 percent confidence levels for $\Delta\Sigma(R)$ obtained from our CLF MCMCs for both the WMAP3 (blue) and WMAP1 (green) cosmologies. Indeed, as anticipated, for the WMAP1 cosmology we obtain excess surface densities that are significantly higher than for the WMAP3 cosmology, in agreement with the higher mass-to-light ratios (cf. Fig. 6.6). A comparison with the SDSS data clearly favors the WMAP3 cosmology over the WMAP1 cosmology. In fact, for the latter our best-fit CLF model yields a reduced χ^2 of 29.5, much larger than for the WMAP3 cosmology ($\chi^2_{\rm red} = 3.1$). Note that the cosmological parameters for these two cosmologies are very similar: $\Omega_{\rm m}$ and σ_8 differ only by ~ 20 percent (in addition to a ~ 5 percent difference in n). Yet, we can very significantly favor one cosmology over the other. This indicates that the combination of clustering data and g-g lensing data can be used to put tight constraints on cosmological parameters. A detailed analysis along these lines is deferred to the next section. Note, additionally, that the most recent results from the WMAP mission (hereafter WMAP5, Dunkley et al. 2009) favor a cosmological model with $(\Omega_{\rm m}, \sigma_8) = (0.258, 0.796)$. Since these two cosmological parameters are those mostly responsible for the amplitude of the g-g lensing predictions and since their values lie in between those of the WMAP1 and WMAP3 cosmologies, we expect the lensing predictions also to lie in between those corresponding to WMAP1 and WMAP3.

We can conclude that galaxy clustering and galaxy-galaxy lensing probe the galaxy-dark matter connection in complementary ways. Since the clustering of dark matter haloes depends on cosmology, the halo occupation statistics inferred from the observed clustering properties of galaxies are degenerate with the adopted cosmology. Consequently, different cosmologies imply different mass-to-light ratios for dark matter haloes. Galaxy-galaxy lensing, on the other hand, yields *direct* constraints on the actual mass-to-light ratios of dark matter haloes. Combined, clustering and lensing therefore offer the opportunity to constrain cosmological parameters.

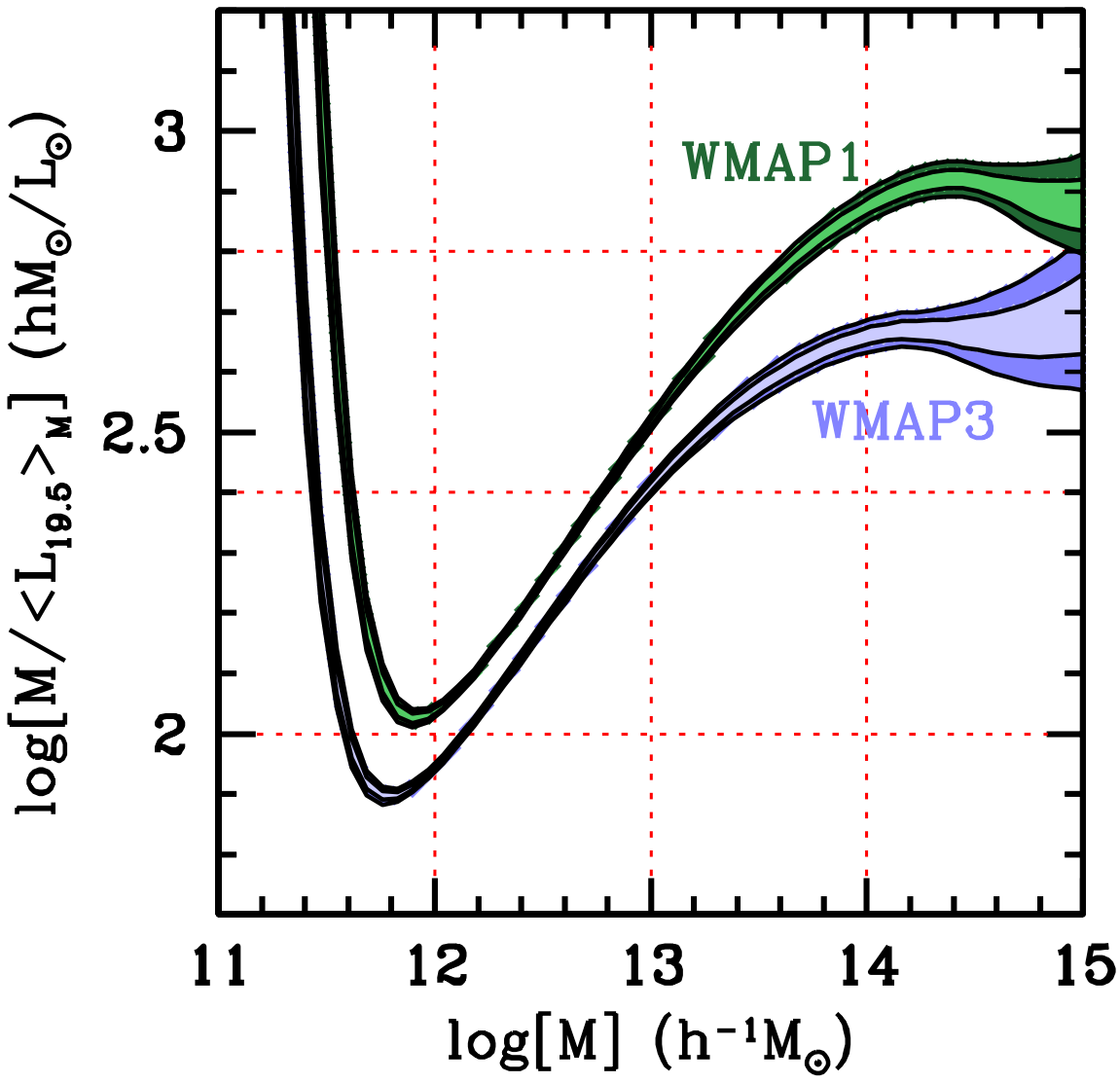

Figure 6.6: The 68 and 95 percent confidence levels for the mass-to-light ratios, $M/\langle L_{19.5}\rangle_M$, obtained from the CLF MCMCs for the WMAP3 and WMAP1 cosmologies.

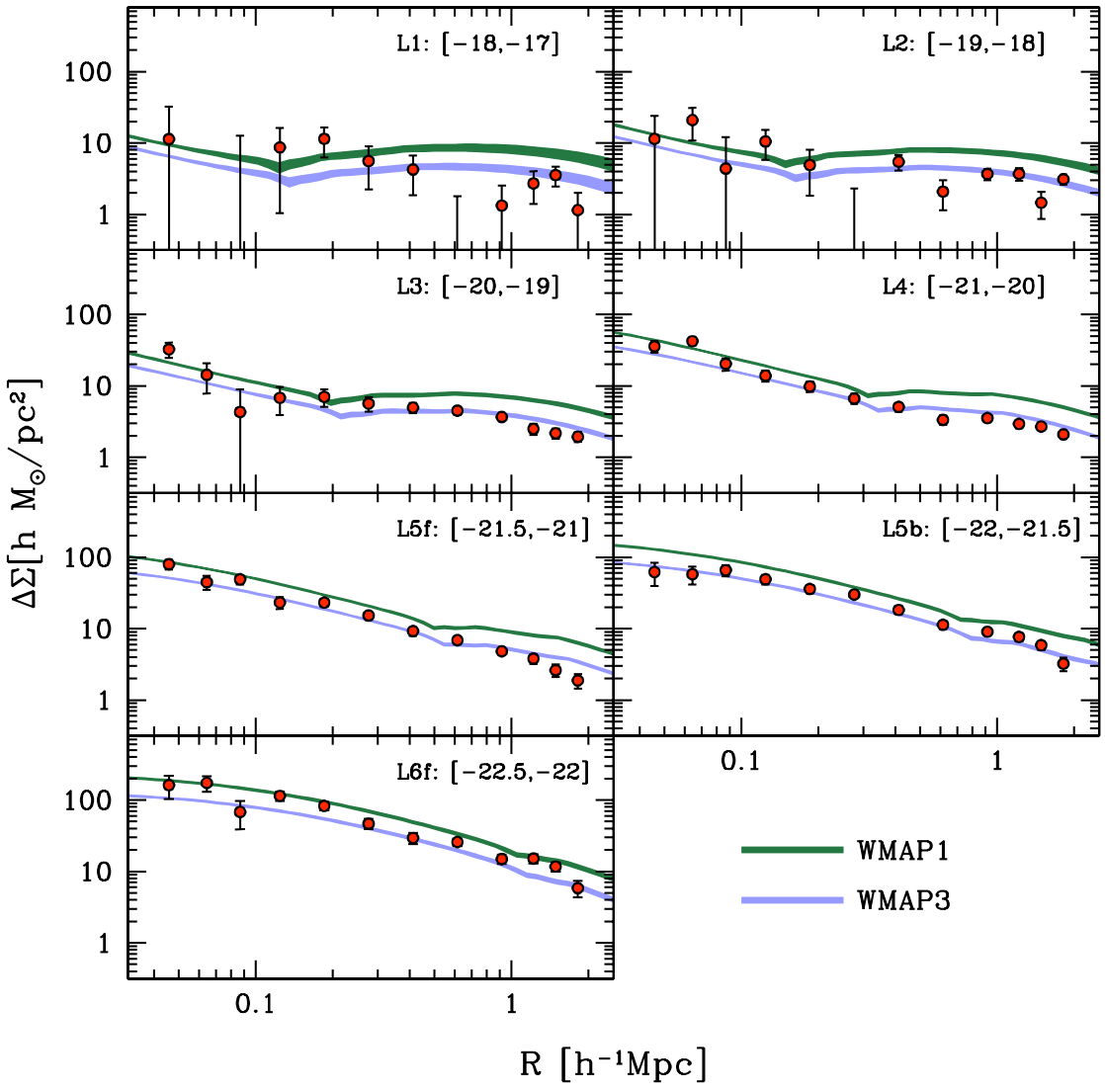

Figure 6.7: The predictions for the lensing signal, $\Delta\Sigma(R)$, are shown for two different sets of cosmological parameters (WMAP1 and WMAP3, see text). The green (blue) shaded area corresponds to the 95% confidence level of the WMAP1 (WMAP3) model. Note that, although the cosmological parameters of these two cosmologies only differ by < 20 percent (see Table 6.1), the ESD predictions are very different, and can easily be discriminated.

6.2 A Novel Method to Constrain Cosmological Parameters

As extensively illustrated in chapter 1, in the last decade many different approaches have been used to constrain the values of the cosmological parameters which identify a specific cosmological model. Among them, the analysis of the temperature fluctuations in the cosmic microwave background (CMB) has certainly been incredibly productive. Non-CMB experiments have also become extremely useful in what they are crucially needed to break the parameter degeneracies intrinsic in CMB analyses (e.g. Eisenstein, Hu & Tegmark 1999; Efstathiou & Bond 1999; Spergel et al 2007). Note, however, that galaxy clustering analyses aiming to constrain cosmological models have focused only on large scales due to the lack of confidence on the knowledge of the non-linear clustering and galaxy biasing (see e.g. Wang et al. 2002; Tegmark et al. 2004; Pope et al. 2004). So far, only a few studies have been presented in which the properties of galaxies on small scales are used to infer information on the underlying cosmology (see Seljak et al. 2005; Yoo et al. 2006 and references therein).

In this section, we present a joint analysis of galaxy clustering and g-g lensing which aims to give indications about the underlying cosmology. This approach makes use of the information encoded in the purely non-linear regime of galaxy distribution. It is motivated by the fact that g-g lensing is extremely sensitive to the underlying cosmology (as demonstrated in the previous section). We investigate the feasibility of such a joint analysis to constrain the values of $\Omega_{\rm m}$ and σ_8 because these are the two cosmological parameters that mostly influence galaxy clustering and g-g lensing. We also investigate all the systematics involved in such a joint analysis and we conclude that combining galaxy clustering and g-g lensing is indeed a novel and competitive method to constrain the value of cosmological parameters.

In chapter 5 and in the previous section, we have shown that given a set of cosmological parameters one can look for the CLF parameters which best fit the data. In those cases, the data were the galaxy luminosity function, $\Phi(L)$ and the clustering length as a function of luminosity, $r_0(L)$. Furthermore, additional constraints from the galaxy group catalogue of Yang et al. (2007) have been used. We have seen how the best fit based on WMAP1 set of cosmological parameters was not able to reproduce g-g lensing data, whereas the best fit based on WMAP3 cosmology was successful in reproducing g-g lensing data without any additional tuning. In this section, we repeat the analysis which looks for the CLF parameters which best fit the data but we now include the g-g lensing data as constraints. Moreover, we will not simply use $r_0(L)$ but the projected correlation function, $w_p(R)$, for galaxies selected according to their luminosity and in a range of scales which goes from ~ 0.1 to $\sim 20h^{-1}Mpc$ (see chapter 2 for a detailed description of the data).

It is worth stressing that the new set of data used to constrain the CLF parameters is much larger than the previous one (used in chapter 5 and in the previous section). In fact, as described in chapter 2, the galaxy luminosity function is sampled in 41 magnitudes, the correlation length in 6, and the data points from the galaxy group catalogue are 27. This gives a total number of 74 constraints. In the new approach, we still have the 41 data points for the galaxy luminosity function and the 27 for the galaxy

group catalogue but in addition we have 72 data points[1] for the excess surface density, $\Delta\Sigma$, and 72 data points for the projected correlation function, w_p. This sums up to a total of 212 data points. This gives already the idea of the power of the new approach: the model must reproduce a large number of data which hold an enormous amount of information about the galaxy-dark matter connection.

6.2.1 The best fit model

In order to find the CLF parameters which best fit the complete set of data, we run an algorithm which, for any given cosmology, looks for the set of parameters which maximizes the likelihood (i.e. minimizes the χ^2). Here $\chi^2 = \chi^2_\Phi + \chi^2_{w_p} + \chi^2_{\Delta\Sigma} + \chi^2_{\rm GC}$ with χ^2_Φ and $\chi^2_{\rm GC}$ already defined in chapter 5 and

$$\chi^2_{w_p} = \sum_{j=1}^{6} \sum_{i=1}^{12} \left[\frac{w_{p,j}(R_i) - \hat{w}_{p,j}(R_i)}{\Delta\hat{w}_{p,j}(R_i)} \right]^2 , \tag{6.2}$$

and

$$\chi^2_{\Delta\Sigma} = \sum_{j=1}^{6} \sum_{i=1}^{12} \left[\frac{\Delta\Sigma_j(R_i) - \Delta\hat{\Sigma}_j(R_i)}{\Delta\Delta\Sigma_j(R_i)} \right]^2 . \tag{6.3}$$

Here $\hat{\cdot}$ indicates an observed quantity, the subscripts 'w_p' and '$\Delta\Sigma$' refer to the 2-point galaxy projected correlation function and the excess surface density, respectively. Note also that the index 'j' runs over the six luminosity bins, whereas 'i' runs over the twelve spatial points per luminosity bin.

The minimization algorithm searches for the combination of the 11 CLF parameters which gives the best agreement between theoretical predictions and observational data. Such a χ^2-minimization is performed by assuming an underlying cosmology. This means that a set of cosmological parameters is fixed at the beginning of the minimization procedure and the model with the minimum χ^2 is found by the algorithm. One can perform this minimization assuming different underlying cosmologies (meaning different sets of cosmological parameters) and then compare the lowest χ^2 obtained for any given cosmology. According to the result of chapter 5, one expects that some cosmological models will allow a better fit than others. This means that the joint analysis of clustering and g-g lensing will support some cosmological models above others. It is worth stressing that in our approach a cosmological model is defined by the set of 5 cosmological parameters $(\Omega_{\rm m}, \sigma_8, n, h, \Omega_{\rm b})$[2]. However, we expect $\Omega_{\rm m}$ and σ_8 to be the two cosmological parameters that affect galaxy clustering and g-g lensing the most. For this reason, in this section we consider only changes in these two parameters. A detailed analysis which explores the complete cosmological parameter space is deferred to § 6.4.

In what follows, we describe the results of the χ^2-minimization procedure. In particular, we focus on discussing which combination ($\Omega_{\rm m}$, σ_8) is favored by the data. As reference we use the results from the analysis of the temperature fluctuations in the CMB performed by the WMAP team. In particular, we compare our results with the confidence levels on the values of $\Omega_{\rm m}$ and σ_8 suggested

[1] There are 12 data points for a given luminosity bin times six luminosity bins (see chapter 2 for a detailed description).
[2] Since we assume a flat geometry, the value of Ω_Λ is fixed by $\Omega_\Lambda = 1 - \Omega_{\rm m}$.

by the fifth year data release (Dunkley et al. 2009). Fig. 6.8 shows the 68%, 95% and 99% confidence levels (three tones of blue) on the values of Ω_m and σ_8 derived from our minimization algorithm for the fiducial model. In our model, the pair $(\Omega_m, \sigma_8) = (0.225, 0.72)$ is the one which provides the best fit to the data ($\chi_{red} = 3.41$). Note that our best fit value for Ω_m agrees with the one suggested by WMAP5 on one-sigma level. The value of σ_8 is in agreement with WMAP5 on a two-sigma level. In fig. 6.8, we also plot the value of the best-fit models suggested by the WMAP1 and WMAP3, (0.3, 0.9) and (0.27, 0.74) respectively. Note that, as expected from the result in chapter 5, our results are closer to that suggested by WMAP3 than that of WMAP1. A major remark is required here. The analysis of the WMAP team consists of a MCMC over all cosmological parameters. In our approach, we have explored only the parameter sub-space (Ω_m, σ_8) and we have kept fixed the values of all other cosmological parameters. This surely limits the degree of freedom of our model and it is partly responsible for the extremely small confidence region. It is to be expected that a full MCMC over all the cosmological parameters would result in a rather larger confidence region. However, note that all the cosmological parameters which have been kept fixed are expected to have a minor impact on the galaxy clustering and g-g lensing, this means that the (Ω_m, σ_8) which allows our model to best fit the data is not expected to change drastically.

We emphasize here that this result is the clear proof that a joint analysis of clustering and g-g lensing may be used as a new method to constrain cosmological parameters. Note the complementarity of the method illustrated here with the one on which WMAP analysis is based. In fact, the analysis of the CMB consists of the study of the temperature fluctuations as an imprint from the cosmological contents in the early stage of the universe ($z \sim 1000$). On the other hand, our method is based on the analysis of the statistical properties of galaxies observed with SDSS ($z \sim 0.1$). This means that the information we are using is based on the latest evolutionary stages of the universe. The fact that these two methods based on two different fields of astrophysics agree remarkably well on the values of the cosmological parameters can be seen as great success of the concordance cosmological model.

Once we have determined the pair (Ω_m, σ_8) which allows the model predictions to best fit the data, we construct a MCMC for that given cosmology. This allows us to study spread around the best fit model allowed by the given cosmology. Technically speaking, we construct a chain of 5 million models. Each element of the chain is a model consisting of 11 parameters. At any point in the chain we generate a new trial model by drawing the shifts in the 11 free parameters from 11 independent Gaussian distributions centered on the current value of the corresponding model parameter. The chain is thinned by a factor of 2500 to remove the correlations between neighboring models. The end result is a MCMC of 2000 independent models properly sampling the posterior distributions of the 11 free parameters. The probability of accepting the trial model is

$$P_{accept} = \begin{cases} 1.0 & \text{if } \chi^2_{new} < \chi^2_{old} \\ \exp[-(\chi^2_{new} - \chi^2_{old})] & \text{if } \chi^2_{new} \geqslant \chi^2_{old} \end{cases} \tag{6.4}$$

where the χ^2 is defined at the beginning of this section.

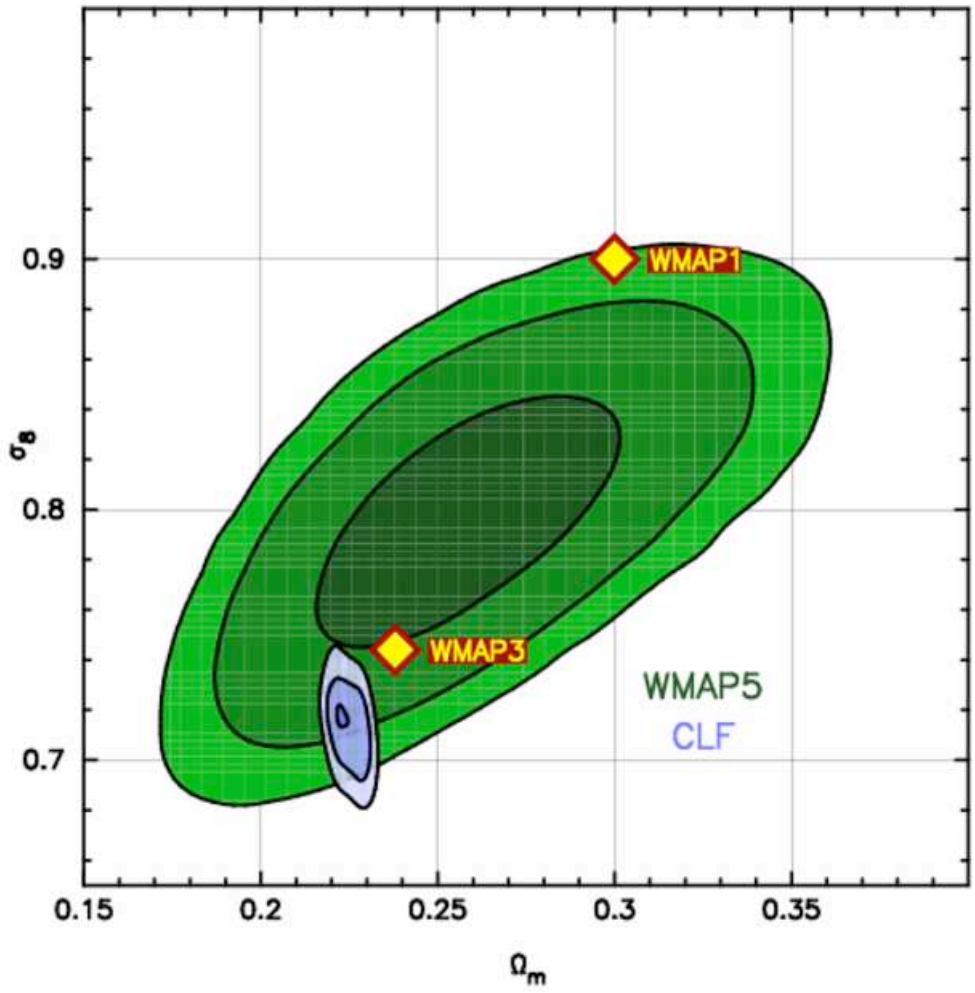

Figure 6.8: The 68%, 95% and 99% confidence levels on the values of Ω_m and σ_8 derived by WMAP5 (three tones of green) and our joint analysis of galaxy clustering and g-g lensing (three tones of blue). Note that the best fit value from WMAP1 and WMAP3 are also shown with yellow diamonds.

Using the output of the MCMC and the best fit model obtained via the minimization routine, we discuss the quality of the fit to the data in the case of the best-fit cosmology $(\Omega_{\rm m}, \sigma_8) = (0.225, 0.72)$. In particular, we show how the model is able to reproduce the excess surface density, $\Delta\Sigma$, the projected correlation function, w_p, the galaxy luminosity function, Φ, and the *"data"* from the galaxy group catalogue. In order to perform this comparison, we plot the theoretical predictions of the best fit model together with the data and with the 68%, 95% and 99% confidence levels derived from the MCMC for the best fit cosmology. Fig. 6.9 show the excess surface density, $\Delta\Sigma$, as a function of the comoving separation in the sky, R, for six luminosity bins. The data are the same presented in chapter 2 and used as observational constraints to find our best fit model. The black solid line represents the theoretical prediction based on the best fit parameters found via the minimization procedure explained in the previous section. The two different tones of blue indicates the 68%, 95% confidence levels obtained via the MCMC for the best fit cosmology[1]. These confidence levels give the idea of the amount of spread around the predictions of the best fit model. Note that fainter galaxies have a ESD which is tighter constrained than brighter galaxies. Fig. 6.10 shows the comparison between model predictions and data of the projected galaxy 2-point correlation function, w_p. Note that very faint and very bright galaxies show a relatively large scatter around the best fit model than intermediate luminosity galaxies. Moreover, in all cases, the large scale correlation is more tightly constrained than the small scale correlation. This is expected because the large scale clustering is linearly related with the dark matter clustering and it does not depend on the complicated non linear processes affecting the small scale. It is worth noticing that the quality of the clustering data is such that $w_p(R)$ has a strong constraining power. We have indeed verified that the deviation of the reduced χ^2 from unity is nearly entirely due to the deviations of the model predictions from the galaxy correlation data in the luminosity bin $[-21, -20]$.

Fig. 6.11 shows the mass-to-light ratio and the satellite fraction as predicted by the best fit model and by the MCMC for the best fit cosmology. Note that, at the low mass end, the prediction on the mass-to-light ratio is quite constraining, whereas at large masses ($M > 10^{14} h^{-1} M_\odot$) the scatter around the best fit is quite substantial. Notice also that the satellite fraction has, overall, a quite large scatter around the best fit model. In fig. 6.11 also the galaxy luminosity function is displayed. Note that, at the bright end, the predictions have larger scatter around the best fit model, whereas the faint end is quite tightly constrained.

Fig. 6.12 shows the *"data"* of the galaxy group catalogue for the best fit cosmology together with the best fit CLF and the MCMC predictions. It is remarkable how accurate the best fit CLF predictions are. All data points are reproduced with high accuracy and only at the massive end of the $\phi_{\rm s} - M$ relation, few data points are reproduced on a 2-sigma level. In general, the spread around the theoretical predictions is as large as the errors in the data. Note that the galaxy group catalogue data differ from those presented previously (see e.g. fig. 5.1), even though the group catalogue is the same. This owes to the fact that, as already mentioned, the halo mass assignments of the groups are cosmology dependent (see Y07 for details).

[1] the 99% confidence level is indicated in green

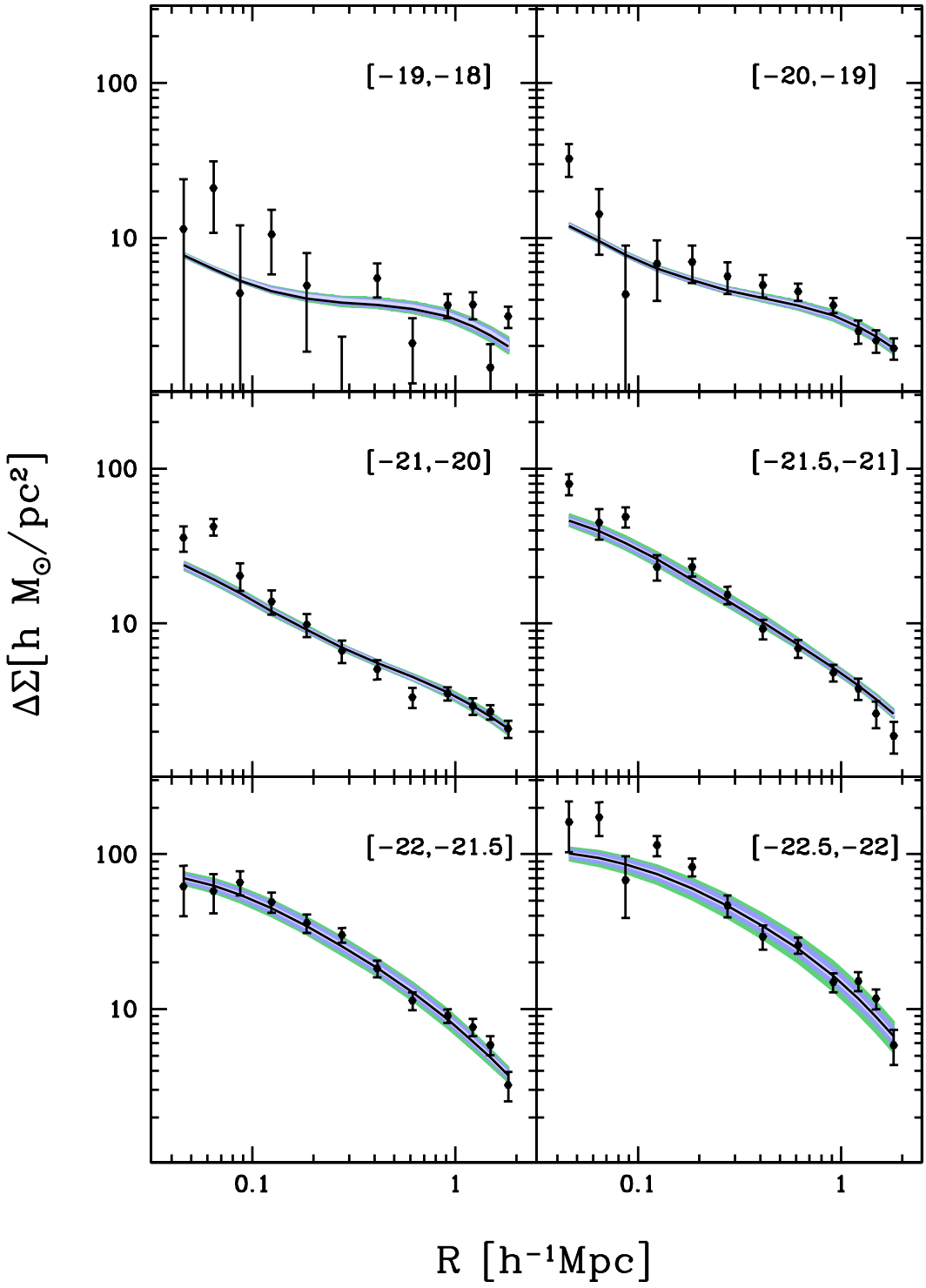

Figure 6.9: The predictions on the $\Delta\Sigma$ from the best fit model (black solid line), and the MCMC for the best fit cosmology (two tones of blue and green). For comparison, the data are shown as filled points with error bars.

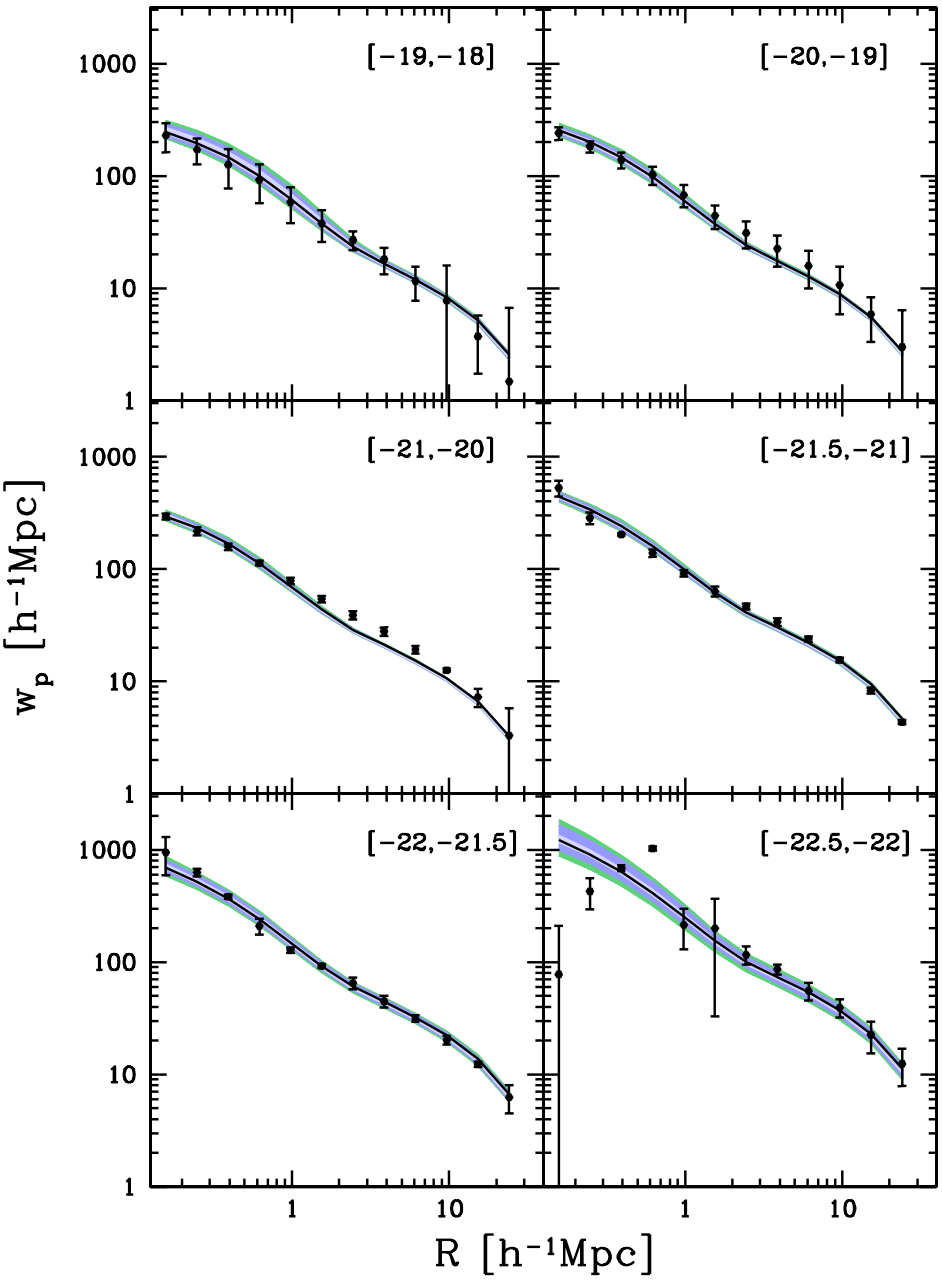

Figure 6.10: The predictions on the w_p from the best fit model (black solid line), and the MCMC for the best fit cosmology (two tones of blue and green). For comparison, the data are shown as filled points with error bars.

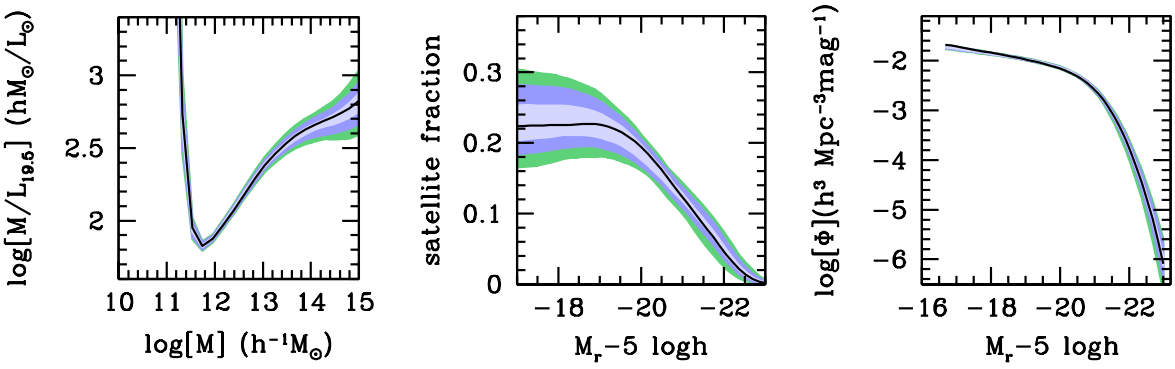

Figure 6.11: The predictions on the mass-to-light ratio (left panel), satellite fraction (middle panel) and luminosity function (right panel) are shown for the best fit model (black solid line), and the MCMC for the best fit cosmology (two tones of blue and green).

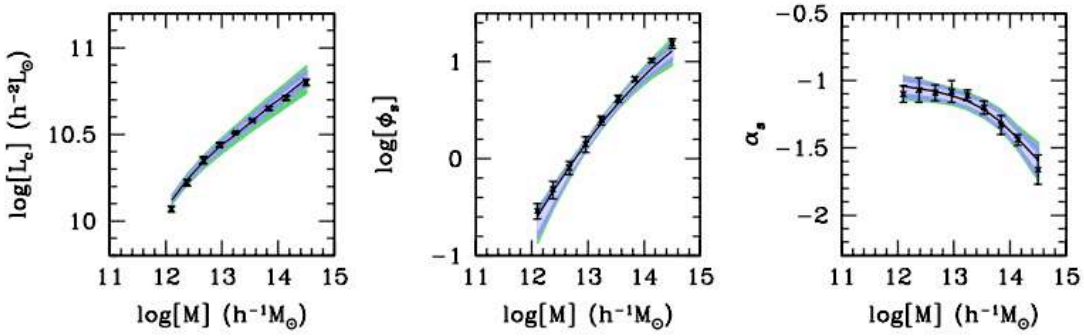

Figure 6.12: The predictions on L_c (left panel), ϕ_s (middle panel) and α_s (right panel) are shown for the best fit model (black solid line), and the MCMC for the best fit cosmology (two tones of blue and green). For comparison, the data are shown as points with error bars.

6.3 The Uncertainties of the Method

In determining the cosmology which allows the CLF to fit the data best, a number of assumptions has been made. in this section, we show how these assumptions affect the feasibility of a joint analysis of clustering and g-g lensing to constrain the value of cosmological parameters.

6.3.1 The Dark Matter Halo Concentration

Among the assumptions which define the fiducial model we have that the concentration-halo mass relation is given by the model suggested by Macciò et al. 2007. However, since this relation is based on results from numerical simulations which show a relatively large scatter (~ 0.25 in dex), we test how strongly the uncertainty on this relation affects the final result on the value of the pair $(\Omega_{\rm m}, \sigma_8)$. We analyze the case in which the $c_{\rm dm}(M)$ relation is a factor of two higher (lower) than the one suggested by Macciò et al. Since the scatter around the relation is expected to be much smaller than a factor of two, the test presented here aims to explore extreme cases in order to see what degree of limitation the $c_{\rm dm}(M)$ relation represents for constraining cosmological parameters. Fig. 6.13 shows the result of the minimization algorithm when this model variation is applied. In particular, the different tones of yellow (red) show the different confidence levels on the value of $\Omega_{\rm m}$ and σ_8 for the case in which the halo concentration is a factor of two higher (lower) than the one used for the fiducial model. Although the best fit $(\Omega_{\rm m}, \sigma_8)$ of these two variations is beyond the 3-σ confidence level of the fiducial model, we conclude that a relatively strong change in the concentration-halo mass relation does not result in a equally strong change in the value of the cosmological parameters. In fact, one has to bear in mind that a factor of two is well beyond the uncertainty of any realistic model for $c_{\rm dm}(M)$. Moreover, the size of the confidence regions derived by WMAP5 is much larger than the displacement in the $(\Omega_{\rm m}, \sigma_8)$ space obtained with our test.

We perform an additional test which concerns the concentration-halo mass relation. In the last decade, different analytical formula have been proposed to describe the results of numerical simulations. Although the qualitative behavior of all these different relations is very similar, small differences are present and the question arise whether this has an impact on determining the value of the cosmological parameters $(\Omega_{\rm m}, \sigma_8)$. We address this question by running our minimization algorithm using the $c_{\rm dm}(M)$ relation proposed by Bullock et al. (2001) and by Eke et al. (2001). Fig. 6.14 shows the confidence levels on the value of $(\Omega_{\rm m}, \sigma_8)$ obtained in these two cases. In particular, the different tones of yellow show the result when the model of Eke et al. is used and the different tones of red refer to the model of Bullock et al. Note that the two models agree relatively well with the fiducial model. In fact, the confidence regions overlap on a 2-σ level. Note also that in the case of Eke et al. $c_{\rm dm}(M)$ model, the χ^2 surface is particularly steep and this causes a much smaller confidence region than in the other cases.

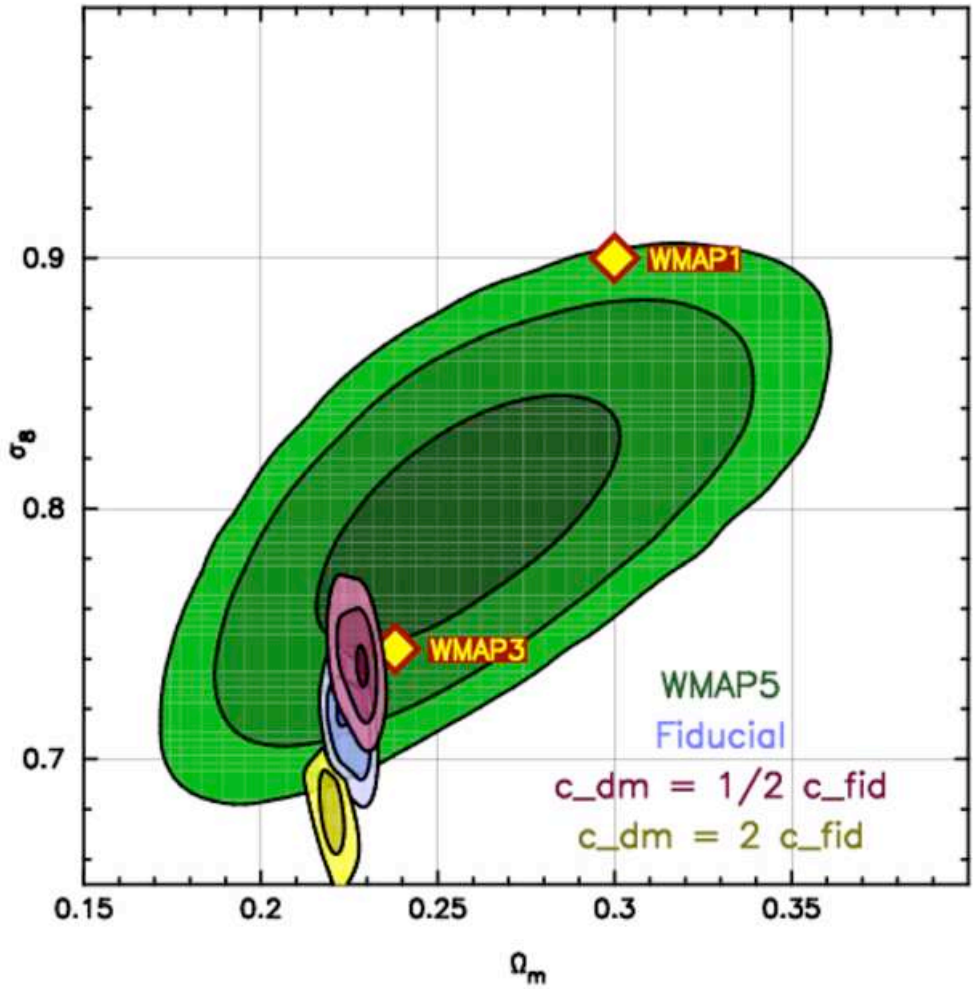

Figure 6.13: The 68%, 95% and 99% confidence levels on the values of Ω_m and σ_8 derived by WMAP5 (three tones of green) and our joint analysis of galaxy clustering and g-g lensing (three tones of blue for the fiducial model). The different tones of red (yellow) delimitate the model variation with the halo concentration assumed to be a factor of 2 lower (higher) than the fiducial model. Note that the best fit value from WMAP1 and WMAP3 are also shown with yellow diamonds.

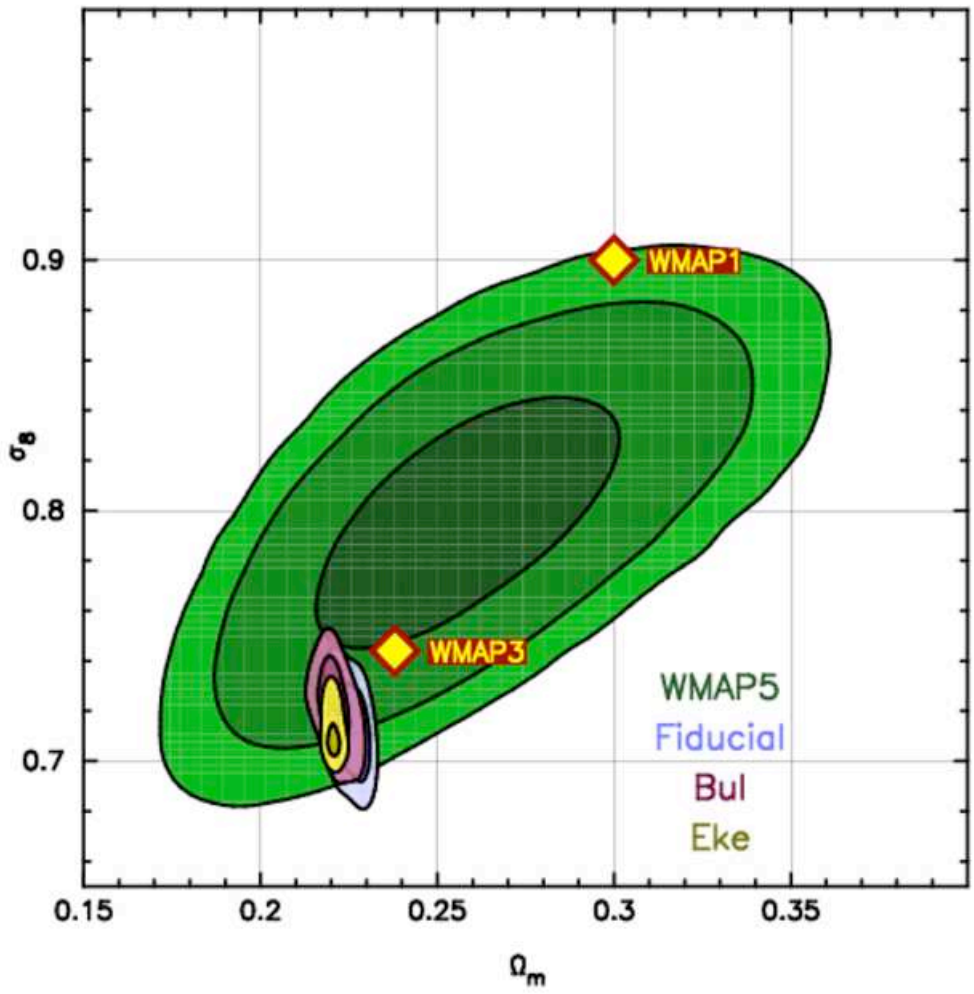

Figure 6.14: The 68%, 95% and 99% confidence levels on the values of Ω_m and σ_8 derived by WMAP5 (three tones of green) and our joint analysis of galaxy clustering and g-g lensing (three tones of blue for the fiducial model). The different tones of red (yellow) delimitate the model variation with the halo concentration assumed to follow the Bullock et al. (Eke et al.) fitting function. Note that the best fit value from WMAP1 and WMAP3 are also shown with yellow diamonds.

6.3.2 The Radial Distribution of Satellite Galaxies

An assumption which characterizes the fiducial model is that satellite galaxies follow the same distribution of the dark matter inside haloes. In other words, one can describe the radial distribution of satellite inside haloes with a NFW profile. We explore the impact of this assumption in determining the value of the cosmological parameters (Ω_m, σ_8) by changing the concentration of this profile. Following the same notation presented in chapter 5, we can assume that satellite galaxies are two times more centrally concentrated than dark matter by using $\mathcal{R}$ equals 0.5. Accordingly, $\mathcal{R} = 2$ corresponds to a satellite galaxy distribution which is two times less centrally concentrated. As it can be seen in fig. 6.15, these two models do not result in statistically different values for (Ω_m, σ_8). In fact, the corresponding results agree with the fiducial model on a 2 to 3 σ level. Note also that the model with $\mathcal{R}$ equals 0.5 has a steep χ^2 surface which results in an extremely small confidence region.

6.4 Towards a comprehensive analysis

So far, only Ω_m and σ_8 have been considered as free cosmological parameters. The values of all other cosmological parameters has been fixed to that suggested by WMAP3. However, some of these cosmological parameters may play a relatively important role in shaping the quantities which describe the galaxy-dark matter relation. If that is the case, the result on the value of the pair (Ω_m, σ_8) may be affected by having fixed the values of the other cosmological parameters. However, it is worth noticing that the remaining cosmological parameters are the amount of baryons in the universe, Ω_b, the value of the Hubble constant, h, and the power-law index of the power spectrum of the initial density fluctuations in the universe, n. The first two parameters are very tightly constrained by primordial nucleosynthesis models (Burles, Nollet & Turner 2001), from CMB anisotropy measurements (Hanay et al 2000; de Bernardis et al. 2002; Pryke et al. 2002; Netterfield et al 2002) and by the Hubble Key Project (Freedman et al. 2002), respectively. The largest uncertainty is present only on the value of n. Different inflationary models suggest different values of this parameter. Note, anyway, that the results from the analysis of the temperature fluctuations in the CMB seem to suggest values of n close to unity. In particular, over the last years, the WMAP team has supported values of n which may vary from 0.9 to 1.0 (corresponding to different analyses as well as statistical and systematic errors).

As a first preliminary test, we run the same minimization algorithm as in the previous section but we now assume two different values of n. We use $n = 0.925$ and $n = 0.975$ which are values 2-σ away from the best fit value in WMAP3 results. We again draw the 68%, 95% and 99% confidence levels on the values of the parameters (Ω_m, σ_8). Fig. 6.16 shows the results in comparison with the fiducial model and the WMAP5 results. It is encouraging that even a drastic change in the value of n is not affecting the overall result about the favored (Ω_m, σ_8) pair. Note, in fact, that the results of the two model variations are within three sigma of the fiducial model. This test shows that using Ω_m and σ_8 as the two cosmological parameters mostly responsible for the galaxy clustering and g-g lensing is indeed a fair assumption. It also shows that the value of n, if within the range suggested by WMAP, is not significantly affecting the result of the best fit (Ω_m, σ_8).

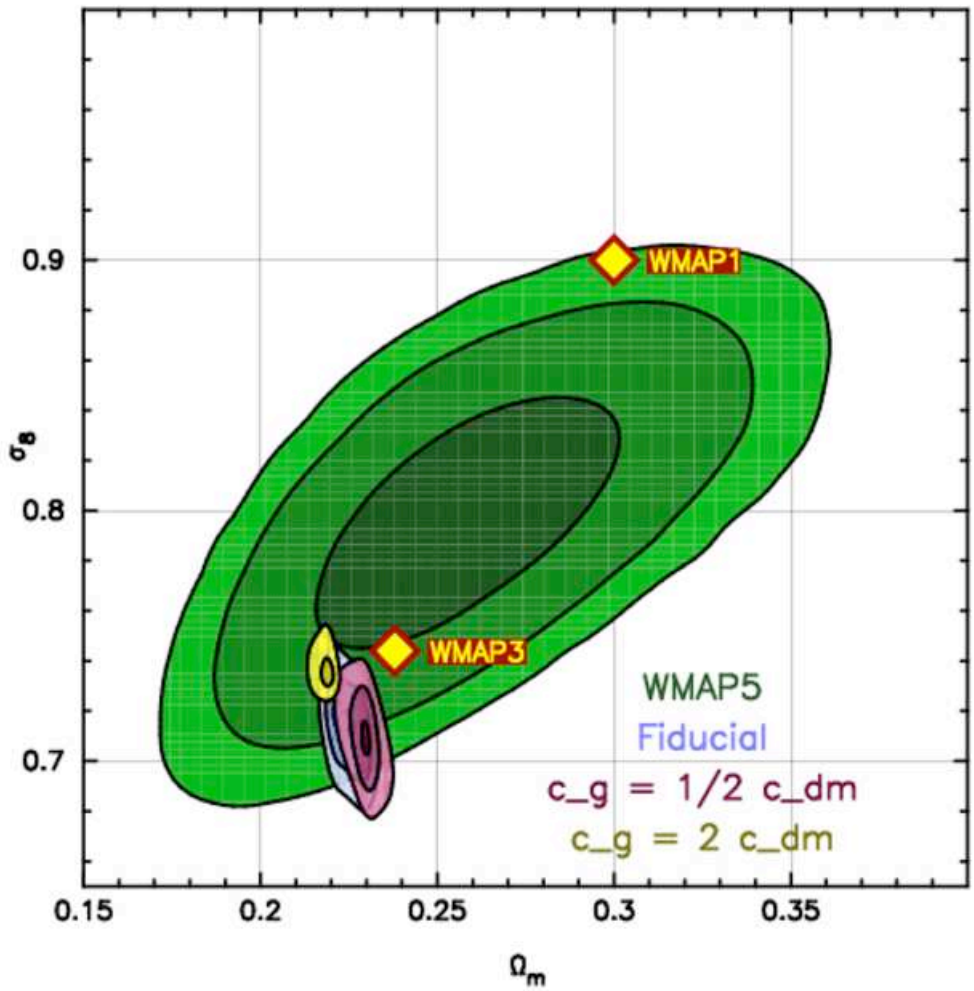

Figure 6.15: The 68%, 95% and 99% confidence levels on the values of Ω_m and σ_8 derived by WMAP5 (three tones of green) and our joint analysis of galaxy clustering and g-g lensing (three tones of blue for the fiducial model). The different tones of red (yellow) delimitate the model variation with the galaxy concentration assumed to be a factor of 2 lower (higher) than the fiducial model. Note that the best fit value from WMAP1 and WMAP3 are also shown with yellow diamonds.

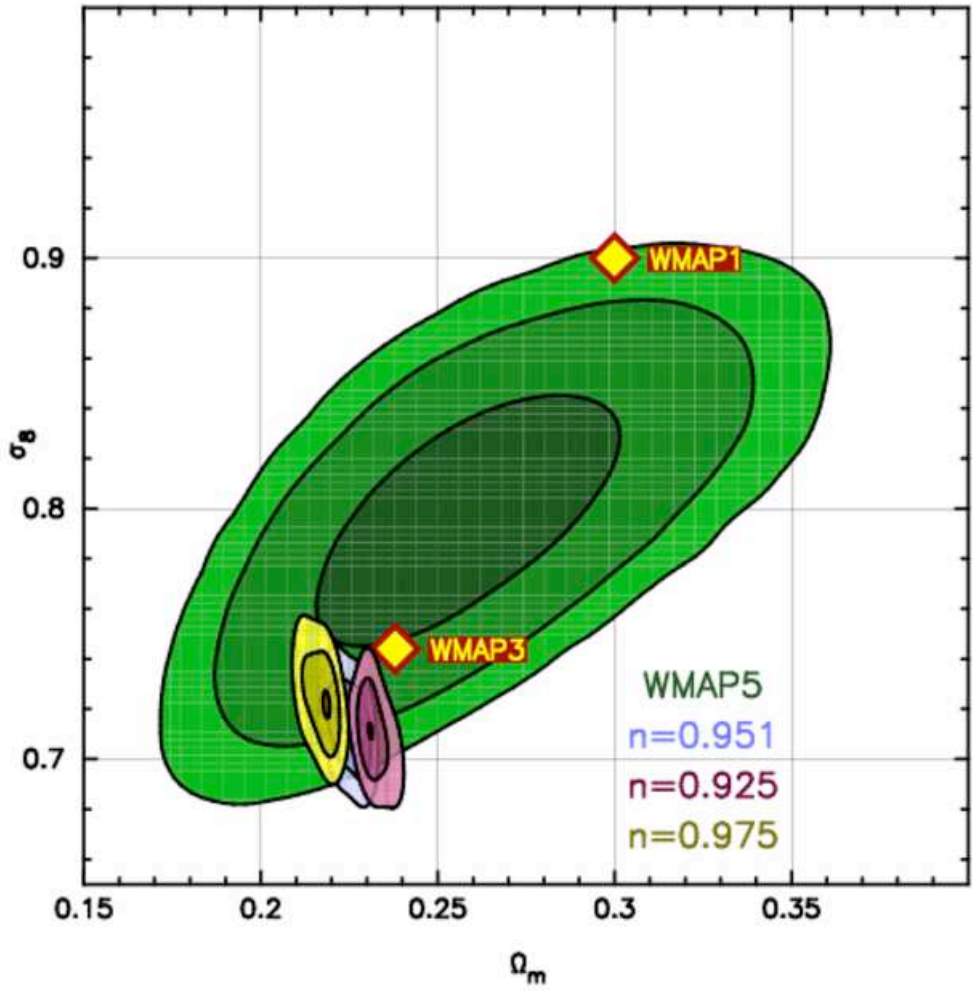

Figure 6.16: The 68%, 95% and 99% confidence levels on the values of $\Omega_{\rm m}$ and σ_8 derived by WMAP5 (three tones of green) and our joint analysis of galaxy clustering and g-g lensing (three tones of blue for the fiducial model). The different tones of red (yellow) delimitate the model variation with $n = 0.925$ ($n = 0.975$). Note that the best fit value from WMAP1 and WMAP3 are also shown with yellow diamonds.

Bearing this last result in mind, we can relax the assumption of having only two leading cosmological parameters and we can embed the joint analysis of galaxy clustering and g-g lensing in a cosmological framework which is defined by the five cosmological parameters: $(\Omega_{\rm m}, \sigma_8, n, h, \Omega_{\rm b})$. We can thus run a MCMC in a sixteen free parameter space. Eleven parameters define the CLF, whereas five define the cosmology. The technical detail about this MCMC is the same as in the previous ones with the only difference that we have increased the number of model evaluations in the chain by a factor of two. The end result is thus a MCMC of 4000 independent models properly sampling the posterior distribution of the sixteen free parameters. Note that all previous MCMC have always assumed flat prior on the values of the parameters. This means that no *"a priori"* constraints have been required for the value of the parameters. In what follows, however, we will restrict the range of few cosmological parameters following the results of previous studies. In particular, we will impose a top-hat prior on the value of n, h, and $\Omega_{\rm b}$. Note that the choice of the top-hat prior is motivated by the fact that these cosmological parameters are already well constrained within a given range by other independent studies. Since the ultimate goal is to compare all available cosmological probes to use the strength of each of them in breaking degeneracies, it is sensible and not restrictive to incorporate a top-hat prior every time a cosmological parameter is "a priori" well constrained. We then have $\Omega_{\rm m}$ and σ_8 as completely free cosmological parameters, whereas we impose the prior $n \in [0.9, 1.1], h \in [0.64, 0.8]$, and $\Omega_{\rm b} \in [0.03, 0.05]$. Fig. 6.17 shows the result of this analysis. The three different tones of blue shows the 68%, 95%, and 99% confidence level on the value of $(\Omega_{\rm m}, \sigma_8)$. Note that, as expected, the confidence region is larger than the one previously presented. This derives from having treated all five cosmological parameters as free. However, note also that the confidence region is still much smaller than the one derived by the WMAP analysis. Having performed a complete marginalization over all cosmological parameters, we can provide statistical errors, we thus have $\Omega_{\rm m} = 0.237 \pm 0.015$ and $\sigma_8 = 0.71 \pm 0.02$, where the errors indicate 95% confidence level. We emphasize that these statistical errors are extremely small compared to those provided by other cosmological probes (e.g. Dunkley et al. 2009). As already shown in fig. 6.16, there is still evidence for a degeneracy between n, σ_8 and consequently $\Omega_{\rm m}$: lower values of n yielding towards lower values of σ_8 and higher values of $\Omega_{\rm m}$. Furthermore, the posterior distribution of the cosmological parameters $\Omega_{\rm b}$ and h is flat. This means that our analysis does not further constrain the values of these two parameters.

We conclude that combining galaxy clustering and g-g lensing can be used as a novel, competitive and complementary method to constrain the value of cosmological parameters.

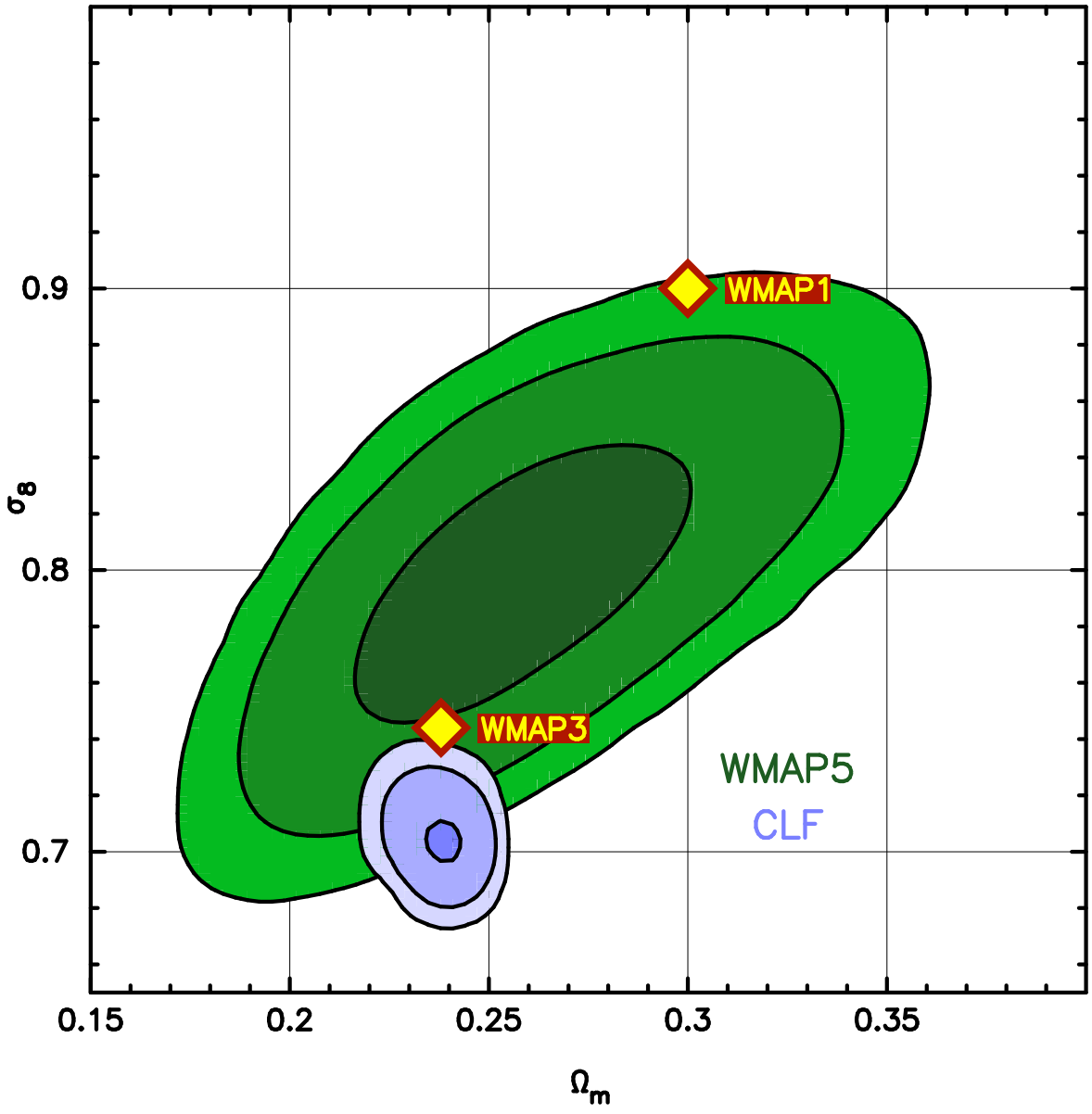

Figure 6.17: The 68%, 95% and 99% confidence levels on the values of Ω_m and σ_8 derived by WMAP5 (three tones of green) and the complete MCMC of the joint analysis of galaxy clustering and g-g lensing (three tones of blue). Note that the best fit value from WMAP1 and WMAP3 are also shown with yellow diamonds.

7

Conclusions

We conclude this thesis summarizing the main aspects of the method used to study the properties of the galaxy-dark matter connection with emphasis on the original results and achievements. Furthermore, we speculate about possible future applications which can be considered as natural follow-up projects of this Ph.D. thesis.

7.1 Summary

Although the advent of wide and deep surveys has resulted in clear detections of galaxy-galaxy lensing, a proper interpretation of these data in terms of the link between galaxies and dark matter haloes has been hampered by the fact that the lensing signal can only be detected when stacking the signal of many lenses. Since not all lenses reside in haloes of the same mass, the resulting signal is a non-trivial average of the lensing signal due to haloes of different masses. In addition, central galaxies (those residing at the center of a dark matter halo) and satellite galaxies (those orbiting around a central galaxy) contribute very different lensing signals, even when they reside in haloes of the same mass (e.g., Yang et al. 2006). This has to be properly accounted for, and requires a detailed understanding of the way galaxies populate dark matter haloes. Furthermore, galaxy-galaxy lensing probes the galaxy-dark matter connection in complementary way to galaxy clustering. In fact, since the halo occupation statistics inferred from the observed clustering properties of galaxies are degenerate with the adopted cosmology, different cosmologies imply different mass-to-light ratios for dark matter haloes. Galaxy-galaxy lensing, on the other hand, yields *"direct"* constraints on the actual mass-to-light ratios of dark matter haloes. Combined, clustering and lensing therefore offer the opportunity to constrain cosmological parameters.

In this thesis, we have modeled galaxy-galaxy lensing with the CLF, $\Phi(L|M)$, which describes the average number of galaxies of luminosity L that reside in a halo of mass M. This CLF is ideally suited to model galaxy-galaxy lensing. In particular, it is straightforward to account for the fact that there is scatter in the relation between the luminosity of a central galaxy and the mass of its dark matter halo.

This represents an improvement with respect to previous attempts to model the g-g lensing signal obtained from the SDSS, which typically ignored this scatter (e.g. Seljak et al.2005; Mandelbaum et al.2006). However, in agreement with Tasitsiomi et al.(2004), we have demonstrated that the scatter in this relation has an important impact on the g-g lensing signal and cannot be ignored. We also improved upon previous studies by modelling the two halo term (the contribution to the lensing signal due to the mass distribution outside of the halo hosting the lens galaxy), including an approximate treatment for halo exclusion.

Following Cooray & Milosavljević (2005), we split the CLF in two components: one for the central galaxies and one for the satellites. This facilitates a proper treatment of their respective contributions to the g-g lensing signal. The functional forms for the two CLF components are motivated by results obtained by Yang et al.(2008) from a large galaxy group catalogue. For a given cosmology, the free parameters of the CLF are constrained using the luminosity function, the correlation lengths as function of luminosity, and some properties extracted from the group catalogue. We have performed our analysis for the so-called WMAP3 cosmology and we have obtained a CLF that can accurately fit the abundances and clustering properties of SDSS galaxies. The halo occupation statistics corresponding to this CLF can be used to predict the g-g lensing signal. The WMAP3 CLF predictions of the g-g lensing signal are in excellent agreement with the data obtained with SDSS with no need for additional tuning of the parameters. In order to test the robustness of our results we have performed a number of tests. In particular, we have shown that small uncertainties in the expected concentrations of dark matter haloes, or in the radial number density distributions of satellite galaxies, only have a very small impact on the predicted lensing signal. In addition, although our treatment of halo exclusion is only approximate, we have demonstrated that it is sufficiently accurate. Finally, making the oversimplified assumption that dark matter haloes are spherical rather than ellipsoidal also has a negligible impact on the lensing predictions. We thus conclude that our method yields accurate and reliable predictions for g-g lensing.

We emphasize that it is not trivial that a single halo occupation model can be found that can simultaneously reproduce the galaxy abundance, the clustering and lensing properties of galaxies and the results obtained from a group catalogue. This can be read as a proof that the analytical approach based on the CLF is able to catch the main properties of the way galaxies populate dark matter haloes adequately accounting for all the intrinsic uncertainties.

We have repeated the exercise of constraining the CLF using the abundances and clustering of galaxies assuming a slightly different cosmological model (WMAP1). Although the cosmological parameters of the WMAP1 and WMAP3 cosmologies only differ at most at a 20 percent level, the CLF for the WMAP1 cosmology predicts excess surface densities that are much higher than observed. We have used this result as an evidence that a combined analysis of galaxy clustering and g-g lensing is extremely sensitive to the underlying cosmology. We thus have performed a detailed study aimed to understand to which extend a joint analysis of galaxy clustering and g-g lensing may be used as a new technique to constrain the values of cosmological parameters. It is worth noting that such a method has not been explored in the literature because of the lack of confidence in the current

understanding of the properties of galaxies and, in general, of the non-linear regime of the structures in the universe. However, the success of the CLF model in reproducing the statistical properties of galaxies (i.e. abundances, clustering and lensing) is encouraging and it can be read as an achievement in the understanding of the behavior of *average* quantities which relate galaxies with their dark matter haloes.

Our approach in testing the feasibility of the method has been quite pedagogical, as often required in exploring novel techniques. In fact, we have started by limiting to the minimum the amount of freedom of the model and we have progressively relaxed the assumptions until a full comprehensive analysis has been performed. More specifically, we have focused on constraining the values of only two cosmological parameters: $\Omega_{\rm m}$ and σ_8. These two parameters are, in fact, believed to be the most important in shaping the galaxy clustering and g-g lensing relations. All other cosmological parameters have been kept fixed at the value suggested by WMAP3. In this simplified scheme, the pair $(\Omega_{\rm m}, \sigma_8)$ which allows one to best fit the data is in relatively good agreement with results from other independent probes as the temperature fluctuations in the CMB, the cluster abundance, the BAO in the galaxy correlation function and combined analyses (see e.g. Spergel et al. 2007). Note that the fact that all cosmological parameters except $\Omega_{\rm m}$ and σ_8 were fixed or, in other words, the lack of marginalization over all cosmological parameters is expected to result in narrow confidence regions around the best fit $(\Omega_{\rm m}, \sigma_8)$. We thus provide 68%, 95% and 99% confidence levels only as a measure of the uncertainty in the determination of the two cosmological parameters in such a simplified approach. Furthermore, we investigate the impact of some crucial assumptions which enter the model. Basically, we study the impact of the assumptions about the radial distribution of satellite galaxies and about the concentration of dark matter haloes. We conclude that, within the current uncertainties, these two uncertain quantities do not affect the results on the value of $(\Omega_{\rm m}, \sigma_8)$ in a statistical sense. As a successive step, we relax the assumption about the *"a priori"* knowledge of the value of the power law index, n, of the initial power spectrum of density fluctuations. We show that, if we limit the changes in n to the range suggested by CMB experiments, the final result is not statistically affected. However, there is evidence for a degeneracy such that the lower is n the lower is σ_8 and thus the higher is $\Omega_{\rm m}$. Finally, we perform a full analysis in which also the value of the Hubble constant, h, and of the baryon energy density, $\Omega_{\rm b}$, is free to vary within the range suggested by previous studies (e.g. Spergel et al. 2007; Dunkley et al. 2009). As expected, this has negligible impact on our method and it thus does not significantly affect the final result on the $(\Omega_{\rm m}, \sigma_8)$ pair. From this analysis, we obtain $(\Omega_{\rm m}, \sigma_8) = (0.237, 0.71)$ as the cosmological pair that allows the model to best fit the data. Having performed a complete marginalization over all cosmological parameters, we can provide statistical errors, we thus have $\Omega_{\rm m} = 0.237 \pm 0.015$ and $\sigma_8 = 0.71 \pm 0.02$, where the errors indicate 95% confidence level. We emphasize that these statistical errors are extremely small compared to those provided by other cosmological probes (e.g. Dunkley et al. 2009). This is mostly due to the fact that galaxy clustering and g-g lensing probe the galaxy-dark matter connection in complementary ways.

We conclude that the joint analysis of galaxy clustering and g-g lensing can be used as a new, complementary and competitive method to constrain the value of cosmological parameters.

7.2 Possible Developments

The technique illustrated in this thesis opens the avenue to possible future applications. Interestingly, both astrophysical and cosmological aspects can be further investigated. However, at this stage, we simply mention preliminary ideas whose actual feasibility needs to be tested. In the following sections, we provide qualitative considerations for any of these applications.

7.2.1 Constraints on galaxy formation and evolution

The ultimate reason for being interested in measurements as galaxy clustering and/or g-g lensing is to relate galaxies to dark matter haloes in a way that provide informative tests about galaxy formation theory. In this thesis, we have seen that the CLF approach basically consists of converting galaxy clustering and g-g lensing into relations between galaxies and their dark matter haloes. The amount of information that we gain from this kind of analysis is strongly dependent on the quality of the data. Although the measurements of the clustering of galaxies as a function of their properties have become relatively accurate, galaxy-galaxy lensing measurements are still very poor. It is worth reminding, though, that g-g lensing is a relatively new technique with a first clear detection obtained only in 1996 (Brainerd et al.). The state of the art is represented by the measurements obtained with SDSS. It is to be expected that the quality of g-g lensing data will dramatically improve soon. For example, a g-g lensing analysis of the COSMOS field (Leauthaud et al. 2007) is underway. Due to the stability of the point spread function and the deepness of the survey, this analysis is extremely promising for detailed measurements of the ESD around lens galaxies of different observed properties. Note that two are the aspects that make this analysis particularly exciting: (i) the possibility to measure a g-g lensing signal on scales smaller than $40h^{-1}kpc$ (the current limit in SDSS data); (ii) the extremely small error bars over the all range of measurements. On one hand, the first of this aspects will allow us to probe the dark matter (sub-)haloes of satellite galaxies, helping in quantifying the amount of stripped mass in sub-haloes hosting satellite galaxies. Moreover, one can quantitatively study the impact of the baryon distributions on the lensing signal. Furthermore, one may try to constrain the dark matter density profile on scales of interest for probing theoretical predictions based on different kind of dark matter (e.g. cold vs warm). Although these three studies will be mostly degenerate, the modeling of g-g lensing via the CLF can be surely used as a tool to give insights about the relative contributions. On the other hand, aspect (ii) is of interest mainly for probing the radial distribution of satellite galaxies with the potential of quantifying the effect of luminosity segregation (i.e. the possibility that satellite galaxies with different luminosity reside, on average, in different part of the host halo). As shown in detail in chapter 4, the analytical model used throughout the thesis encodes all is required to properly

account for this complication and, once the data will be released, it will be straightforward to compare them with model predictions.

An important aspect that we have not emphasized yet is that the study of the galaxy-dark matter connection should be conducted as a function of time in order to understand the co-evolution of galaxies and dark matter haloes. One can, in principle, relate galaxies to dark matter haloes using the CLF approach at different redshifts. Relying on the idea that the evolution of dark matter haloes can be studied with numerical simulations (e.g. Springel et al. 2006), the evolution in the halo occupations statistics may provide valuable constraints on baryonic physics in galaxies over time (e.g. accretion and consumption of gas, star formation and so on). A first step in this direction could be to repeat the exercise presented in chapter 5 but at redshift higher than $z \sim 0.1$ (the average redshift of SDSS galaxies). This is feasible thanks to the effort that has been placed on surveys as DEEP2 (Davis et al. 2003), COMBO-17 (Wolf et al. 2004), NDWFS (Jannuzi & Dey 1999) and VVDS (Le Fevre et al. 2005) and on forthcoming g-g lensing analysis (see e.g. Leauthaud et al. 2007, Bernstein 2009 and reference therein). It is worth emphasizing that our method benefits of the information derived from g-g lensing mostly for its complementarity to the one derived from galaxy clustering. Therefore, the ideal observational situation would be to have both clustering and g-g lensing data with relatively high accuracy and at different redshifts. Once this is achieved, the evolution of the halo occupation statistics studied via the CLF approach would provide a transparent way to look at the evolution of statistical properties of galaxies and their relation with dark matter haloes, practically constraining galaxy formation scenarios.

7.2.2 Constraints on cosmological parameters: degeneracies and systematics

Although the results presented in chapter 6 allows us to conclude that the joint analysis of galaxy clustering and g-g lensing can be used to constrain the value of the cosmological parameters $\Omega_{\rm m}$ and σ_8, different directions can be taken to further study the feasibility of this novel technique. It is important to remind that the functional form used to describe the CLF requires eleven parameters which may be partly degenerate with the cosmological parameters. In general, one could adopt the Fisher information matrix to assess how well the data can constrain the value of the model parameters. Moreover, it would be ideal to combine the analysis presented in this thesis with other well established cosmological probes such as CMB temperature fluctuations, and cluster abundance. In general, different probes are sensitive to different parameters or they are affected by degeneracies along different directions in the parameter space. Thus, a combined analysis is the best way to simultaneously break degeneracy and learn about systematics.

7.2.3 Constraints on dark energy models

When treated as standard candles, the apparent dimming of distant Type Ia supernovae surprisingly pointed towards an accelerating expansion of the Universe (Riess et al. 1998; Perlmutter et al. 1999). This was the first piece of observational evidence in favour of the presence of a negative pressure

component in the energy budget of the Universe. Independent support for this component, called dark energy, came from the combination of CMB measurements and large scale structure data (Efstathiou et al. 2002; Tegmark et al. 2004). Understanding the nature of dark energy has become one of the most important problems in physics today since it has strong implications for our understanding of the fundamental physical laws of the Universe. The simplest possibility is that the dark energy corresponds to the vacuum energy, in which case it behaves analogously to the cosmological constant, Λ, with equation of state $p = w_{\rm de}\rho_{\rm de} = -1\rho_{\rm de}$, which is basically the assumption used throughout this thesis. However, several alternative models are possible. One way to narrow down the wide range of possible models is to obtain constraints on the dark energy equation of state parameter $w_{\rm de}$. Note that one could, in principle, allow for variations in the value of $w_{\rm de}$ and relax the assumption that it is constant with time by exploring the possible redshift dependence of this parameter. One approach to constrain the time variation in the equation of state parameter is to express $w_{\rm de}$ as a first order Taylor expansion around $z = 0$, e.g. $w_{\rm de}(z) = w_{\rm de}(z = 0) + w'_{\rm de}\, z$. The advantage of this method is the simplicity; any value of $w'_{\rm de} \neq 0$ would indicate a non-constant equation of state parameter. The most promising approach is to combine information at different epochs. In general, the cosmic acceleration driven by dark energy entails particular connections between the expansion history, spatial geometry, and the growth of fluctuations, and hence between the observables measured at low redshift (galaxy abundances, galaxy clustering and gravitational lensing), and at high redshift (spectrum of the CMB temperature fluctuations). It is definitively worth exploring the possibility to embed the analysis presented in this thesis in a more general framewiork which allows us to incorporate the dark energy equation of state. In particular, it would be interesting to constrain $w_{\rm de}(z)$ by using the combined analysis of galaxy clustering and g-g lensing for different redshifts.

Acknowledgements

I acknowledge the support of International Max Planck Research School (IMPRS) for Astronomy and Cosmic Physics at the University of Heidelberg. I also acknowledge R. Mandelbaum for providing the lensing data in electronic format and for her kind co-operation. Furthermore, I thank Alexie Leauthaud for useful discussions, as well as Nikhil Padmanabhan and Joanne Cohn for fruitful comments during the visit at UC Berkeley/LBNL. It is my pleasure to thank the Dublin Institute for Advanced Studies (DIAS) for hosting me during the writing of this thesis.

I am enormously grateful to Frank. His enthusiasm for astronomy has been really inspiring for me. From him I have learned how to question scientific results and, more important, I have learned to be judgmental. It is with great pleasure that I thank Matthias because he always had time and valuable comments for me.

I thank Surhud, more than a Ph.D. fellow: an inexhaustible source of peace and positive thinking. His joy has helped me a lot during the difficult phases of the last three years. Thanks to Anupreeta for making Surhud so happy. I thank Cassie for her "joie de vivre". Thomas for pancakes during weekends. Ros because it's good to know that there is someone who speaks *proper English*. Christian because I never thought I could have a good friend with so many diametrically opposite views. I thank Yaroslav for the profound discussions during downhill walking. And I am happy to thank Kris and Ramin: the two most european north american I have ever met. I also want to thank Claudia for sharing with me the hobby of photography. Thanks to the *"italian coffee crew"* at MPIA, because sometimes a cup of espresso can make me feel home. At the same time I really want to thank Simon, my flatmate: a Bavarian and a Sicilian under the same roof give birth to a really *interesting* atmosphere. I am grateful for all the nice people I have met during these three years in MPIA and in Heidelberg in general.

A very special thank goes to Lisa. Thanks for being so sweet and understanding. Since I met you everything has been extremely nice. I really hope we will keep this and make it grow.

Un ringraziamento particolare va anche a chi ha seguito questa mia vicenda a distanza. Grazie alla mia famiglia che ripone in me fiducia senza limiti. Grazie a mia sorella ed a mio fratello: disponibili, solidali e sempre capaci di incoraggiarmi. Un grazie a Maria Carmela e Fabrizio che sento vicini come sorella e fratello. Ed ancora un grazie

di cuore ai miei quattro nipoti che mi hanno insegnato tanto e con i quali ho condiviso momenti di gioia di un'intensità unica. Ed ancora voglio ringraziare chi si è dimostrato un compagno di viaggio anche a distanza: grazie Francesco. Grazie ad Ale, per i mesi stupendi trascorsi ad Heidelberg. Grazie a Federico, Luca, Maurizio, Corrado, Silvio e Livio perché sono con me nelle mie scelte (anche quando non lo sanno).

Grazie a tutti quelli che hanno contribuito alla mia felicità e alla mia crescita.

Thanks to all those people I met and I shared happy moments with.

Bibliography

Abazajian K., et al., 2004, AJ, 128, 502

Abell G.O., Eastwood T.S., BAAS, 7, 426

Adelman-McCarthy J.K., et al., 2006, ApJS, 162, 38

Astier P., et al., 2006, A&A, 447, 31

Bacon D.J., Taylor A.N., 2003, MNRAS, 344, 1307,

Bahcall N.A., 1977, ARA&A, 15, 505

Bahcall N.A., et al., 2003, ApJ, 585, 182

Balbi A., et al., 2000, ApJ, 545L, 1

Baudis, 2006, cosp, 36, 3621

Becker M.R., et al., 2007, ApJ, 669, 905

Beers T.C., Tonry J.L., 1986, ApJ, 300, 557

Bennett D.P., at al., 1991, BAAS, 23Q1331

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

Berlind A. A., et al., 2006, ApJS, 167, 1

Bernabei et al., 2008, NIMPA, 592, 297

Biviano A., 2000, cucg.confE, 1

Blake C., Collister A., Bridle S., Lahav O., 2007, MNRAS, 374, 1527

Blanton M.R., et al. 2003a, ApJ, 592, 819

Blanton M.R., et al. 2003b, AJ, 125, 2348

Brainerd T.G., & Specian M.A., 2003, ApJ, 593L, 7

Brainerd T.G., Wright O.C., 2002, ASPC, 283, 177

Brainerd T.G., Blandford R.D., Smail I., 1996, ApJ, 466, 623

Bonvicini V., et al., 2001, NIMPA, 461, 262

BIBLIOGRAPHY

Borgani S., et al., 2001, ApJ, 561, 13

Bosma A., 1978, PhDT, 195

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

Burles S., Nollett K.M., Turner M. S., 2001, ApJ, 552L, 1

Carlberg R.G., Yee H.K.C., Ellingson E., 1997, ApJ, 478, 462

Cebrián., et al. 2000, NJPh, 2, 13

Chen J., 2007, preprint (arXiv:0712.0003)

Cole S., Lacey C., 1996, MNRAS, 281, 716

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

Cooray A., Milosavljević, 2005, ApJ, 627, L89

Cooray A., 2006, MNRAS, 365, 842

Conroy C. et al., 2007, ApJ, 654, 153

Crook A.C., et al., 2007, ApJ, 655, 790

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

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

deBernardis P., et al., 2002, ApJ, 564, 559

De Lucia G., Kaufmann G., Springel V., White S.D.M., Lanzoni B., Stoehr F., Tormen G., Yoshida N., 2004, MNRAS, 348, 333

Dunkley et al., 2009, ApJS, 180, 306

Efstathiou G., Bond J. R.,1999, MNRAS, 304, 75

Efstathiou G., et al., 2002, MNRAS, 330, 29

Einasto J., et al., 2007, A&A, 462, 811

Eke V.R., Cole S., Frenk C.S., Navarro J.F., 1996, MNRAS, 281, 703

Eke V.R., Navarro J.F., Steinmetz M., 2001, ApJ, 554, 114

Eke V.R., Navarro J.F., Steinmetz M., 2001, ApJ, 554, 114

Eke V.R., et al., 2004, MNRAS, 355, 769

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

Eisenstein D. J., Hu W., Tegmark M., 1999, ApJ, 518, 2

Eisenstein D. J., et al., 2005, ApJ, 633, 560

Erickson L.K., Gottesmann S.T., Hunter Jr. J.H., Nature, 325, 779

Faber S.M., Gallagher J.S., 1979, ARA&A, 17, 135

Freedman W.L., et al., 2001, ApJ, 553, 47

Fukugita, M., Ichikawa, T., Gunn, J. E., Doi, M., Shimasaku, K., & Schneider, D. P. 1996, AJ, 111, 1748.

Geller M.J., Huchra J.P., 1983, ApJS, 52, 61

Gerke B.F., et al., 2005, ApJ, 625, 6

Griffiths R.E., Casertano S., Im M., Ratnatunga K.U., 1996, MNRAS, 281, 1159

Gunn, J., et al., 1998, AJ, 116, 3040.

Gunn, J., Gott J.R.I., 1972, ApJ, 176, 1

Guzik J., Seljak U., 2001, MNRAS, 321, 439

Guzik J., Seljak U., 2002, MNRAS, 335, 311

Guzzo L., et al., 2000, A&A, 355, 1

Hamana, T., 2002, ASPC, 257, 207

Hamilton A.J.S., Kumar P., Lu E., Matthews A., 1991., ApJL, 374, L1

Hanany S., et al., 2000, ApJ, 545L, 5

Hayashi E., White S.D.M., 2007, preprint (arXiv:0709.3933)

Heymans C., et al., 2006, MNRAS, 371, L60

Hoekstra H., Franx M., Kujiken K., Carlberg R.G., Yee H.K.C., 2003 , MNRAS, 340, 609

Hoekstra H., Yee H.K.C., Gladders M.D., 2004 , ApJ, 606, 67

Hogg D. W., 1999, astro-ph/5116

Hudson M.J., Gwin S.D.J., Dahle H., Kaiser N., 1998, ApJ, 503, 531

Jain B., Mo H.J. , White S.D.M., 1995, MNRAS, 276, 25L

Jain B., Bertschinger E., 1998, ApJ, 509, 517

Jannuzzi B.T., Dey A., 1999, AAS, 195, 1207

Jarvis M., et al., 2003, AJ, 125, 1014

Jing Y. P., Mo H.J., Börner G., 1998, ApJ, 494, 1

Johnston D.E. Sheldon E.S., Tasitsiomi A., Frieman J.A., Wechsler R.H., McKay T.A., 2007, ApJ, 656, 27

Komatsu E., et al., 2009, ApJS, 180, 330

Koester B.P., et al., 2007, ApJ, 660, 239

Lange A.E., et al., 2001, PhRvD, 63, 2001

Leauthaud A., et al., 2007, ApJS, 172, 219

Lebedenko V.N., 2008, preprint (arXiv:0812.1150)

LeFevre O., et al., 2005, AAS, 207, 6334

Lewis A., Bridle S., 2002, PhRvD, 66j3511

Li R., Mo H.J., Fan Z., Cacciato M., van den Bosch F.C., Yang X., More S., 2008, in preparation

Limousin M, Kneib J.P., Natarajan P., 2005, MNRAS, 356, 309

Limousin M, Kneib J.P., Bardeau S., Natarajan P., Czoske O., Smail I., Ebeling H., Smith G.P., 2007, A&A, 461, 881

Lin Y.-T., Mohr J.J., Stanford S.A., 2004, ApJ, 610, 745

Ma C., Fry J.N., 2000, ApJ, 543, 503

Macciò A.V., Dutton A.A., van den Bosch F.C., Moore B., Potter D., Stadel J., 2007, MNRAS, 378, 55

Macciò A.V., Dutton A.A., van den Bosch F.C., Moore B., 2008, preprint (arXiv:0804.1926)

Magliocchetti M., Porciani C., MNRAS, 2003, 346, 186

Mandelbaum R., Tasitsiomi A., Seljak U., Kravtsov A.V., Wechsler R.H., 2005a, MNRAS, 362, 1451

Mandelbaum R., Hirata C. M., Seljak U., Guzik J., Padmanabhan N., Blake C., Blanton M.R., Lupton R., Brinkmann J., 2005b, MNRAS, 361, 1287

Mandelbaum R., Seljak U., Kauffmann G., Hirata C.M., Brinkmann J., 2006, MNRAS, 368, 715

Mandelbaum R., Seljak U., Hirata C.M., 2008, preprint (arXiv:0805.2552)

McIntosh D.H., Zabludoff A.I., Rix H.-W., Caldwell N., 2005, ApJ, 619, 193

McKay et al., 2001, preprint (arXiv: astro-ph/0108013)

McKay T.A., et al., 2002, ApJ, 571L, 85

Merchán M., Zandivarez A., 2002, MNRAS, 335, 216

Merchán M., Zandivarez A., 2005, ApJ, 630, 759

Miralda-Escudé J., 1991, ApJ, 370, 1

Mo H.J., Jing Y.P., Börner G., 1997, MNRAS, 286, 979

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524L, 19

More S., van den Bosch F.C., Cacciato M., 2009a, MNRAS, 392, 917

More S., van den Bosch F.C., Cacciato M., Mo H.J., Yang X., Li R., 2009b, MNRAS, 392, 801

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

Natarajan P., Kneib J.P., 1997, MNRAS, 287, 833

Natarajan P., Kneib J.P., Smail I., 2002, ApJ, 580, L11

Natarajan P., De lucia G., Springel V., 2007, MNRAS, 376, 180

Natarajan P., Kneib J.P., Smail I., Treu T., Ellis R., Moran S., Limousin M., Czoske O., 2007, preprint (arXiv:0711:4587)

Netterfield C.B., et al., 2002, ApJ, 571, 604

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

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

Norberg P., Frenk C.S., Cole S., 2008, MNRAS, 383, 646

Padmanabhan, N., et al., 2007, MNRAS, 378, 852

Parker L..C., Hoekstra H., Hudson M.J., van Waerberke L., Mellier Y., 2007, ApJ, 669, 21

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

Peacock J.A., Doods S.J., 1994, MNRAS, 267, 1020

Peacock J.A., Doods S.J., 1996, MNRAS, 280L, 19

Peebles P.J.E., 1982, ApJ, 263L, 1

Percival W.J., et al., 2007, MNRAS, 381, 1053

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

Pier J.R. et al., 2000, AAS, 197, 1310

Pierpaoli E., Scott D., White M., 2002, ASPC, 268, 90

Pope A. C., et al., 2004, ApJ, 607, 655

Prada F., et al., 2003, ApJ, 598, 260

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

Pryke C., 2002, ApJ, 568, 46

Reiprich T. H., Böhringer H., 2002, luml,. conf 84

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

Rood H.J., Turnrose B.E., 1968, ApJ, 152, 1057

Rubin V.C, Ford W.K.J., Thonnard N., 1980, ApJ, 238, 471

Rubin V.C, Ford W.K.J., Thonnard N., Burstein D., 1982, ApJ, 261, 439

BIBLIOGRAPHY

Ruhl J.E., et al., 2003, ApJ, 599, 786

Sanglard V., et al., 2005, PhRvD, 71l2002

Schechter P., 1976, ApJ, 203, 297

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

Schneider P., Rix H.-W., 1997, ApJ, 474, 25

Scoccimarro R., Sheth R.K., Hui L., Jain B., 2001, ApJ, 546, 20

Seljak U., Zaldarriaga M., 1996, ApJ, 469, 437

Seljak U., 2000, MNRAS, 318, 203

Seljak U., 2002, MNRAS, 337, 769

Seljak U., et al., 2005, Phys Rev D, 71, 043511

Sheldon E.S., et al., 2004, AJ, 127, 2544

Sheldon E.S., et al., 2007a, preprint (arXiv:0709.1153)

Sheldon E.S., et al., 2007b, preprint (arXiv:0709.1162)

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

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

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

Smith R.E., et al., 2003, MNRAS, 341, 1311

Spergel D.N., et al., 2003, ApJS, 148, 175

Spergel D.N., et al., 2007, ApJS, 170, 377

Springel V., Frenk C.S., White S.D.M. 2006, Nature, 440, 1137

Tago E., et al., 2006, AN, 327, 365

Tasitsiomi A., Kravtsov A.V., Wechsler R.H., Primack J.R., 2004, ApJ, 614, 533

Tegmark M., et al., 2004, PRD, 69, 103501

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

Tinker J.L., Weinberg D.H., Zheng Z., 2006, MNRAS, 368, 85

Treu T., Koopmans, L., 2002, AAS, 201, 3201

Tyson J.A., Valdes F., Jarvis J.F., Mills A.P., 1984, ApJ, 281, L59

Udalski A., et al., 1994, Acta Astron., 44, 165

Uomoto A., Barkhouser R., Smee S., 2002, AAS, 201, 1804

van den Bosch F.C., Yang X., Mo H.J., 2003, MNRAS, 340, 771

van den Bosch F.C., Mo H.J., Yang X., 2003, MNRAS, 345, 923

van den Bosch F.C., Norberg P., Mo H.J., Yang X., 2004, MNRAS, 352, 1302

van den Bosch F.C., Yang X., Mo H.J., Norberg P., 2005a, MNRAS, 356, 1233

van den Bosch F.C., Weinmann S.M., Yang X., Mo H.J., Li C., Jing Y.P., 2005b, MNRAS, 361, 1203

van den Bosch F.C., et al. 2007, MNRAS, 376, 841

van den Bosch F.C., Pasquali A., Yang X., Mo H.J., Weinmann S.M., McIntosh D.H., Aquino D., 2008, preprint (arXiv:0805.0002)

van der Marel R.P., Magorrian J., Carlberg R.G., Yee H.K.C., Ellingson E., 2000, AJ, 119, 2038

van Waerbeke L., Tereno I., Mellier Y., Bernardeau F., 2002 (astro-ph/12150)

Viana P. T. P., Nichol R.C., Liddle A.R., 2002, ApJ, 569L, 75

Wang Y., Yang X., Mo H.J., van den Bosch, F.C., 2007, ApJ, 664, 608

Wang X., Tegmark M., Zaldarriaga M., 2002, Phys. Rev. D, 65 123001

Warren M.S., Abazajan K., Holz D.E., Teodoro L., 2006, ApJ, 646, 881

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

Wolf C., et al., 2004, A&A, 421, 913

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

Yang X., 2004,Mo H.J., Jing Y.P., van den Bosch F.C., Chu Y., MNRAS, 350, 1153

Yang X., Mo H.J., van den Bosch F.C., Weinmann S.M., Li C., Jing Y.P., 2005, MNRAS, 362, 711

Yang X., Mo H.J., van den Bosch F.C., Jing Y.P., Weinmann S.M., Meneghetti M., 2006, MNRAS, 373, 1159

Yang X., Mo H.J., van den Bosch F.C., Pasquali A., Li C., Barden M., 2007, ApJ, 671, 153 (Y07)

Yang X., Mo H.J., van den Bosch F.C., 2008, ApJ, 676, 248 (YMB08)

Yoo J., Tinker J.L., Weinberg D. H., Zheng Z., Katz N., Davé R., 2006, ApJ, 652, 26

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

Zaritsky D., Smith R., Frenk C., White S.D.M., 1993, ApJ, 405, 464

Zaritsky D., Smith R., Frenk C., White S.D.M., 1997, ApJ, 478, 39

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

Zheng Z., et al., 2005, ApJ, 633, 791