University of Sussex

Galaxy clustering using the GAMA survey

Leonidas Christodoulou

Submitted for the degree of Doctor of Philosophy

University of Sussex

September 2012

Declaration

I hereby declare that this thesis has not been and will not be submitted in whole or in part to another University for the award of any other degree.

Signature:

Leonidas Christodoulou

UNIVERSITY OF SUSSEX

Leonidas Christodoulou, Doctor of Philosophy

Galaxy clustering using the GAMA survey

Summary

We present a study of the clustering of galaxies in the local Universe ($z < 0.4$) using the SDSS and GAMA galaxy surveys. Using GAMA spectroscopic redshift we construct a large photometric redshift catalogue from the SDSS imaging data. We then measure the two-point angular correlation function as a function of photometric redshift, absolute magnitude and colour. For all our samples, we estimate the underlying redshift and absolute magnitude distributions using Monte-Carlo resampling. A linear relation between relative bias and L/L^* is found to hold down to luminosities $L \sim 0.03L^*$. We find that the redshift dependence of the bias of the L^* population can be described by the passive evolution model of linear bias. We confirm an increase in clustering strength for sub-L^* red galaxies compared with $\sim L^*$ red galaxies at small scales in all redshift bins, whereas for the blue population the correlation length is almost independent of luminosity for $\sim L^*$ galaxies and fainter. We proceed by studying the redshift space correlation function from GAMA as functions of luminosity and redshift. For $L \gtrsim L^*$ galaxies we obtain an almost constant pairwise velocity dispersion $\sigma_{12} \approx 400$ km s^{-1}, whereas for $L < L^*$ galaxies the pairwise velocity dispersion increases as we go fainter. When measured in different redshift slices the pairwise velocity dispersion as a function of luminosity shows no signs of evolution, however it does present some scale dependence. Our measurements of the growth rate parameter are consistent with the standard ΛCDM + GR cosmological model.

Imagine a vast sheet of paper on which straight Lines, Triangles, Squares, Pentagons, Hexagons, and other figures, instead of remaining fixed in their places, move freely about, on or in the surface, but without the power of rising above or sinking below it, very much like shadows - only hard and with luminous edges - and you will then have a pretty correct notion of my country and countrymen. Alas, a few years ago, I should have said "my universe": but now my mind has been opened to higher views of things.

Edwin A. Abbott, *Flatland*

Acknowledgements

First and foremost I would like to thank my supervisor Jon Loveday, for giving me the chance to work within the GAMA survey, for suggesting the projects undertaken in this thesis and for his valuable guidance. I would also like to thank my second supervisor Andrew Liddle whom, although we did not directly work together, has given me many a kind advice and feedback, regarding the direction of this thesis and of my studies in general. I am very thankful to my collaborator at Durham University Peder Norberg who, despite being based in another institution, has worked closely with me during the last two years of my thesis and who has helped me shape and clarify many of the results that are presented herein.

It is a pleasure to thank the other members of the GAMA survey: Ivan Baldry, Andrew Hopkins, Simon Driver, John Peacock, Aaron Robotham, Sarah Brough, Joss Bland-Hawthorn have all provided valuable feedback and stimulating discussions.

I would like to thank the members of the Astronomy Centre for making my time there most memorable. I would like to thank Seb Oliver, Isaac Roseboom, David Parkinson as well as the other postgraduate students and members of staff whom we share offices and many hours working or at the pub or both: Leon Baruah, Mafalda Dias, Kevin Falls, Ippocratis Saltas, Donough Regan, Gemma Anderson, Albert Asawaroengchai, Antony Lewis, Marisa March, Owain Young, Bruno Henriques, Mark Frost, Matt Thomson, Naomi Dubois, Charlotte Clarke, Sorour Shamshiri, Will Watson, Peter Hurley and Antonio Vazquez. Special thanks to Nicola Mehrtens for her encouragement and support.

I acknowledge financial support from the Greek State Scholarship Foundation trustee of the Nik. D. Chrysovergis legacy. I would like to thank Elefteria Tsaousi for her help with my scholarship's official documentation and Nektarios Vlahakis for his reports.

On a more personal level, even after four years I still miss my grandmother Maria and my dear friend Pavlos. They both had a deep and sincere interest in my physics studies. Many thanks to Kyriakos for his interest and encouragement, although I have always been unwilling to elaborate. I would like to thank Vicky for sharing with me the ups and downs of our lives and for making me a better person that I would have been, had I never met her.

Last but not least, I owe everything to my family in Cyprus who has patiently supported me throughout the years: My parents Andreas and Despina, my sister Louiza, my aunt Isabella, my uncle Michalis and my grandmother Eleni. I would especially like to thank my parents for always encouraging me to look for what makes me whole as a person, even when I failed to understand and take heed of their advices. Everything that I have learnt from them has been my guide during my postgraduate studies and it will continue to be my guide in the future.

Official acknowledgements

While working on my thesis I was financially supported by the Greek State Scholarship Foundation, trustee of the Nik. D. Chrysovergis legacy.

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including GALEX MIS, VST KIDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is http://www.gama-survey.org/ .

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web site is http://www.sdss.org. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, the Astrophysical Institute Potsdam, the University of Basel, Cambridge University, Case Western Reserve University, the University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences, Los Alamos National Laboratory, the Max Planck Institute for Astronomy, the Max Planck Institute for Astrophysics, New Mexico State University, Ohio State University, the University of Pittsburgh, the University of Portsmouth, Princeton University, the US Naval Observatory, and the University of Washington.

This research has made use of NASA's Astrophysics Data System.

To my parents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Observing the universe

It is one of the most humbling quests to contemplate the nature of the cosmos. Remarkably the last 100 years, after the publication of Einstein's general theory of relativity, this quest has undoubtedly become part of natural science. This scientific enterprise is called cosmology, famously derived from the greek word $\kappa\grave{o}\sigma\mu o\varsigma$ which means "world" and has the same root as the word gem, and it aims to understand the origin and evolution of the Universe as a whole.

Cosmology can be formally studied using a theory of gravity and thus the beginning of modern cosmology coincides with the publication of Einstein's general theory of relativity. Along with Einstein's efforts to apply general relativity in order to describe the universe one of the most important scientific discoveries of all times came from Hubble (1929), who found the proportional relation of the recessional velocity υ, of galaxies and their distance d, from us

$$\upsilon = H_0 d, \tag{1.1}$$

where the constant of proportionality H_0 is rightly called the "Hubble constant". Thus, galaxies are receding from us and the Universe is expanding. It is worth mentioning that Lemaître (1927, 1931a) made a somewhat similar observation, backed up by the first derivation of the Hubble law using general relativity. It was thereby established that the universe wasn't static. Nevertheless considerable intellectual effort was undertaken in order to establish whether the Universe was in an eternal steady state or it has started from a different, condensed state and then evolved into what we see today. Observations came to answer conclusively this question. An unavoidable conclusion for an expanding Universe described by Hubble's law is that it must have been hotter and denser in the

past and that radiation in the past was in thermal equilibrium. This cosmic radiation and its black body spectrum was discovered accidentally by Penzias and Wilson (1965) while the construction of a dedicated radiometer was in the works (this story is detailed in Peebles, 1993). It was thereby established that the Universe had a definite beginning which nowdays is known as the "Big Bang"[1]. Ever since, the cosmic microwave background (CMB) radiation has become the most important cosmological probe.

Compared with other natural sciences cosmology is subject to a major limitation already known from earlier astronomical studies. We are only able to observe the Universe, strictly speaking only our past light-cone, instead of performing controlled experiments. However, using simulations we can effectively create extensive reruns of the evolution of the Universe and thus assess the significance of our findings, under carefully chosen constraints (Springel et al., 2005). Notwithstanding this obstacle, extragalactic astronomy has delivered some of the most exciting results in the history of physics.

Firstly, it offers a stage for applying (and testing) general relativity on the largest possible scales. By combining CMB experiments and large-scale structure measurements (e.g. Tegmark et al., 2006) it has been established that the Universe is almost spatially flat and its constituents are approximately 25% of baryonic and dark matter and 75% of dark energy, a smooth fluid with negative pressure. Out of the entire matter budget only 5% consists of normal baryonic matter responsible for everything that emits light in the Universe. Although in the standard cold dark matter scenario dark matter gravitates like normal matter, it is pressureless and collisionless and it can only create virialized structures once it undergoes gravitational collapse. On the other hand, dark energy behaves very similarly to Einstein's cosmological constant and it is therefore expected to eventually dominate completely the Universe. Notwithstanding the unknown to us origin and exact nature of these dark fluids, their energy densities and in particular their ~ 1 ratio, pose another serious challenge for physicists. Dark energy and its relation with particle physics seems to be the most important and to this day a vast literature on the subject of dark energy exists, nicely summarized in various review articles (Weinberg, 1989; Carroll, 2001; Peebles and Ratra, 2003; Padmanabhan, 2008).

Secondly, there is now substantial understanding of how structures are formed in the Universe (Padmanabhan, 1993). Aside from observationally deducing the expansion of the universe Edwin Hubble also made the important observation that the distribution of galaxies appears to be homogeneous down to an increasing faint magnitude limit (Hubble,

[1]Coincidentally, Lemaître (1931b) also proposed the possibility of the beginning of the Universe from a primeval fireball.

1926, 1934). This raises the question why the Universe is distinctively inhomogeneous on small scales and what governs the transition between the two regimes. The explanation is gravitational instability, where small perturbations in an otherwise homogeneous field, get amplified by gravity and subsequently create the observed structures that we observe today like galaxies and clusters of galaxies.

In this picture galaxies act as point particle tracers, albeit biased ones, of the cosmic density field, which is dominated by dark matter. Furthermore, if the Universe has started from a smooth state and evolved into the complex network of structures that we observe today, then clearly by studying these structures and understanding the interplay between cosmology and extragalactic astronomy, one is able to make valuable cosmological inferences. This is the aim of this thesis as we pursue a study of the clustering of galaxies in the local Universe ($z < 0.4$)[2]. It was transparently shown by Peebles (1980) that the statistics of galaxy clustering can used as a probe of the constituents of the Universe, as well as an experimental test of its physical properties. In order to achieve this we take advantage of the most recent galaxy redshift surveys and we use sophisticated statistical methods to analyze the available data. Chapters 4, 5 and 6 present our results. In the rest of this Chapter we present an overview of the open problems in cosmology and extragalactic astronomy (Section 1.2) and we briefly describe structure of the thesis (Section 1.3).

Cosmology has traveled a long road since Hubble's discovery. It has not only the legitimate status of physical science today, but has also taught us an important lesson, namely that there is more to the Universe than meets the eye.

1.2 The status of cosmology today and an incomplete list of open problems

One of the biggest strengths of cosmology today is the large suite of cosmological observations which all point towards the same "concordance" ΛCDM model. Remarkably, a variety of experiments using different wavelengths, parts of the sky, with different systematics primarily agree on the value of the matter density today, giving $\Omega_{m,\text{today}} \approx 0.3$ (Peebles and Ratra, 2003; Peebles, 2012). Despite of these achievements, some of the ΛCDM pillars, like dark matter and dark energy, imply poorly understood physics or even

[2]For standard cosmology and $H_0 = 70$ km s^{-1}Mpc^{-1} this corresponds to a distance of 1543 Mpc or 5×10^9 light years.

modifications of general relativity on cosmological scales[3] (Peebles, 2003; Amendola and Tsujikawa, 2010). The challenge now is to take advantage of the present and forthcoming galaxy surveys and CMB experiments to sharpen the constraints on the cosmological parameters and hunt for indications of new physics.

In order to use galaxies for measuring background cosmological parameters, we have to understand the precise way galaxies trace the cosmological dark matter density field (Tegmark et al., 2004). Much of this thesis is building on previous work for measuring the bias of galaxies in a systematic way as a function of redshift, colour and luminosity. It is also crucial to reveal the scales for which galaxies can be reliably used for cosmological parameter constraints.

Extragalactic astronomy holds the promise of testing general relativity at scales many orders of magnitude larger than it has hitherto been tested before (Guzzo et al., 2008). The issue however is plagued with systematics and cumbersome uncertainty estimations. This is a large ongoing project for many groups working in the field and this thesis presents some work towards that direction. It goes without saying that the implications of any result favoring a modification of the gravity theory will be colossal and hence considerable effort has been put into devising appropriate experimental tests (Linder, 2005; McDonald and Seljak, 2009; White et al., 2009; Simpson and Peacock, 2010; Amendola et al., 2012).

Moving on to the galaxy formation and evolution front, we notice that it is currently dominated by the low redshift statistically powerful results coming from large wide-field surveys, with much sparser intermediate and high redshift results. Yet, it is important to extend these studies in order to capture more phenomena and understand evolutionary processes. We note that much of the input that goes into large dark matter simulations comes from observations and therefore improved observational constraints will help us create more "realistic" simulations. A particular example of such parameter is the galaxy velocity bias in dark matter haloes (Tinker, 2007).

1.3 Structure of this thesis

In Chapter 2 we introduce the theoretical framework of this work. We discus the Copernican principle (Section 2.1) and the constraints it imposes on the space-time metric. We then introduce the Friedmann equations and the cosmological redshift (Section 2.2). Cosmological distances and the age of the Universe are introduced in Section 2.3. Going

[3]Barrow (2005) described the situation as follows: "Ironically the less we know about a particular part of the Universe's make up, the more abundant it seems to be!"

beyond the background level we present a short introduction of the linear perturbation theory in Section 2.4. Statistical quantities, namely two-point functions are presented in Section 2.5. The very important halo model of large-scale structure is discussed next (Section 2.6) and the Chapter ends with a description of the cosmological parameters that are most relevant to this work (Section 2.7).

Chapter 3 presents the datasets that we use in this work, with particular attention to the GAMA survey. We begin with the description of SDSS in Section 3.1 and the description of GAMA, which consists of galaxies selected from the much larger SDSS sample, follows in Section 3.2. GAMA mock catalogues are presented in Section 3.3 and a short presentation of some of the first GAMA scientific results is given in Section 3.4.

In Chapter 4 we present a new large photometric redshift catalogue for large-scale structure studies. We begin by introducing the concept of a photometric redshift in Sections 4.1 and 4.2. The construction of the new catalogue where we used GAMA data as a training set is described step by step in Section 4.3. As photometric redshifts come with a variety of systematics we explore some of those we feel are most relevant, as well as the impact they may have in on out clustering analysis in Section 4.4. Section 4.5 presents our conclusions.

Chapter 5 presents the clustering study undertaken using the photometric redshifts from Chapter 4. After a short introduction of the physical problems we wish to tackle (Section 5.1), we outline the sample selection in Section 5.2. We discuss technical issues like masking and jackknife errors in Section 5.3. Results for the luminosity and redshift dependence of galaxy clustering are presented in Section 5.4 and colour, luminosity and redshift dependent clustering follows in Section 5.5. Our analysis also aims for a statistical description of the faint blue galaxy population, whereof spectroscopy is inefficient, presented in Section 5.6. Although we don't fit explicit halo models to our data, we include a discussion (Section 5.7) where we interpret our results using the more physically motivated halo modeling. Finally, we find that our results become more transparent and easy to interpret if use the relative bias instead of the correlation length and thus Section 5.8 presents some more quantitative clustering results. We conclude the Chapter in Section 5.9.

In Chapter 6 we present a different type of clustering analysis, namely an analysis in redshift space, where we solely use GAMA data. After a brief introduction of the aims of the Chapter in Section 6.1, we discuss the theory of redshift space distortions in Section 6.2. Our correlation function measurements in redshift space are presented in Section 6.3 and

the analysis follows in Section 6.4. In Section 6.5 we compare our results with previous results from the literature and finally, our conclusions are given in Section 6.6.

The main conclusions of the thesis are reiterated in Chapter 7. We split them in the photometric redshift and angular clustering part (Section 7.1) and the redshift space distortion part (Section 7.2). We close in Section 7.3 with a small, but important in our view, presentation of the possible future directions that this thesis has hereby uncovered.

For the remainder of this thesis we attempt to understand using theory and observations how galaxies can be related to dark matter and what can we learn from that. From a long point view, this thesis is just another steeping stone to the observational programme laid out in detail by Peebles (1980) which comprises of using large-scale structure in order to understand the physical properties our Universe. One cannot easily overstate the outstanding success of that programme so far. For on the one hand, it has changed forever our ideas about the Universe, its material content and our place within it, and on the other, more than 30 years since it's publication, it is still timely and a most promising route for even more progress in physical cosmology.

Chapter 2

Theory

This Chapter presents the theoretical framework of observational cosmology. Section 2.1 introduces the main results from applying general relativity to cosmology and the fundamental Friedmann equations, Section 2.2 relates these results with the cosmological redshift and subsequently Section 2.3 shows the translation from redshifts to cosmological distances. Linear cosmological perturbation theory is briefly introduced in Section 2.4. Section 2.5 introduces the statistical measures of clustering. The phenomenological non-linear halo modeling is presented in Section 2.6. The Chapter ends with a discussion (Section 2.7) on the values of the cosmological parameters that are most relevant for this thesis.

2.1 The homogeneous and isotropic Universe

The central assumption of cosmology is that all places in the Universe are equally privileged. This is a generalization of the age-old idea suggested by Copernicus that we should not expect the Earth to be the center of the Solar System. Almost every scientific inference in the field of cosmology is based on this assumption, which of course can only hold on sufficiently large (cosmological) scales, and consequently bears the well-deserved name Copernican Principle.

The Copernican principle asserts that the metric which describes the expanding Universe should be invariant under translations and rotations and therefore homogeneous and isotropic in space (Carroll, 2004). These symmetries put strong constraints on the functional form of the global spacetime metric, which eventually takes the following form in

spherical coordinates (t, r, θ, ϕ) (Carroll, 2004)

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right]. \tag{2.1}$$

This is the Friedmann - Lemaître - Robertson - Walker (FLRW) metric (after the names of the people who developed it) and it is used to describe the background evolution of the Universe, which is believed to respect these symmetries. In equation 2.1 a is the scale factor, the quantity that describes the expansion of the Universe and k is the global curvature of spacetime. Possible k values are -1, 0 and 1 for open, flat and closed Universe respectively. Once curvature is determined the evolution of the Universe as a whole is described only by the functional form of a. For astrophysical purposes it is convenient to rescale a to unity today, so that $a(t = t_0) = a_0 = 1$. This rescaling removes the freedom to use the set $\{-1, 0, 1\}$ as the only possible curvature values; k now is a continuous variable. The metric signuture convention is $(-, +, +, +)$. For this Chapter we also assume $c = 1$. We drop that convention in later Chapters, where we deal with observations.

In order to compute cosmological distances it is more convenient to change the radial coordinate r, to a new one χ via the following transformation

$$d\chi = \frac{dr}{\sqrt{1-kr^2}}, \tag{2.2}$$

so that

$$r = S_k(\chi) \tag{2.3}$$

and

$$S_k(\chi) = \begin{cases} \frac{1}{\sqrt{k}}\sin(\chi), & \text{if } k > 0 \\ \chi, & \text{if } k = 0 \\ \frac{1}{\sqrt{k}}\sinh(\chi), & \text{if } k < 0. \end{cases} \tag{2.4}$$

Equation 2.4 can be compactly written as (Amendola and Tsujikawa, 2010)

$$S_k(\chi) = \frac{1}{\sqrt{-k}}\sinh\left(\sqrt{-k}\chi\right). \tag{2.5}$$

In equation 2.5 $S_k(\chi) = \chi$ at the limit $k \to^- 0$.

The energy-momentum tensor in the FLRW spacetime also has to be isotropic and can be given as follows

$$T^\mu_\nu = (\rho + p)u^\mu u_\nu + p\delta^\mu_\nu, \tag{2.6}$$

where $u = (-1, 0, 0, 0)$ is the four-velovity for comoving observers, ρ is the density of the fluid and p its pressure. Note that this particular form of the metric (equation 2.1)

and the energy-momentum tensor (equation 2.6) apply only for a special set of observers, called comoving or fundamental observes. These are the observers who are at rest with the expansion of the Universe. Observers not satisfying that property need not describe the Universe with the FLRW metric or they should apply the respective coordinate transformations to the smooth background.

Equipped with a metric and an energy-momentum tensor, we can plug equations 2.1 and 2.6 into Einstein's equations

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi T_{\mu\nu}, \tag{2.7}$$

in order to obtain the evolution of a. In equation 2.7 $R_{\mu\nu}$ is the Ricci curvature tensor and R is the Ricci scalar, both obtained from contractions of the Riemann tensor. By solving equation 2.7 one obtains the Friedmann equations (Carroll, 2004)

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \tag{2.8}$$

$$\frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 = 4\pi G\left(\rho + p\right) + \frac{k}{a^2}. \tag{2.9}$$

Equation 2.8 introduces the important quantity $H \equiv \dot{a}/a$ which is called the Hubble parameter and is the observable that describes the expansion (or contraction) of the Universe. The value of the Hubble parameter today in known as the Hubble constant and its value is found to be $H_0 = 72 \pm 8$ km s^{-1}Mpc^{-1} (Freedman et al., 2001). Since the exact value of H_0 usually appears as a constant multiplication or additive factor in calculations it is customary to present results in an "H_0 independent fashion" where $H_0 = 100h$ km s^{-1}Mpc^{-1} and h incorporates the uncertainty in the value of Hubble's constant.

From the two equations 2.8 and 2.9 it is possible to derive a third one called the conservation equation,

$$\dot{\rho} + 3H(\rho + p) = 0, \tag{2.10}$$

which of course is not independent from the two Friedmann equations. Note that equation 2.10 can be directly derived from the Bianchi identities (Carroll, 2004). Therefore there are three unknowns, namely $a(t)$, $\rho(t)$ and $p(t)$ in the two Friedmann equations. In order to solve this system we introduce the equation of state $w = p(t)/\rho(t)$ of the ideal fluid, which relates its density with its pressure. The equation of state for each cosmic fluid describes its behavior, which doesn't have to be monotonic.

We now recast equation 2.8 in the following form, very useful for observational analyses

$$\Omega_m + \Omega_k = 1, \tag{2.11}$$

Table 2.1: The equation of state ($w = p/\rho$) and power law dependence on the scale factor ($\rho = \rho_0 a^q$) for the most important matter and energy sources in Universe.

Source	w	q
matter	0	-3
radiation	$1/3$	-4
vacuum	-1	0
curvature	$-1/3$	-2

where $\Omega_m \equiv 8\pi G\rho/3H^2$ and $\Omega_k \equiv -k/a^2H^2$. For a flat Universe ($\Omega_k = 0$) one obtains the critical density required to achieve that,

$$\rho_{\rm crit} \equiv \frac{3H^2}{8\pi G} = 1.88h^2 \times 10^{-29} \text{ g cm}^{-3}. \tag{2.12}$$

As mentioned in Chapter 1 it is now believed that there is a component in the Universe that behaves very much like the cosmological constant Λ (Carroll, 2001). Including Λ in the Einstein's equations, results of the following Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{2.13}$$

and subsequently equation 2.11 generalizes as

$$\Omega_m + \Omega_k + \Omega_\Lambda = 1, \tag{2.14}$$

where the density parameter for the cosmological constant is given by $\Omega_\Lambda = \Lambda/3H^2$. In equation 2.13 H, a and ρ are the only time-dependent quantities, whereas in equation 2.14 time-dependence has moved to energy densities. The Universe that we live has a non-zero cosmological constant or some other component enters Einstein's field equations with very similar behavior (various alternatives are discussed in Amendola and Tsujikawa, 2010). Similarly it is possible to add a radiation component in the total energy-density of the Universe. Radiation has an equation of state of the form $p_{rad} = 1/3\rho_{rad}$.

The evolution of the scale factor depends on how the various matter components scale with time. Pressureless matter ($p_m = 0$) and radiation dilute as the Universe expands so their densities scale as a^{-3} and a^{-4} respectively. The cosmological constant has by definition a constant density at all times, whereas curvature scales as a^{-2} (from Friedmann equation). We summarize the physical properties of all the popular components that enter the energy-momentum tensor in Table 2.1. Substituting the present values for energy

density sources equation 2.14 becomes

$$H^2 = H_0^2 \left(\Omega_\Lambda + \Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_k a^{-2}\right), \tag{2.15}$$

where H_0 is the present value of the Hubble parameter.

2.2 The cosmological redshift

As the expansion proceeds, the proper distance l, between two positions is modified by the scale factor, $l(t) = l_0 a(t)$. Taking the time derivative of that expression we recover Hubble's law:

$$v = \dot{l} = l_0 \dot{a}(t) = l\frac{\dot{a}}{a} \equiv Hl. \tag{2.16}$$

The expansion of the Universe implies that photons lose energy as they traverse it and therefore when absorbed have larger wavelengths than when emmitted. Formally this can be shown as follows. Firstly for nearby objects ($v/c \ll 1$), their wavelength changes as

$$\frac{\lambda_0}{\lambda} = 1 + \frac{v}{c} = 1 + z \tag{2.17}$$

and so $v = cz$.

For larger distances it is necessary to use the metric (equation 2.1) in order to derive the redshift relation, as curvature effects become important (Liddle, 2003). In fact, the cosmological redshift is an outcome of the FLRW metric itself (Carroll, 2004). This is shown as follows. First, photons travel along null paths and therefore in equation 2.1 we have $ds^2 = 0$. Second, we assume that two photons have been emitted from a distant galaxy at the instances t_1 and $t_1 + dt_1$ and received on our detector at t_0 and $t_0 + dt_0$ then the distances that they travel are

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1-kr^2}} \tag{2.18}$$

$$\int_{t_1+dt_1}^{t_0+dt_0} \frac{dt}{a(t)} = \int_0^{r_1} \frac{dr}{\sqrt{1-kr^2}}. \tag{2.19}$$

Between t_1 and $t_1 + dt_1$ the galaxy positions haven't change much and therefore we can equate the two left hand sides, which gives

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{t_1+dt_1}^{t_0+dt_0} \frac{dt}{a(t)}. \tag{2.20}$$

Rearranging the limits of the integral from t_0 to $t_0 + dt_0$ and from t_1 to $t_1 + dt_1$ equation 2.20 becomes

$$\int_{t_0}^{t_0+dt_0} \frac{dt}{a(t)} = \int_{t_1}^{t_1+dt_1} \frac{dt}{a(t)}, \tag{2.21}$$

which to a zeroth order approximation yields

$$dt_0 \frac{1}{a(t_0)} = dt_1 \frac{1}{a(t_1)} \Rightarrow \frac{dt_0}{dt_1} = \frac{a(t_0)}{a(t_1)}. \tag{2.22}$$

For electromagnetic radiation $dt \propto \lambda$[1], therefore

$$\frac{\lambda_0}{\lambda_1} = \frac{a(t_0)}{a(t_1)}. \tag{2.23}$$

For an expanding Universe $a(t_0) > a(t_1)$ and therefore the photon's wavelength has been stretched because of the expansion. Note that the position of the galaxy emitting the photon hasn't change much between t_1 and $t_1 + dt_1$.

The cosmological redshift is a fundamental quantity in observational cosmology and considerable effort is put towards measuring redshifts of galaxies. It is however rather expensive to obtain using spectra. In Chapter 4 we make use of a small but representative spectroscopic galaxy catalogue, to calibrate *photometric* redshifts for a much larger imaging galaxy catalogue.

2.3 Distances in cosmology and the age of the Universe

The concept of distance in an expanding Universe is considerably more involved than distances in static spacetimes. Nevertheless, apart from the Hubble parameter, distances provide another set of observable quantities, that can determine the properties of the Universe on the background level. In this section we assume $c = 2.998 \times 10^5$ km s^{-1}.

For $z \ll 1$, where Hubble's law applies, the natural Hubble distance is given by

$$D_{\rm H} \equiv \frac{c}{H_0} = 2998 h^{-1} {\rm Mpc}, \tag{2.24}$$

where h as usual parametrizes our uncertainty on the actual value of Hubble's constant. $D_{\rm H}$ is the characteristic distance scale of the expanding Universe.

Using the metric and setting again $ds = 0$ we get $dt^2 = a^2(t)dr^2$, we can define the comoving distance, as

$$d_C = \int \frac{cdt}{a(t)} = \int_0^z \frac{cd\tilde{z}}{H(\tilde{z})} = D_{\rm H} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})}, \tag{2.25}$$

[1] $c = \lambda f \propto \lambda / dt \Rightarrow \lambda \propto dt$

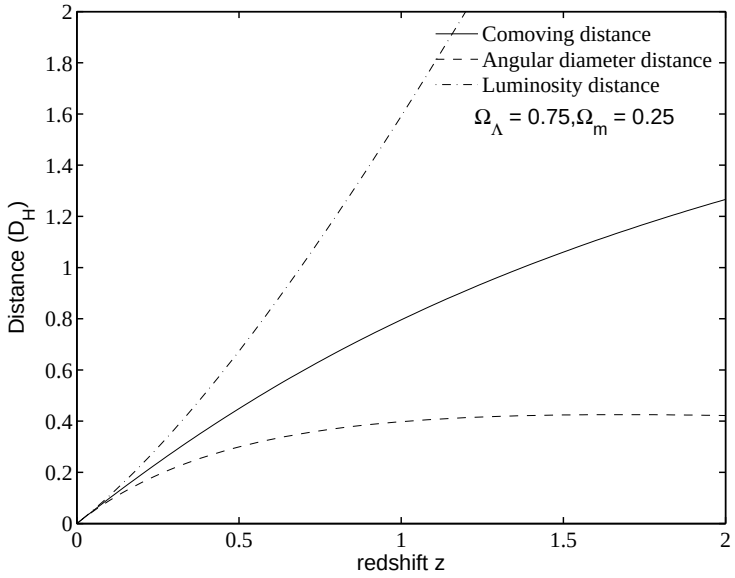

Figure 2.1: The comoving, angular diameter and luminosity distance for typical values of the cosmological parameters (Equations 2.15 and 2.25). Distances are depicted dimensionless, but can be transformed to astronomical units using $D_H = 2998h^{-1}$Mpc. The physical meaning of the turnover of the angular diameter distance is that objects with fixed physical size were apparently larger in past, when the Universe was smaller.

where $E(z) \equiv H(z)/H_0$. For non-flat Universes the comoving distance expression is more involved. Again, using only the radial part of the metric we get the following relation

$$\int_0^{d_C} \frac{dr}{\sqrt{1 \pm kr^2}} = \int_0^z \frac{cd\tilde{z}}{H(\tilde{z})}. \tag{2.26}$$

Performing the integrals yields

$$d_C = \frac{1}{\sqrt{k}} S_k \left[\sqrt{k} D_H \int_0^z \frac{d\tilde{z}}{E(\tilde{z})} \right], \tag{2.27}$$

which is usually expressed in terms of the curvature parameter Ω_k, therefore

$$d_C = D_H \frac{1}{\sqrt{\Omega_k}} S_k \left[\sqrt{\Omega_k} \int_0^z \frac{d\tilde{z}}{E(\tilde{z})} \right]. \tag{2.28}$$

Throughout, the redshifts that we use for the analysis of observational data are always converted to comoving distances. There exist however other convenient distance measures, most notably the luminosity distance d_L (useful for objects with known absolute luminosity and for the apparent - absolute magnitude relation) and the angular diameter distance d_A (useful for objects with known angular extent). These are directly related to comoving distance d_C, via (Carroll, 2004)

$$d_L = (1+z) d_C = (1+z)^2 d_A. \tag{2.29}$$

All three distance measures as a function of redshift for concordance cosmology are shown in Fig. 2.1. We note that fits on luminosity distance obtained from distant supernovae gave the first conclusive evidence of the late time cosmic acceleration (Riess et al., 1998; Perlmutter et al., 1999).

Lastly, one can also define the lookback time simply as the time elapsed as a photon was emitted at redshit z and observed today. Using the metric

$$t_o - t_e = \int_{t_o}^{t_e} dt = \int_{a_e}^{1} \frac{da}{aH(a)} = H_0^{-1} \int_0^z \frac{d\tilde{z}}{(1+\tilde{z})E(\tilde{z})}. \tag{2.30}$$

Thus, for the values of the cosmological parameters that we measure today $t_0 \sim H_0$. By plugging in the values for the cosmological parameters given in Section 2.7, equation 2.30 yields that the time elapsed since the Big Bang is 13.7 billion years.

2.4 Linear perturbation theory

Going beyond the background level, which is solely described by the FLRW metric (equation 2.1), it is also interesting for a number of reasons to study the clumpy Universe. First, galaxies are believed to form inside dark matter haloes which are the result of the primordial inhomogeneities of the dark matter density field. Thus understanding the physics of galaxy formation and evolution depends crucially on understanding the evolution of perturbations in the Universe and vice versa, since we use galaxies to make inferences about the dark matter density field. Second, inhomogeneities are related to the primordial density fluctuations which in turn are likely to be related to inflationary models for the early Universe and the fundamental theories of physics. Third, the cosmological parameters that we measure depend on inhomogeneities and their evolution, for example the larger the matter density in the Universe then the higher the amplitude of the two point correlation function, the statistic that is commonly used to describe the perturbed matter field.

Hot Big Bang cosmology asserts that perturbations are very small at high-z and therefore can they be studied with linear theory. Since structure is seeded from these small perturbations it is reasonable to expect that at sufficiently large scales linear theory provides a good description for our observations today. Cosmological perturbation theory is generally involved because of the gauge invariance (Padmanabhan, 1993). Nevertheless, for analyzing current observations it is sufficient to work in the Newtonian gauge where observers are moving with the Hubble flow and subsequently observe peculiar velocities due to matter falling into structures.

This section aims for the derivation of the relevant cosmological quantities, mainly the linear growth of structure and its time derivative, which are going to be compared with observations in Chapters 5 and 6. It follows closely the discussion from Peebles (1980, § 6, 7, 8, 10).

The Newtonian gauge allows us to use the weak field approximation of general relativity. The 00 component of the perturbed metric gives the Poisson equation and it is given by

$$g_{00} = 1 + 2\Phi, \tag{2.31}$$

where Φ is a small perturbation over the smooth background. The discussion is valid for scales below the Hubble scale $cH_0^{-1} \sim 10^{28}$ cm, but it does implicitly assume that general relativity is valid up to these scales. Then, for an ideal fluid the Poisson equation is given by

$$\nabla^2_{\mathbf{r}}\Phi = 4\pi G(\rho + 3p) - \Lambda. \tag{2.32}$$

Since we are working with observers that move with the Hubble flow it is useful to rewrite equation 2.32 in expanding coordinates. The proper separation $\mathbf{r}$, between two objects is given by

$$\mathbf{r} = a(t)\mathbf{x} \tag{2.33}$$

and their proper velocity $\mathbf{u}$, comes from the time derivative of the previous equation

$$\mathbf{u} = \mathbf{x}\dot{a}(t) + a(t)\dot{\mathbf{x}}. \tag{2.34}$$

In equation 2.34 the two components of cosmic velocity $\mathbf{u}$ emerge, as the first term on right hand side describes the Hubble expansion[2] and the second term describes the peculiar velocity field. The derivative of the peculiar velocity component of equation 2.34 gives the peculiar acceleration

$$\mathbf{g(x)} = \frac{d\mathbf{u}}{dt} + \mathbf{u}\frac{\dot{a}}{a} = -\frac{\nabla\phi}{a}. \tag{2.35}$$

One can then define the new potential, in comoving coordinates, ϕ given by $\phi = \Phi + 1/2a\ddot{a}x^2$ and after this transformation the Poisson equation reads (Peebles, 1980)

$$\nabla^2_{\mathbf{x}}\phi = 4\pi G\rho a^2 + 3a\ddot{a}. \tag{2.36}$$

Hereafter we drop the subscript $\mathbf{x}$, as the derivative will always be understood with respect to comoving separations. For this discussion we also assume that $p = 0$ and $\Lambda = 0$.

[2]Setting $\dot{\mathbf{x}} = 0$ we get $v = \dot{a}x = (\dot{a}/a)\, r = H_0 r$, i.e. Hubble's law.

Equation 2.32 can be easily solved in spherical coordinates for a pressureless ideal fluid and zero cosmological constant

$$\Phi_b = \frac{2}{3}\pi G\rho_b(t)r^2 = \frac{2}{3}\pi G\rho_b(t)a^2(t)x^2. \tag{2.37}$$

In comoving coordinates the equation of motion of a is

$$\frac{d^2a}{dt^2} = -\frac{4}{3}\pi G\rho_b(t)a, \tag{2.38}$$

where the subscript b refers to background quantities. Equations 2.36, 2.37 and 2.38 give the following expression

$$\nabla^2\phi = 4\pi Ga^2\left[\rho(\mathbf{x},t) - \rho_b(t)\right]. \tag{2.39}$$

It is also easier to work with the dimensionless density contrast of the matter field δ, defined as

$$\delta(\mathbf{x},t) = \frac{\rho(\mathbf{x},t) - \rho_b(t)}{\rho_b(t)}. \tag{2.40}$$

Thus, the perturbed background is the source of the gravitational potential ϕ given by the equation

$$\nabla^2\phi = 4\pi Ga^2\rho_b(t)\delta. \tag{2.41}$$

We now need a model for the behaviour of matter in the Universe. A very good guess is an ideal fluid which is governed by the energy conservation equation

$$\left(\frac{\partial\rho}{\partial t}\right)_{\mathbf{r}} + \nabla_{\mathbf{r}}\cdot\rho\mathbf{u} = 0, \tag{2.42}$$

where the time derivative is taken at a fixed point $\mathbf{r}$, and the Euler equation

$$\rho\left[\left(\frac{\partial\mathbf{u}}{\partial t}\right)_{\mathbf{r}} + (\mathbf{u}\cdot\nabla_{\mathbf{r}})\,\mathbf{u}\right] = -\nabla_{\mathbf{r}}p - \rho\nabla_{\mathbf{r}}\Phi, \tag{2.43}$$

where Φ is given from the Poisson equation 2.32. Before we perturb equations 2.42 and 2.43 it is convenient to write them in comoving coordinates. The mass conservation equation in comoving coordinates reads as (Peebles, 1980)[3]

$$\frac{\partial\rho}{\partial t} + 3\frac{\dot{a}}{a}\rho + \frac{1}{a}\nabla\cdot\rho\mathbf{v} = 0, \tag{2.44}$$

where $\mathbf{v} = a\dot{\mathbf{x}}$. In the absence of peculiar velocities ($\mathbf{v} = 0$) equation 2.44 reduces to equation 2.10. Equation 2.44 describes the conservation of matter in comoving coordinates $\mathbf{x}$, but it is however non-relativistic. For small perturbations around the mean background density ρ_b, we use δ from equation 2.40, so equation 2.44 becomes

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot[(1+\delta)\mathbf{v}] = 0. \tag{2.45}$$

[3]In the following we use the transformation $\left(\frac{\partial\rho}{\partial t}\right)_{\mathbf{r}} = \frac{\partial\rho}{\partial t} - \frac{\dot{a}}{a}\mathbf{x}\cdot\nabla\rho$.

Applying the same procedure to the Euler equation (2.43) we obtain

$$\frac{\partial \mathbf{v}}{\partial t} + \frac{\dot{a}}{a}\mathbf{v} + \frac{1}{a}\left(\mathbf{v}\cdot\nabla\right)\mathbf{v} = -\frac{1}{a}\nabla\phi. \tag{2.46}$$

The last step is to combine equations 2.45 and 2.46. This can be achieved by taking the time derivative of the former, the divergence of the latter and then subtract them. The third term on the left hand side of equation 2.46 is a second order term and it can be omitted. The result is

$$\frac{\partial^2\delta}{\partial t^2} + 2\frac{\dot{a}}{a}\frac{\partial\delta}{\partial t} = 4\pi G\rho_b\delta, \tag{2.47}$$

where we also made use of the Poisson equation 2.32. Linearizing equation 2.45 by setting $\mathbf{v}\delta \approx 0$ we obtain the important equation

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\mathbf{v} = 0. \tag{2.48}$$

Equations 2.48 and 2.46 are the final results and they describe the linear evolution of matter δ and velocity fields[4] $\nabla\cdot\mathbf{v}$, as well as their relation in the linear regime. These relations generally hold for small density perturbations $\delta \ll 1$ and for length scales d that satisfy $(\upsilon H_0^{-1}/d)^2 \ll \delta$.

The solutions of equation 2.47 are of considerable importance as they will eventually relate the aforementioned theoretical approach with observations. Moreover, they are cosmology dependent so equation 2.47 can also be used to constrain cosmological parameters. The simplest case arises for the Einstein-de Sitter Universe for which

$$\Omega_{\rm tot} = \Omega_m = 1. \tag{2.49}$$

In this case $a \propto t^{2/3}$ and the solution of equation 2.47 is (Peebles, 1993)

$$\delta = At^{2/3} + Bt^{-1}, \tag{2.50}$$

where A and B are constants and perturbation δ has two independent modes, one growing and one decaying mode. The growing mode evolves simply as $D \propto 1/1+z$. For different cosmologies than Einstein-de Sitter equation 2.50 becomes

$$\delta = A(\mathbf{x})D_1(t) + B(\mathbf{x})D_2(t), \tag{2.51}$$

where $D_1(t)$ and $D_2(t)$ are again linearly independent. Generically there will be a fast growing mode and a slow growing mode. Equation 2.48 suggests that the time derivative

[4]Technically this is the divergence of the velocity field but it is common to refer it as velocity field since this is the quantity that is directly related with the velocity power spectrum.

of D is of considerable importance, since it relates matter and velocity fields. It is usual to work with the dimensionless quantity

$$f_g = \frac{a}{\dot{a}}\frac{\dot{D}}{D} = \frac{d\ln D}{d\ln a}, \tag{2.52}$$

known as the growth rate of structure.

Solutions of equation 2.51 are of course cosmology dependent. For a wide range of cosmologies the approximation $f_g \approx \Omega_m^{0.6}$ is sufficient as the growth rate depends primarily on the matter density. A more accurate approximation is given by Carroll et al. (1992) and a fortran package for numerical integration of equation 2.51 is provided by Hamilton (2001). The growth of structure $D(t)$ is studied in Chapter 5 along with the time evolution of the linear bias, whereas measurements of the growth rate are presented in Chapter 6.

The growth rate of structure $f_g(\Omega_m)$ is a fundamental quantity to test alternative theories of gravity. As shown by Wang (2008) dark energy models with identical expansion history $H(z)$ can be distinguished from measurements of the growth rate as a function of redshift $f(\Omega_m, z)$. This phenomenon is shown in Fig. 2.2, although it should be noted that it is possible to come up with modified gravity models that are also degenerate with general relativity with respect to the growth rate. Since general relativity has only been tested in the Solar system it is important to justify with the best possible evidence its use on cosmological scales. Note that assuming the correctness of GR on cosmological scales it is possible to perform null tests on ΛCDM using the growth rate as shown from the green and black line of Fig. 2.2. Such cross-checks are important because they test the model with a different methodology, as well as different set of assumptions.

Finally, it is worth mentioning that the density field δ can be decomposed into Fourier modes such that $\delta = \sum \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}$. In the linear approximation, there is no coupling between different Fourier modes and therefore each $k-$mode in equation 2.47 evolves independently. This is a strong advantage for the use of the power spectrum, instead of the real space correlation function, but only in the linear regime (Hamilton, 2000).

2.5 Statistical description of cosmological density fields

2.5.1 The fair sample hypothesis

Current observations favour a Universe that is infinite[5]; either spatially flat or open. Yet it is only possible to draw conclusions about the small portion of the Universe which we can observe, i.e. sources whose light had enough time to arrive at our telescopes. The

[5]Assuming a trivial topology

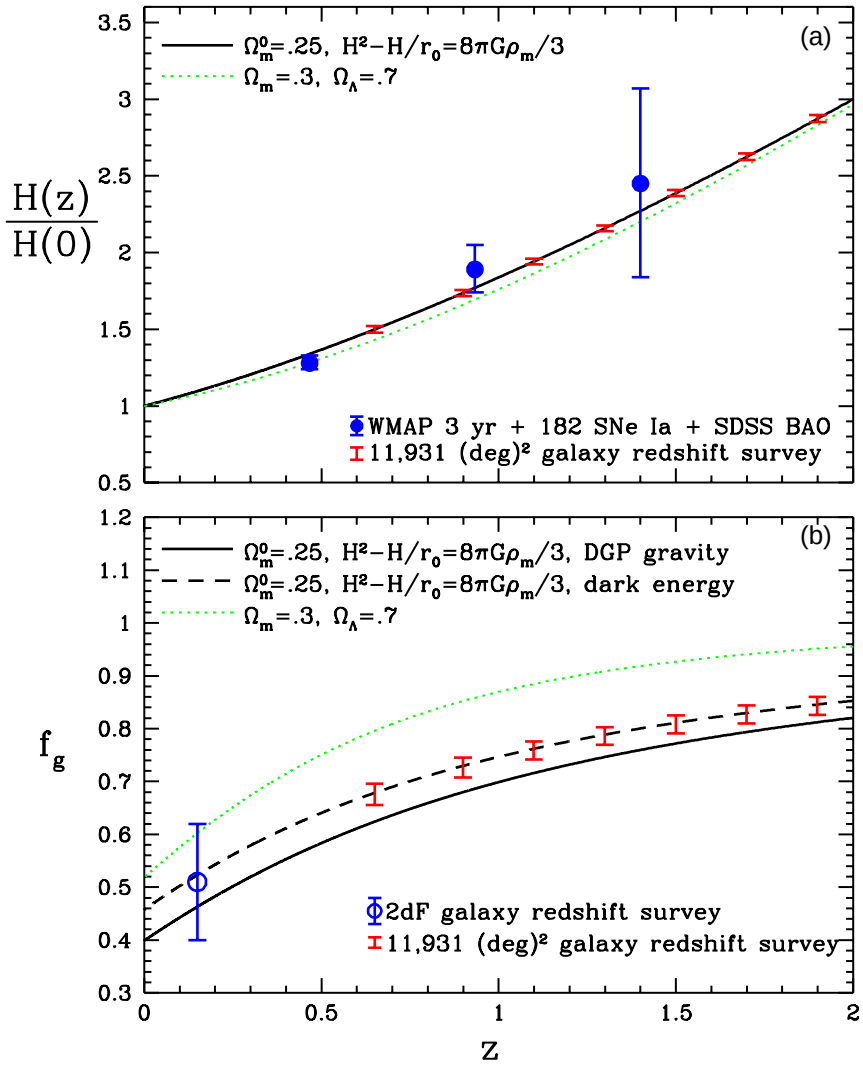

Figure 2.2: The importance of growth rate measurements in cosmology. The above plot shows how the degeneracy in the expansion rate $H(z)$ between different dark energy models (upper panel) is broken using the growth rate $f(z)$ (lower panel). [Figure credit: Wang (2008).]

assumption that the observable Universe is a representative sample of the whole and thus suitable for scientific inferences is known as the fair sample hypothesis (Peebles, 1980, § 30).

The fair sample hypothesis has profound consequences on the way that we study inhomogeneities in the Universe as well as their meaning, because of the averaged quantities that are introduced for their description (Peacock, 1999) e.g. the average density contrast. Observationally we are faced with the matter density field described at each point with the quantity $\delta(\mathbf{x})$, but this is only one realization thereof. Yet if we hypothetically average over all realizations of the density field of the Universe (which can be done using simulations) we will eventually get $\langle\delta(\mathbf{x})\rangle = 0$. As it is impossible to do so, we simply assume that all density perturbations average to zero. A stronger assumption is that the Universe is ergodic and therefore averaging over sufficiently far away regions is equivalent to the ensemble average.

We now proceed with the mathematical definitions of the two point correlation function, the power spectrum and related quantities of interest in cosmology.

2.5.2 Spatial correlation function

The important statistic known as the two point spatial correlation is defined as

$$\xi(\mathbf{r}) = \langle\delta(\mathbf{x})\delta(\mathbf{x}+\mathbf{r})\rangle, \tag{2.53}$$

where $\delta(\mathbf{x})$ is defined in equation 2.40. Alternatively, $\xi(\mathbf{r})$ can be defined as the excess probability, over the Poisson distribution, of finding a pair of galaxies at a separation $\mathbf{r}_{12}$ (Peebles, 1980)

$$dP = n^2\left[1+\xi(\mathbf{r}_{12})\right]dV_1 dV_2, \tag{2.54}$$

where dV_1 and dV_2 are the volume elements in which the two galaxies lie and n is the mean density of galaxies. Homogeneity and isotropy implies that the separation between the two galaxies is described by a scalar quantity, so equation 2.54 becomes

$$dP = n^2\left[1+\xi(r_{12})\right]dV_1 dV_2. \tag{2.55}$$

The definition for the two point correlation function easily generalizes to higher order statistics, such as the three point correlation function and the four point correlation function (Peebles, 1980). However, for a Gaussian field all odd correlation functions vanish and all higher order even correlation functions can be derived from combinations of the two point function. Note that even if the initial conditions of the early Universe

were indeed Gaussian, gravitational clustering would make odd correlation functions non-zero and since we are observing the evolved density field today, we expect to observe a measurable signal of the galaxy three point correlation function (Bernardeau et al., 2002; Gaztañaga et al., 2009). Notwithstanding this caveat, the two point correlation function is clearly fundamental for our understanding of the statistics of the matter density field.

What does the two point correlation function of galaxies look like? To a good approximation it can be parametrized by a power law of the form

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma}, \tag{2.56}$$

where r_0 is the clustering length which describes the strength of the observed clustering (the larger the r_0 the more clustered the galaxies). Alternatively one can use the fact that the shape of the linear dark matter $P(k)$ is known at any redshift from linear theory using either analytical approximations (Eisenstein and Hu, 1998) or numerical calculations (Lewis et al., 2000). One then obtains the configuration space quantity $\xi(r)$ by Fourier transforming $P(k)$ as explained in the next Section. Of course galaxies will be biased and therefore wouldn't follow the exact shape of the the matter correlation function but it is reasonable to expect that on large scales biasing is linear and deterministic. On the other hand on small scales non-linear effects will alter the shape of the linear power spectrum. This effect can be studied by simulations which can provide corrections to the linear power spectrum (Smith et al., 2003).

In Fig. 2.3 we show the linear $\xi^{\rm lin}(r)$ and the non-linear $\xi^{\rm nl}(r)$ with added non-linear corrections on small scales (Smith et al., 2003) for standard cosmology. Overlaid is a fiducial power law approximation, with $r_0 = 5\ h^{-1}$Mpc and $\gamma = 1.8$ (Peebles, 1980). An important feature of the matter correlation function is the baryon acoustic peak which can be observed at comoving separations of $\sim 100\ h^{-1}$Mpc. It presents the sound horizon at the epoch of recombination. Once CMB photons decoupled from the baryons the sound horizon remained frozen and its signal is observed in CMB and large-scale structure measurements (Eisenstein et al., 2005).

The normalization of the two point correlation function, which fixes its amplitude, is usually defined as the density variation in spheres of radius of $8\ h^{-1}$Mpc when smoothed with a top-hat window function

$$\sigma_r^2 = \frac{3}{r^3}\int_0^r \tilde{r}^2 \xi(\tilde{r}) d\tilde{r}, \tag{2.57}$$

where $\xi(r)$ is the linearly extrapolated correlation function at $z = 0$. A traditional value for r in equation 2.57 is $r = 8\ h^{-1}$Mpc and the density variation is denoted as σ_8. It

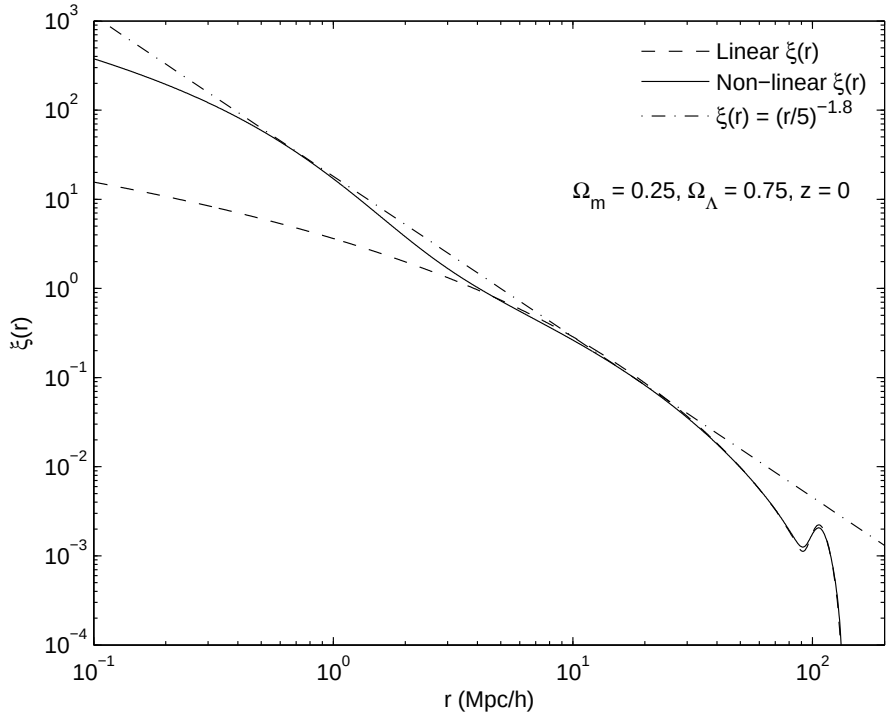

Figure 2.3: Real space dark matter two point correlation function at $z = 0$. Solid line is the non-linear correlation function (Smith et al., 2003), dashed line is the linear correlation function and dotted dashed line a power law obtained from observations. On scales $r > 20\ h^{-1}$Mpc the linear and non-linear correlation functions agree as expected, whereas on smaller scale the fiducial power law provides a good phenomenological description of $\xi(r)$. Linear and non-linear correlation functions were obtained using CAMB (Lewis et al., 2000). For a discussion about the baryon acoustic oscillation peak at $r \sim 100\ h^{-1}$Mpc see text.

follows then that $\xi(r) \propto \sigma_8^2$. Note that in its simplest form galaxy bias is a constant modification of the amplitude of the matter correlation function since $\delta^{(g)}(\mathbf{x}) = b\delta(\mathbf{x})$ and therefore $\xi^{(g)}(r) \propto b^2$. Therefore, in the linear regime the galaxy two point correlation function amplitude depends on both the bias and σ_8. There exist various ways to break this degeneracy. In this thesis we choose to work with the relative bias, the bias of one galaxy population with respect to another, which automatically cancels σ_8, since all co-spatial galaxies trace the same dark matter density field (Chapter 5) or to present constraints on the $b\sigma_8$ combination (Chapter 6).

2.5.3 Power spectrum

The density field can be decomposed in harmonic waves using Fourier analysis as follows

$$\delta = \sum \delta_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{x}}. \tag{2.58}$$

If we plug equation 2.58 into equation 2.53 we obtain the following relation between $\xi(r)$ and δ_k

$$\xi(\mathbf{r}) = \frac{V}{(2\pi)^3} \int |\delta_{\mathbf{k}}|^2 e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k \tag{2.59}$$

and therefore the power spectrum $P(k) = \langle|\delta_{\mathbf{k}}|^2\rangle$ is the Fourier pair of the two point correlation function. Once again, isotropy allows us to use spherical symmetry and reduce the three integrals of equation 2.59 to the following one

$$\xi(r) = \frac{V}{(2\pi)^3} \int P(k) j_0(kr) 2\pi k dk, \tag{2.60}$$

where $j_0(kr) = \sin(kr)/kr$ is the zeroth order Bessel function.

We mention in Section 2.4 that in the linear regime different $\delta_{\mathbf{k}}$ modes evolve independently. Fitting codes are provided (Eisenstein and Hu, 1998) to evolve the matter power spectrum from high redshift to today. However, operationally it is sometimes easier to analyze observations in configuration space. Since measurements of $\xi(r)$ span over many orders of magnitude and r is also expressed with logarithmic spacing, Fourier transforming $\xi(r)$ is cumbersome, as it requires vast numbers of data points [6]. Yet it is necessary to do so for comparing data with theory. Fortunately, Hamilton (2000) has presented an analytical solution to this problem and has implemented it in a publicly available code that we use extensively throughout.

[6] Usual FFT methods require set of data with evenly space points (Press et al., 1992), whereas the correlation function and power spectrum span over several orders of magnitude making the problem impractical.

2.5.4 Angular correlation function

In an analogous way with the spatial correlation function one can define the angular correlation function. The excess of clustering for objects that lie on the surface of a sphere is given by the two dimensional analogue of equation 2.55

$$dP = \bar{n}^2 \left[1 + w(\theta_{12})\right] d\Omega_1 d\Omega_2, \tag{2.61}$$

where $d\Omega_1$ and $d\Omega_2$ are the the solid angle elements enclosing the two galaxies respectively and n is the mean surface density of galaxies, such that $\bar{n} = N/\Omega$.

The angular correlation function is just the projection of the spatial correlation function on a sphere. Thus, if we know radial distribution of sources dn/dr, we can project $\xi(r)$ to $w(\theta)$ as follows

$$w(\theta) = \int_0^\infty \frac{dn_1}{dr_1} dr_1 \int_0^\infty \frac{dn_2}{dr_2} \xi\left(R, \frac{r_1 + r_2}{2}\right) dr_2, \tag{2.62}$$

where $R = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta)}$, is the cosine rule. Written in this form, equation 2.62 gives the angular correlation between two populations with different radial selection functions and it also makes the implicit assumption that $\xi(r)$ doesn't evolve in the overlapping volume of dn_1/dr_2 and dn_2/dr_2. At low-z, it is rare to have a situation where the spatial correlation function evolves considerably to violate that assumption.

In fact, it is possible to simplify the considerably involved double integral of equation 2.62 if we introduce some further assumptions (Limber, 1953; Simon, 2007). The first assumption is that dn/dr is only a function of $\bar{r} = (r_1 + r_2)/2$. The second assumption is that $\xi(r)$ varies significantly only across the separations $\Delta r = r_2 - r_1$. Finally the third assumption is a geometrical one and implements the well-known small angle approximations $\sin\theta \approx \theta$ and $\cos\theta \approx 1$. The result is known as Limber's equation

$$w(\theta) = \int_{-2\bar{r}}^{2\bar{r}} \frac{dn_1}{d\bar{r}} \frac{dn_2}{d\bar{r}} d\bar{r} \int_0^\infty \xi\left(\sqrt{\bar{r}\theta^2 + \Delta r^2}, \bar{r}\right) d\Delta r. \tag{2.63}$$

Equation 2.63 can be used for any $\xi(r)$. In most conceivable cases, the second integral is over all pair separations and a sensible upper limit is of order of 100 Mpc, nevertheless it can also be defined from $-\infty$ to ∞ since in any case the weight functions dn/dr will fall rapidly outside the survey limits. Changing the limits does have some mathematical advantages for analytically evaluating the inner integral of equation 2.63 for a power law $\xi(r)$ given by equation 2.56 (Peebles, 1980, § 52). Switching to the $\bar{r}$ and Δr pair of variables and making use of the distance-redshift relation $d\bar{r} = D_H dz/E(z)$ (equation 2.25) to facilitate with observational radial selection functions which are expressed in

cosmological redshift, equation 2.63 becomes

$$w(\theta) = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \frac{\int \left(\frac{dN}{dz}\right)^2 \frac{dz}{d\bar{r}} \bar{r}^{1-\gamma}(z) F[\bar{r}(z)] r_0^{\gamma}(z) dz}{\left[\int \left(\frac{dN}{dz}\right) dz\right]^2} \theta^{1-\gamma}, \tag{2.64}$$

where $F(x) = 1 - kx^2$ and we assumed that the galaxy pairs come from the same selection function dN/dz. Note that in equations 2.63 and 2.64 $\bar{r}$ is the comoving distance at a given redshift.

In conclusion, if the spatial correlation function is a power law, then the angular correlation function is also a power law with a shallower slope and is given by the equation

$$w(\theta) = A_w \theta^{1-\gamma}, \tag{2.65}$$

where A_w is a constant, given from the long pre-factor of the $\theta^{1-\gamma}$ term in equation 2.64 and γ the same parameter that appears in equation equation 2.56. In practice, the angular correlation function in the form of equation 2.63 is the observable and we are interested in extracting the real space correlation function in the form of equation 2.56. From equation 2.64 we can extract the correlation length immediately, under the assumption that it doesn't vary significantly within the dN/dz redshift range, such that $r_0(z_{\rm eff}) = r_0(z)$. In Chapter 5 we make use of this assumption to invert our angular clustering signal to spatial. The full inverted equation can be obtained if we solve equation 2.64 with respect to $r_0(z_{\rm eff})$

$$r_0(z_{\rm eff}) = A_w^{\frac{1}{\gamma}} \left[\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)} \frac{\int \left(\frac{dN}{dz}\right)^2 \frac{dz}{d\bar{r}} \bar{r}^{1-\gamma}(z) F[\bar{r}(z)] dz}{\left[\int \left(\frac{dN}{dz}\right) dz\right]^2} \right]^{-\frac{1}{\gamma}}. \tag{2.66}$$

The Limber (1953) approximations are only valid under the previous stated assumptions. Once these breakdown the clustering results are biased as shown by Simon (2007) and Crocce et al. (2011). However, for the study presented in Chapter 5 these assumptions are still valid and any potential biases are well below our statistical and systematic uncertainties.

Historically, the angular correlation function was the only way of analyzing data from galaxy surveys and an early account of the work done in this area is given by Peebles (1980). With advent of automated spectrographs it became easier to work with the spatial correlation function. Nevertheless, the need for statistical power in cosmological studies has not been eliminated yet and combined with the continually increased use of photometric redshifts, which come with significant uncertainties (Chapter 4), the angular correlation function statistic is still a necessary tool for extragalactic astronomy.

2.5.5 Measuring the angular correlation function

We now describe how we measure the angular correlation function in galaxy surveys. Although the discussion is limited to angular clustering it can be easily generalized to spatial clustering as well. From the definition of the angular correlation function (equation 2.61) we are interested in the number of coherent pairs that are separated by a given angular separation $\theta \rightarrow \theta + \delta\theta$. Operationally this is given from the relation

$$\theta = 2\sin^{-1}\left(\frac{r}{2}\right), \tag{2.67}$$

where r is the cartesian distance between the two objects on the unit sphere. We then compare that number with the expected number of pairs in the absence of any clustering. As survey areas are typically complicated it is significantly easier to find the expected number of pairs by Monte-Carlo integration: We generate random realizations of the data, which cover the exact survey area, but do not have any clustering signal. Then we simply divide the number of data pairs with the number of unclustered points, appropriately normalized for the expected number of pairs, in order to get an estimate for the angular correlation function

$$1 + w(\theta) = \frac{N_{\rm DD}(\theta)}{N_{\rm RR}(\theta)} \frac{n_{\rm R}(n_{\rm R}-1)}{n_{\rm D}(n_{\rm D}-1)} = \frac{DD(\theta)}{RR(\theta)}. \tag{2.68}$$

It is important to note that random points should have the exact angular coverage of the real data. The effects on any areas in the sky obscured by stars, or affected by bad seeing conditions should be reflected accurately in the random points catalogue. These issues are carefully addressed, when we present our angular correlation function measurements in Chapter 5.

Alternative versions of equation 2.68 have been proposed which improve its performance. Davis and Peebles (1983) proposed the following estimator

$$1 + w(\theta) = \frac{N_{\rm DD}(\theta)}{N_{\rm DR}(\theta)} \frac{2n_{\rm D}n_{\rm R}}{n_{\rm D}(n_{\rm D}-1)} = \frac{DD(\theta)}{DR(\theta)}, \tag{2.69}$$

and Hamilton (1993) introduced another one which carefully cancels the uncertainty of the mean galaxy density at first order

$$1 + w(\theta) = \frac{N_{\rm DD}(\theta)N_{\rm RR}(\theta)}{N^2_{\rm DR}(\theta)} = \frac{DD(\theta)RR(\theta)}{DR^2(\theta)}. \tag{2.70}$$

Finally Landy and Szalay (1993) proposed the following estimator, which they showed presents minimum variance compared to the others and account better for the survey edge effects

$$w(\theta) = \frac{N_{\rm DD}(\theta)}{N_{\rm RR}(\theta)} \frac{n_{\rm R}(n_{\rm R}-1)}{n_{\rm D}(n_{\rm D}-1)} - 2\frac{N_{\rm DR}(\theta)}{N_{\rm RR}(\theta)} \frac{n_{\rm R}(n_{\rm R}-1)}{2n_{\rm D}n_{\rm R}} + 1 = \frac{DD(\theta)}{RR(\theta)} - 2\frac{DR(\theta)}{RR(\theta)} + 1. \tag{2.71}$$

In practice, the differences between the Hamilton (1993) and the Landy and Szalay (1993) estimators are almost unnoticeable (Chapter 5.)

As calculating the correlation function is an N^2 algorithmic process, big samples can make estimates of the correlation function very time consuming. In Chapter 5 we overcome this problem by binning galaxies before the pair-counting, such that we only consider galaxies in the same or neighbouring bins in every iteration. This makes the number of calculations to scale approximately as $N\log N$.

Finally, measurements of the angular correlation function statistic are strongly correlated (Tegmark et al., 2002; Cabré et al., 2007; Norberg et al., 2009), despite the fact that they have by construction perfect angular window functions. In Chapter 5 we construct robust covariance matrices which we use for the power law fits that we perform.

2.5.6 Uncertainty estimation

There are generally two methods for estimating uncertainty on clustering measurements and the discussion is somewhat related with the fair sample hypothesis. The fundamental problem in uncertainty estimations in clustering measurements in cosmology is that we are able to observe only one realization of the density field. The best way to overcome this is the use of simulations, which require knowledge of the initial conditions and a rough estimate of the cosmological parameters. Therefore, by creating many realizations of some volume of the Universe we are in a position to quantify statistically the likelihood of the observed parameters. Of course simulations can never include all elements of reality and this issue is even more pressing for galaxy formation and evolution. The alternative method is to use the data itself to construct repeating measurements. Two popular methods which exploit the data for uncertainty estimations are bootstrap and jackknife. For the bootstrap method one splits the data in equal subsets and then randomly removes a subset and then substitutes it with another one from the data. By repeating this process one creates many quasi-indepent realizations of the data. The second method is jackknife, where the data is split again in N equal subsets but now one removes a part and simply treats the remaining $N-1$ subsets as a new realization.

The question of which method is more reliable for uncertainty estimation inevitably depends on the data at hand and the nature of the problem. Thus for the two main datasets that we use in this thesis we follow different approaches: jackknife for the studying the clustering from SDSS and mocks for the studying the clustering from GAMA. Nevertheless, there have also been systematic tests of the uncertainty estimation methods themselves

which serve as guides at the beginning of the scientific analysis, as well as warnings for the limitations of the uncertainty method used (Cabré et al., 2007; Norberg et al., 2009).

2.6 Halo modeling of large-scale structure

There exist various approaches to model the matter and galaxy correlation function beyond the linear regime. They are broadly divided into two categories; the perturbation theory approach (Bernardeau et al., 2002), which is more useful for cosmological applications (e.g. Sánchez et al., 2009) and the halo modeling approach (Cooray and Sheth, 2002), which is more useful from a galaxy evolution point of view (Zehavi et al., 2011, and references therein). Although halo modeling is a more phenomenological approach, it is superior in describing small to intermediate scale clustering, which is the regime where most galaxy surveys can probe robustly. The data presented and analyzed in this thesis are more relevant to the halo modeling of large scale structure, but our comparisons with specific models are mostly qualitative. In this Section we introduce the characteristic features of the halo modeling of large-scale structure, known as halo occupational distribution (HOD).

Since theory suggests that galaxies are formed inside dark matter haloes the first quantity to introduce is the differential comoving number density of gravitationally bound objects $dn(M,z)/dM$ given their virialized mass M, known as the halo mass function (HMF). A classic result is the Press and Schechter (1974) formalism which reads

$$\frac{dn(M,z)}{dM} = \frac{\bar{\rho}}{M} f(\nu) \frac{d\ln\sigma^{-1}(M)}{dM}, \tag{2.72}$$

where $\bar{\rho}$ is the background density of the Universe and $\nu \equiv \delta_{\rm sc}(z)/\sigma(M)$. For Einstein-deSitter cosmology $\delta_{\rm sc}(z=0) = 1.686$ and it is weakly Λ-dependent (Cooray and Sheth, 2002). $\sigma^2(M)$ is the variance in spheres given by equation 2.57 by substituting the virial radius corresponding to halo of mass M, usually evaluated when the halo is ≈ 200 times denser than the background (Cooray and Sheth, 2002)

$$R_{\rm vir} = \left(\frac{3M}{4\pi 200\bar{\rho}}\right)^{1/3}. \tag{2.73}$$

In the Press-Schechter model $f(\nu)$ in equation 2.72 is given by

$$f(\nu) = \sqrt{\frac{1}{2\pi}} e^{-\frac{\nu^2}{2}}. \tag{2.74}$$

The Press and Schechter (1974) formalism doesn't agree well with simulations at the high and low end of the HMF (Jenkins et al., 2001). In the literature there have been proposed

more accurate analytical mass functions (Sheth and Tormen, 1999), as well as mass functions calibrated from simulations (Tinker et al., 2008) which have been shown to cover a wide range of cosmologies and redshifts.

The second ingredient of the halo formalism is the density profile of dark matter haloes. Again this quantity is usually calibrated from simulations and it is therefore a prediction of CDM. A functional form, obtained from CDM simulations has been provided by Navarro et al. (1996)

$$\rho(r|M) \propto \frac{1}{\left(\frac{r}{r_s}\right)\left(1+\frac{r}{r_s}\right)^2}, \tag{2.75}$$

where $r_s = r_{\rm vir}/c$ and c is the concentration parameter, which depends only on the mass of the halo. Equation 2.75 is known as the NFW profile.

Lastly, one needs a recipe to populate dark matter halos with galaxies. It is natural to expect that the brightest and most massive galaxy will reside at the center of its halo and that less massive satellite galaxies will live in the outskirts. Moreover, the mass of the halo will determine how bright the central galaxy can possibly be. On the other hand the satellites are usually taken to have a power law dependence on the mass of the halo. Combined these two terms define the halo occupational distribution number

$$\langle N \rangle_M = \langle N_{\rm cen} \rangle_M + \langle N_{\rm sat} \rangle_M, \tag{2.76}$$

which is the expected number of galaxies for a halo of a given mass M.

To sum up, the HOD formalism is described by three main quantities. The HMF, the NFW profile and the galaxy halo occupation number. One then constructs appropriate convolutions to describe various statistical quantities, like the two- and higher order correlation functions (Cooray and Sheth, 2002). Due to these convolutions the HOD formalism is usually expressed in Fourier space. However, Berlind and Weinberg (2002) provided useful empirical (phenomenological) relations of the HOD formalism in configuration space. The galaxy correlation function has two components, the intra-halo galaxy pairs and the pairs of galaxies which reside in different haloes

$$\xi = 1 + \xi_{\rm 1h} + \xi_{\rm 2h}. \tag{2.77}$$

The extra unity arises from the fact that all terms in equation 2.77 are proportional to the total number of galaxy pairs that they describe (see equation 2.54). In the Berlind and Weinberg (2002) model the HOD number weighted with the HMF enters in the one halo term

$$1 + \xi_{\rm 1h}(r) = \frac{1}{2\pi r^2 \bar{n}_{\rm g}} \int \frac{dn}{dM} \frac{\langle N(N-1) \rangle_M}{2} F(r|M) dM, \tag{2.78}$$

where $\bar{n}_{\rm g}$ is the mean density of galaxies and $F(r|M)$ is the distribution of separations of the galaxy pairs in the halo, which usually follows the density run of the NFW profile. Halo modeling provides formulae for the deterministic bias, which connect the two-halo term with the underlying dark matter density film. However, the two-halo term can be approximated reasonably well by the biased linear correlation function (Fig. 2.3)

$$\xi_{\rm 2h}(r) \simeq b_{\rm eff}^2 \xi^{\rm lin}(r), \tag{2.79}$$

where $b_{\rm eff}$ is an effective bias term which consists of the dark matter halo deterministic bias (Cooray and Sheth, 2002) and the galaxy bias with respect to dark matter.

2.7 Cosmological parameters

Despite some warnings in the past that the FLRW metric doesn't provide any characteristic quantities (Peebles, 1980, § 5) or too few as in Sandage's two numbers (h, Ω_m, Sandage, 1961) quest, it is now widely accepted that observational cosmology can constrain a wide range of physical parameters, as well as objectively selecting between different cosmological models (Lewis and Bridle, 2002; Tegmark et al., 2006). Moreover, cosmological parameters whose presence wasn't suspected in past, like the densities of dark energy and dark matter, have been discovered and are now extremely relevant. In this Section we summarize the status of the cosmological parameters and its relation with galaxy evolution models and this work in particular.

The simplest inflationary, hot big bang, cosmological model has the following free parameters set $\boldsymbol{\theta}$ (Tegmark et al., 2006; Lahav and Liddle, 2010),

$$\boldsymbol{\theta} \equiv (h, \Omega_{\rm tot}, \Omega_\Lambda, \Omega_m, \Omega_b, \Omega_\nu, w, n_s, r, \alpha, A_s, \tau, b). \tag{2.80}$$

First is the hubble parameter h, usually measured from distance measurements in the local Universe. $\Omega_{\rm tot}$ is the total energy and matter budget of the Universe. Spatial curvature in FLRW Universes can be found from the relation $\Omega_k = 1 - \Omega_{\rm tot}$. Ω_Λ and Ω_m give the energy density of dark energy and dark matter. Ordinary, baryonic matter is given from Ω_b and neutrino density from Ω_ν. For a constant dark energy equation of state the relevant parameter is $w = p_\Lambda/\rho_\Lambda$. For the possibility of a time-varying dark energy a popular (but not necessarily the only) parametrization adds one more parameter w_a, via the equation $w(a) = w_0 + w_a(1-a)$. The parameters which describe the initial spectrum of fluctuations, associated with the inflationary era of the Universe are the spectral index n_s, the tensor to scalar ratio r and the amplitude of fluctuations A_s (which fixes σ_8). τ

is the reionazation optical depth. The last parameter in equation 2.80 is the galaxy bias b, which describes the different clustering between different galaxy populations and dark matter. We present detailed measurements of this parameter as a function of galaxy type and redshift in Chapter 5.

The progress on the measurements of the parameters of equation 2.80 in the last 20 years has been remarkable. WMAP and SDSS along with other extragalactic observational campaigns placed precise values on the cosmological parameters. However the issue is complicated due to the degeneracies between them, which requires one either to open the parameter space, beyond what one can constrain or to make extra assumptions about their properties. Fortunately CMB two-point statistics, along with two-point large-scale structure statistics[7] and supernovae distance measurements suffice to break most of these degeneracies (Tegmark et al., 2006; Spergel et al., 2007). The result is the emergence of a standard cosmological model known as flat ΛCDM.

For galaxy clustering in the local Universe the most interesting cosmological parameters are those which define the background expansion, namely Ω_Λ, Ω_m and h. Of particular interest is also σ_8 which sets the amplitude of fluctuations and therefore the amplitude of the power spectrum which also correlates with Ω_m. Due to the fact that not all baryonic mass has turned into galaxies Ω_b is not of particular interest in galaxy clustering, with the exception of baryon acoustic oscillations which have their amplitude modulated by the baryon density before the galaxy formation epoch (Eisenstein et al., 2005). For flat ΛCDM we have by definition $\Omega_{\rm tot} = 1$ and $w = -1$. In table 2.2 we quote the observed values from Reid et al. (2010) that are most relevant for this thesis. In Fig. 2.4 (adopted from Reid et al., 2010) we show the joint constraints on Ω_Λ, w and Ω_m. Throughout, we keep the dependance on h and σ_8 explicit, unless otherwise stated.

The quest for measuring the cosmological parameters has now reached an interesting point. In extragalactic astronomy, for all practical purposes, there is no need for more accurate values of $\boldsymbol{\theta}$, apart perhaps from the bias. As an example, changing Ω_m from 0.25 to 0.3, in concordance cosmology, changes distances at $z = 0.4$ (the outer limit for our galaxy samples) by 1.4 percent, which is a tolerable error for most extragalactic applications. This distance error translates into 0.06 magnitude error, which is also acceptable. On the other hand, models of the early Universe, as well as models of the late time acceleration of the Universe, require even more precise measurements, as well as understanding of the systematics, to discriminate the best available model. Nevertheless, one can already notice

[7]Usually this is just BAO measurements, which have less dependance on non-linear physics (Eisenstein et al., 2005; Percival et al., 2010)

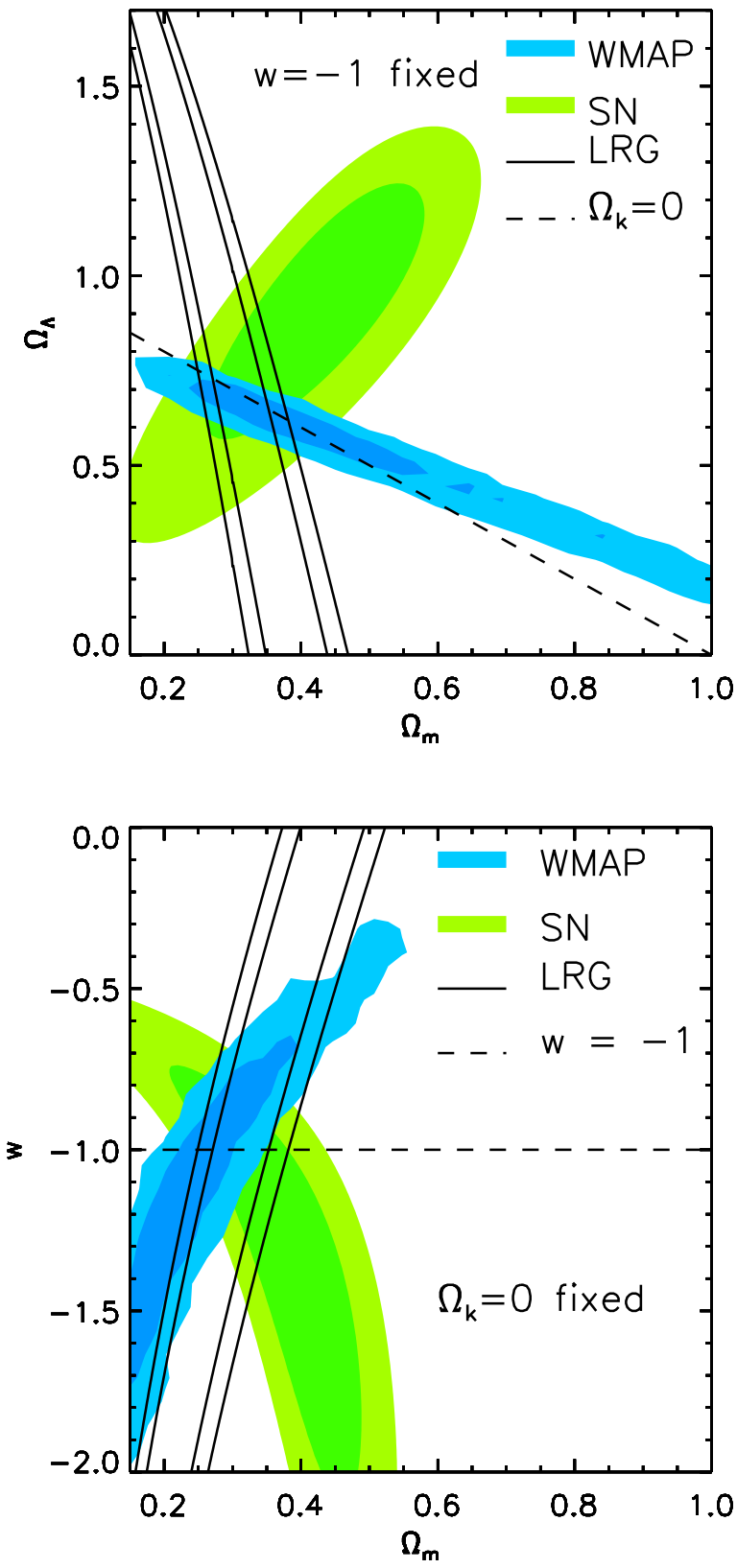

Figure 2.4: Upper panel: Joint constraints on Ω_Λ and Ω_m, from CMB (WMAP), supernovae (Union data set) and large-scale structure measurements (SDSS DR7, LRG's), assuming the concordance ΛCDM model. Lower panel: Joint constraints on w and Ω_m using the same data. Note the consistency between the three contours in both plots, evident from their common intersection. [Figure credit: Reid et al. (2010).]

Table 2.2: Cosmological parameter values from combining measurements from CMB (WMAP), supernovae (Union data set) and large-scale structure measurements (SDSS DR7, LRG's), assuming the concordance ΛCDM model. For a description of the parameters see text. The values are taken from Reid et al. (2010).

Parameter	
h	69.4 ± 1.6
Ω_m	0.289 ± 0.019
Ω_Λ	0.711 ± 0.019
$\Omega_b h^2$	0.02272 ± 0.00058
σ_8	0.824 ± 0.025
Age (Gyr)	13.73 ± 0.13

from Fig. 2.4 that in both panels the three likelihood contours have a common intersecting point, which is powerful consistency check for the standard ΛCDM model.

Whilst the enterprise of measuring the cosmological parameters is blooming and its results are very likely to shape the future of cosmology it is important to check our results and their consistency with a variety of methods. A cautionary article by Croft and Dailey (2011) noted that there are fewer measurements of Ω_Λ that are 2σ away from the "fiducial" WMAP results, than we would statistically expect. If there aren't any unknown correlations in the various datasets, an assumption that is nonetheless very likely to be invalid, then this ought to be a cause of some concern.

Lev Landau famously once said that "cosmologists are often in error but never in doubt". The transition from precision (number of significant figures) cosmology to accurate (take hold of the systematics) cosmology is cosmologists's new mission as described by Peebles (2002). The potential danger if we neglect the latter is clear: 1 per cent bias in a distance measure such as BAO at $z = 1$ translates to 4 per cent error in the value of w (Angulo et al., 2008). A similar situation might apply for supernovae measurements (White, 2007). Therefore ruling out ΛCDM might prove to be a much more difficult task than its establishment, but one can be optimistic that this new route that we may travel will be beneficial for science in general.

Chapter 3

Data

In this Chapter we describe the datasets used in this thesis. Firstly, we describe the public Data Release 7 from the SDSS collaboration (Section 3.1), which we use to study galaxy clustering with photometric redshifts. We then describe the proprietary data from the GAMA collaboration (Section 3.2), which is selected from a previous SDSS release. In Section 3.3 we introduce the first GAMA mocks catalogues, which are used for error analyses. We finish with an overview of some of GAMA scientific results so far (Section 3.4).

3.1 The Sloan Digital Sky Survey

The Sloan Digital Sky Survey II (SDSS, York et al., 2000; Abazajian et al., 2009) is a spectroscopic and multi-band imaging survey which covers over a quarter of the sky. As such, it is the biggest astronomical survey ever accomplished with a variety of successful scientific results ranging from the discovery of new asteroids in the solar system to distant quasars at redshift $z \sim 5$ and from a detailed statistical description of the large-scale structure at redshift $z \sim 0.4$ to a systematic classification of galaxies and their correlations in the local universe.

Data is taken in five photometric bands, known as u, g, r, i and z. In each band the detected flux f, is then transformed into the corresponding apparent magnitude m, through the relation (Lupton et al., 1999)

$$m = -\frac{2.5}{\ln 10}\left[\sinh^{-1}\left(\frac{f/f_0}{2b}\right) + \ln b\right], \tag{3.1}$$

where f_0 is the zero point magnitude ($m = 0$) and b is a softening parameter, approximately one standard deviation noise of the sky. Equation 3.1 has the nice property of reducing to the traditional magnitude relation $m = -2.5\log_{10}(f/f_0)$, for large values of f/f_0, whereas

it is linear in f/f_0 for $f/f_0 \lesssim 1$. The linearity for low values of f/f_0 allows for well defined error distributions on noise-dominated measurements, whereas the $\sinh^{-1}$ function allows for formally negative values of flux. Photometric calibration for SDSS data releases prior to the seventh data release (DR7) was uniform to a 3 percent level (Stoughton et al., 2002) and then shrunk to 1 percent with the implementation of the übercal photometry (Padmanabhan et al., 2008), which we use in this work.

SDSS uses a variety of photometric flux calibration techniques and provides various magnitudes[1] for galaxies and stars. Along with the so called model magnitudes (corresponding to whichever of a de Vaucouleurs or exponential profile provides a better fit to the observed galaxy light profile), SDSS also provides slightly modified Petrosian magnitudes (Petrosian, 1976), which are more appropriate to characterize extended objects like galaxies. As galaxies are extended objects without sharp edges, galaxy samples are selected on Petrosian magnitudes. Colours on the other hand, are usually defined using model magnitudes. SDSS photometry is described in detail in Stoughton et al. (2002). As an input in the photo-z software we also use the Petrosian radii of objects which are defined as the radii that enclose 50% and 90% of the total Petrosian flux. As such, Petrosian radii are a proxy of the distance of the galaxies. The SDSS database also provides corrections for dust attenuation, known as reddening corrections (Schlegel et al., 1998). These are particularly important for photo-z studies. We include these corrections when we extract our sample from the database.

Magnitudes are being used to define objects as galaxies. For the spectroscopic follow up used for the construction of its main galaxy sample, SDSS implements the following main criterion for all objects with $r < 17.77$ (Strauss et al., 2002)

$$m_{\mathrm{PSF},r} - m_{\mathrm{model},r} > 0.3. \tag{3.2}$$

Note that this criterion was loosened up to 0.24 in SDSS DR2 (Abazajian et al., 2004). For the photometric galaxy sample, which we use in this work, the star-galaxy separation criterion is (Stoughton et al., 2002)

$$\sum_i m_{\mathrm{PSF},i} - \sum_i m_{\mathrm{model},i} > 0.145, \tag{3.3}$$

where i runs over the 5 SDSS bands and the sum is actually performed using the fluxes. All magnitudes are de-reddened according to Schlegel et al. (1998).

SDSS's detailed mapping of local galaxies has also revealed the biggest structure ever observed in the universe. It's a supercluster at redshift $z \sim 0.1$ and it was given the

[1] `http://www.sdss.org/dr7/algorithms/photometry.html`

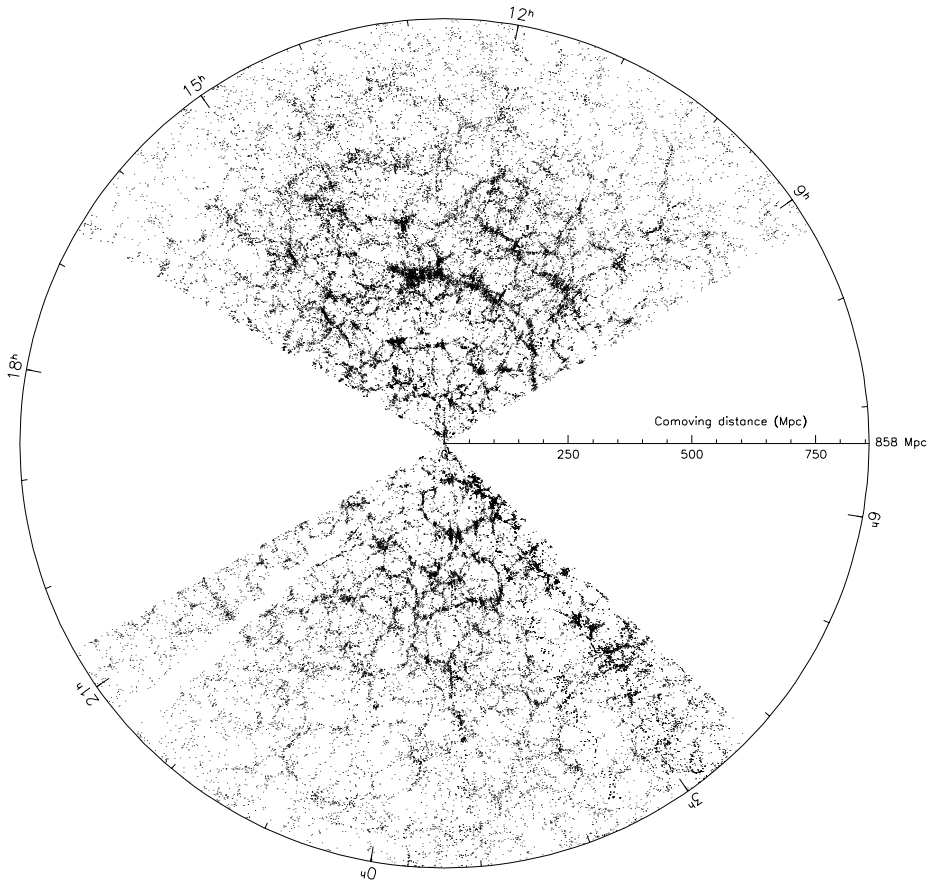

Figure 3.1: The SDSS galaxies footprint. Plotting galaxies as function of their angular and radial position reveals the rich large-scale structure of the low redshift universe. The Sloan Great Wall, the biggest observed so far can be seen at 12h and at a distance of about $\sim$ 300Mpc from us. [Figure credit: Gott et al. (2005).]

name The Great Sloan Wall (Gott et al., 2005) and its position is shown in Fig. 3.1. Notwithstanding their scientific value as such, superstructures also introduce complications in error estimates using internal error estimation methods like jackknife (Norberg et al., 2009, 2011). We address this issue with respect to our clustering measurements from SDSS in Chapter 5.

At the end of its campaign SDSS II has accumulated more than 1 million galaxy spectra with $r < 17.7$ and ~ 5 million galaxies in its photometric catalogue with $r < 19.4$ (henceforth r without a subscript will denote the Petrosian magnitude of a galaxy). Conveniently, a data repository website[2] is provided by the SDSS collaboration for querying and downloading desired datasets, as well as performing appropriate computations and selection cuts. We take advantage of this vast amount information to study galaxy clustering with photometric redshifts aiming to improve and extend the work that has been done by the SDSS team.

3.2 Galaxy And Mass Assembly survey

Galaxy And Mass Assembly[3] (GAMA) is one of the first new generation of galaxy surveys, strategically designed to tackle cutting edge problems in extragalactic astronomy that have the potential to revolutionize physics.

GAMA's main science goal is to provide measurements of the halo mass function (HMF) for the redshifts range $0 < z < 0.5$. The HMF,described in Section 2.6 is a fundamental quantity for dark matter physics and it is therefore crucial for our understanding of structure formation and evolution in the universe. Furthermore, as by the design of the survey galaxies will directly be allocated in their parental dark matter haloes, GAMA will provide invaluable information on galaxy formation as a function of environment.

Other associated science goals of GAMA are to study galaxy evolution, using available multiwavelength observations from other surveys which cover the GAMA fields and to measure the merger rate of galaxies as a function of galaxy type. Also, more relevant to this thesis, GAMA aims to provide measurements of the anisotropic galaxy clustering, which is a promising route for testing models of modified gravity and will provide improved photometric redshift estimates, using its unique galaxy catalogue as a training set.

By virtue of its design, GAMA lies right in between wide-field low redshift surveys and deep high redshift surveys (Fig. 3.2). This will primarily allow GAMA to probe groups of

[2] `http://casjobs.sdss.org/CasJobs/`

[3] `http://www.gama-survey.org`

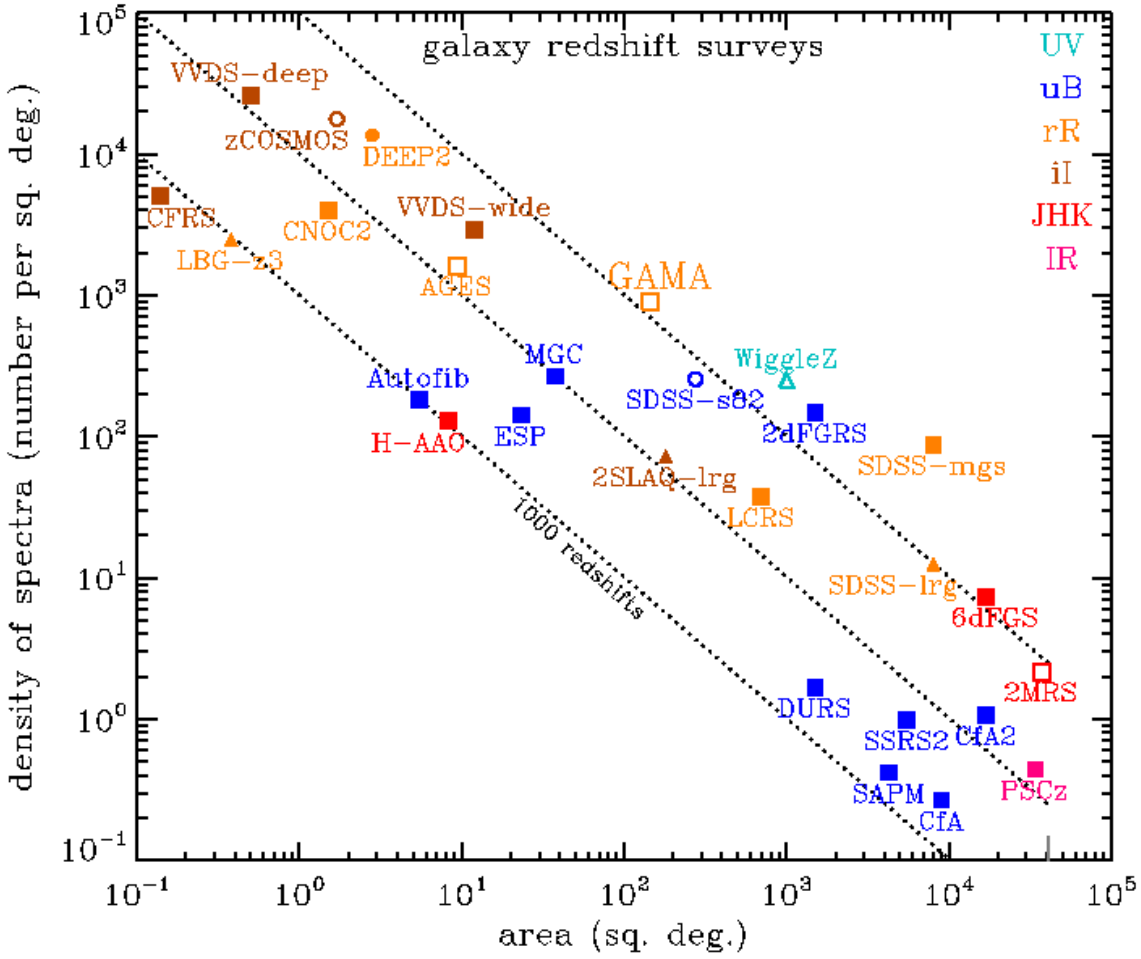

Figure 3.2: Graphic overview of spectroscopic galaxy surveys. Each survey is shown as a function of area coverage and density of spectra. Lying at the middle of the graph, GAMA fills an important gap. [Figure credit: Baldry et al. (2010).]

mass similar to or less than that of the local group ($M \sim 10^{12} M_{\odot}$), something which no other survey has been able to achieve so far.

In this thesis we use data from the completed GAMA I, whilst observations for GAMA II are on their way. GAMA I fields consist of three 4×12 deg^2 regions, named G09, G12 and G15, all of which lie almost on the equator. G09 and G15 magnitude limit is $r_{\rm pet} < 19.4$, whereas G12 is $r_{\rm pet} < 19.8$. The redshift cone diagram of GAMA I is shown in Fig 3.3. Technical papers which describe the survey, include Baldry et al. (2010) for the target selection, Robotham et al. (2010) for the tiling algorithm and Driver et al. (2011) for the first data release. GAMA targets, have been selected according to SDSS DR6 photometry (Adelman-McCarthy et al., 2008). Throughout we use SDSS Petrosian r-band magnitudes to calculate absolute magnitudes and model magnitudes to calculate colours.

Star-galaxy separation for GAMA builds on the SDSS one, with additional infrared

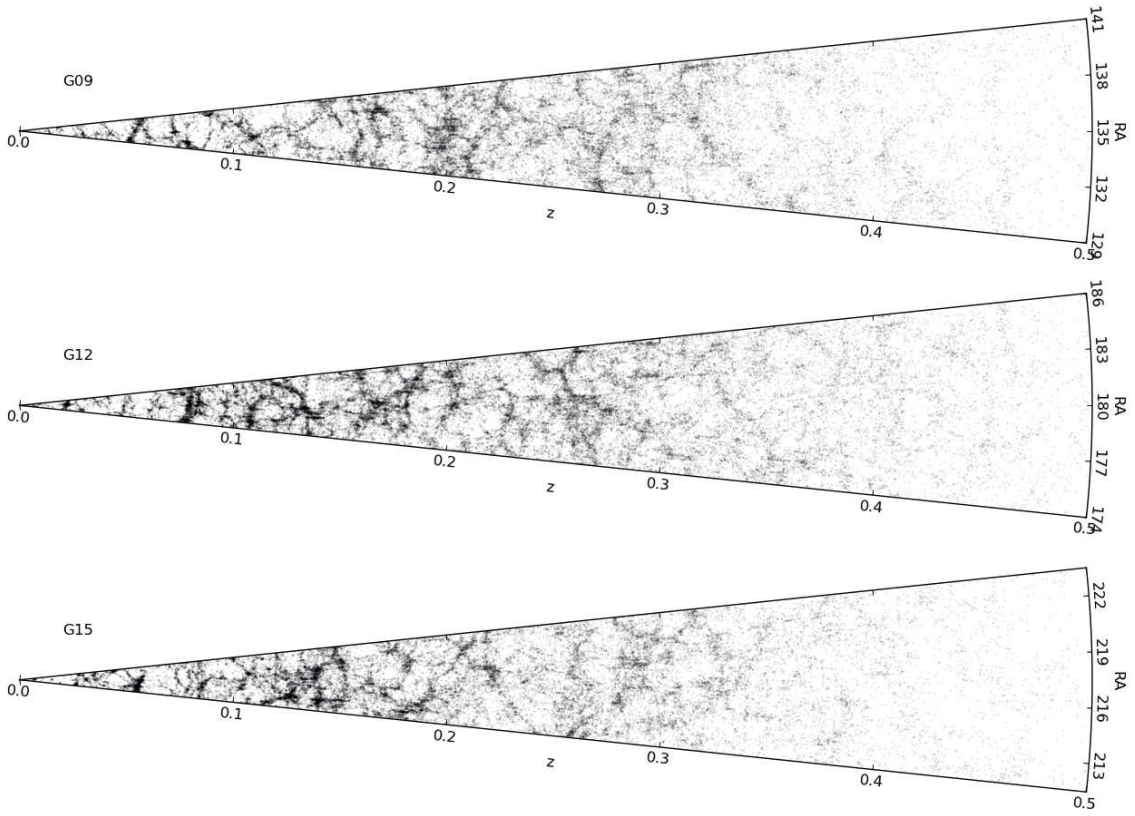

Figure 3.3: Redshift cone diagram of GAMA Phase I galaxies, split in the three GAMA fields. The diagram shows the rich structure up to $z = 0.4$, which we aim to describe using galaxy clustering statistics. (Figure credit: Jon Loveday.)

information from the UKIRT Infrared Deep Sky Survey (UKIDSS, Dye et al., 2006). An object in the GAMA catalogue is classified as a galaxy if it satisfies one of the following three criteria (Baldry et al., 2010):

$$\begin{aligned} &m_{\mathrm{PSF},r} - m_{\mathrm{model},r} > 0.25 \quad \mathrm{OR} \\ &m_{\mathrm{PSF},r} - m_{\mathrm{model},r} > 0.05 \quad \mathrm{AND} \quad J_{\mathrm{AB}} - K_{\mathrm{AB}} - f_{\mathrm{locus}}(g-i) > 0.2 \quad \mathrm{OR} \\ &m_{\mathrm{PSF},r} - m_{\mathrm{model},r} > f_{\mathrm{sg}}(m_{\mathrm{model},r}) \end{aligned} \tag{3.4}$$

where $f_{\mathrm{locus}}(x)$ is the stellar locus and has the following functional form

$$f_{\mathrm{locus}}(x) = \begin{cases} -0.7172 & \text{if } x < 0.3 \\ -0.89 + 0.615x - 0.13x^2 & \text{if } 0.3 < x < 2.3 \\ -0.1632 & \text{if } x > 2.3 \end{cases}$$

whereas, for objects without $J-K$ measurements, f_{sg} is given by

$$f_{\mathrm{sg}}(x) = \begin{cases} 0.25 & \text{if } x < 19.0 \\ 0.25 - \frac{1}{15}(x - 19) & \text{if } 19.0 < x < 20.5 \\ 0.15 & \text{if } x > 20.5 \end{cases}$$

As shown by Baldry et al. (2010) (see their Fig. 6) infrared colour $J-K$ greately improves the selection at low redshifts. Finally, all GAMA targets have been visually inspected as well.

For the extraction of the photometric catalogue, we cannot use UKIDSS information, because it does not cover the whole SDSS area, which we use for this study. We thus simply use the criterion $m_{\mathrm{PSF},r} - m_{\mathrm{model},r} > 0.25$ along with appropriate flags. In Chapter 5 we find that these criteria are not adequate for low redshift, intrinsically faint galaxies and we show how this issue severely limits clustering studies.

Aside from its unique combination of breath and depth, GAMA is also almost 98 percent spectroscopically complete down to $r < 19.4$ and 97 percent down to $r < 19.8$ for G12, which makes it ideal for using it as training set for empirical methods of estimating photometric redshifts. Redshifts were measured by the GAMA team and their quality is characterized by the variable nQ defined in Driver et al. (2011). As also described by Driver et al. (2011), significant effort has been taken by the team to assess objectively the quality of the redshifting process. Since GAMA builds on the legacy of other surveys the final redshift catalogue includes additional external redshifts. In line with all GAMA science papers we only use galaxies for which their redshift quality is $nQ > 2$. Angular masking for GAMA is derived from the SDSS imaging mask along with detailed redshift completeness as function of position and is described in (Driver et al., 2011)

A number of technicalities however, do differ between this work and other GAMA papers. The most important are the k-correction (Hogg et al., 2002) and evolutionary correction (Lin et al., 1999). In this study we use k-corrections for individual objects using `KCORRECT` version 4.1.4 (Blanton and Roweis, 2007) and evolutionary corrections using the passive evolution parameters of Loveday et al. (2012), whereas e.g. Robotham et al. (2011) used global $k + e$-corrections. r-band Petrosian magnitudes for every object are $k + e$-corrected using the correction obtained from model magnitudes.

3.3 GAMA mock catalogues

GAMA benefits state of the art mock galaxy catalogues, specifically designed to match the global clustering and luminosity properties of the survey. We use in total 9 mock galaxy catalogue and we address below the resulting constraints in our analysis imposed by their limited number. The same set of mock catalogues was used by Robotham et al. (2011) and a detailed descriptions can be found in that work and in Merson et al. (in prep.). Mock catalogues were constructed using lightcone outputs from the Millenium simulation (Springel et al., 2005), whereof dark matter haloes were populated using the semi-analytic galaxy formation model GALFORM (Bower et al., 2006). GAMA mocks match the survey selection and luminosity function in the r-band as given by the measurements of Loveday et al. (2012).

Complicated effects like luminosity dependent clustering and peculiar velocities are not accounted as realistically as desired in this first set of mocks (Robotham et al., 2011). Our tests for redshift space clustering showed that discrepancies between GAMA data and the mocks are limited and arise only at small scales. Nevertheless, we only make use of GAMA mocks to assess statistical and systematic errors in our analysis of the anisotropic clustering signal. Thus, we are interested in the relative clustering errors, instead of comparing absolute deviations on the two point correlation function. A further limitation is their statistical independence as all mocks are derived from one simulation output. As the different mocks have separations of $\sim 500\ h^{-1}$Mpc we do not expect possible long wavelength repetitions to have an effect on our correlation function uncertainty estimates which are limited to scales $< 20\ h^{-1}$Mpc.

3.4 Overview of some of the first GAMA scientific results

The GAMA survey has been acquiring data since February 2008 and it's first data release paper was Driver et al. (2011). In the meantime there have been many interesting scientific results produced by the GAMA team. Amongst them are the construction of a galaxy group catalogue (Robotham et al., 2011), luminosity function measurements (Loveday et al., 2012) and an on-going project on local group analogues[4].

The first step towards the measurement of the HMF has been made with the construction of GAMA galaxy group catalogue (G^3Cv1, Robotham et al., 2011). In contrast with the construction of a galaxy group from simulations, observations have to take into account redshift space distortions (a subject discussed in detail in Section 6.2). In order to overcome these complications one has to rely on mock catalogues, which are used to calibrate free parameters of the group finding algorithm. The main result of Robotham et al. (2011) is $14,388$ galaxy groups, along with estimates of their dynamical mass. The similarities of the present work and the galaxy group catalogue of Robotham et al. (2011) include the use of the same GAMA data release (although we work solely with $r_{\rm petro} < 19.4$), as well as same mock catalogues for assessing uncertainties and test the performance of the estimators used.

The breath and depth of the GAMA survey allows one to probe efficiently the faint end of the galaxy luminosity function. Loveday et al. (2012) have studied the galaxy luminosity functions and their evolution in the SDSS *ugriz* bands. Using the parametric Schechter function they provided measurements of the evolution of galaxy number density and luminosity for red and blue galaxies. The evolutionary parameter Q for the r-band absolute magnitude from Loveday et al. (2012) will be used for our $e-$corrections in Chapter 4 and 6. We also use the magnitude dependent colour cut proposed by Loveday et al. (2012) for our colour split in Chapter 5.

One of the most serious challenges of the ΛCDM paradigm is the Milky Way's subhalo structure (Boylan-Kolchin et al., 2012). The problem consists of the disagreement of the CDM predicted number of subhaloes in a Milky Way type halo ($M_{\rm vir,MW} \sim 10^{12} M_{\odot}$) and the actual number (as well as their respective luminosities) of the Milky Way satellite galaxies. Since this problem is at the moment linked with the Milky Way it would be informative to put it in a more cosmological perspective and check whether the Milky Way is a representative group with respect to other galaxy groups. G^3Cv1 is ideal for this exercise as it contains galaxy groups with similar mass to the Milky Way out to $z \sim 0.1$,

[4] A complete list of GAMA publications can be found at `http://www.gama-survey.org/pubs/`

i.e. a cosmologically representative volume. In Robotham et al. (2012) it is shown that at least one more galaxy group with similar properties exists in G^3Cv1, whereas ongoing work (Robotham et al., 2012b, In prep.) is addressing in considerable detail the similarities of the physical properties of the Milky Way group with the global ones of G^3Cv1.

The main focus of this thesis is studying large-scale structure in local Universe using the GAMA. Due to its relatively small volume GAMA cannot be used for placing strong constraints on cosmological parameters. Yet GAMA is $\sim$ 98% complete and therefore it can be used for testing systematic errors in the datasets as well as in the theoretical modelling. Throughout the rest of the thesis we will turn to these issues repeatedly.

Chapter 4

Constructing a photometric redshift catalogue

Parts of this Chapter have been previously published in Christodoulou et al. (2012)

4.1 Prelude

The cosmological redshift is a fundamental quantity in extragalactic astronomy. It is necessary for robustly estimating distances as well as absolute magnitudes. It is however very costly to obtain in abundance for statistical studies and it becomes inefficient for very faint galaxies, the majority of galaxies in the universe, for which too much exposure time is needed in order to get the spectra.

It is however possible to obtain information about the redshift of a galaxy without the use of spectroscopy. This is the concept of a photometric redshift which consists of using easy to obtain observables like broadband photometric information and imaging to get an estimate of the true (spectroscopic) redshift. One can then use the resulting photo-z in clustering studies. The motivation for doing this is twofold. First, we can get large numbers of intrinsically faint galaxies, for which spectroscopy is difficult to obtain en masse. Thus we are able to study the clustering properties of galaxy populations, which are numerous in the universe but rare in flux limited surveys. Second, by increasing the number of galaxies in our sample the statistical errors are reduced and thus more complicated models can be tested[1].

[1]This is not exactly true in the case of cosmological parameter estimation where the effective volume of a survey is the quantity that characterizes the usefulness of the sample (Tegmark et al., 2006). The inevitable degradation of the redshift quality in future multi-imaging photometric redshift surveys will result in a reduction of the effective volume of the survey (Seo and Eisenstein, 2003; Blake and Bridle,

Many ongoing and future galaxy surveys (e.g. DES[2], LSST[3], PanSTARRS[4] and Pau[5]) will rely entirely on photometric redshifts for their science. Unlike SDSS and GAMA for which photo-z's are a byproduct they are designed solely as photo-z surveys. Therefore, beyond the new results presented here, the analysis of this Chapter acts as a prototype for future photo-z studies.

Unfortunately, photometric redshifts come with large errors relative to the true redshifts, which are Gaussian only to a first order approximation and therefore need to be studied carefully. In some cases the estimation is catastrophic (we define catastrophic redshift estimates as those which satisfy $|z_{\rm phot} - z_{\rm spec}| > 3\sigma$) or inherently problematic (e.g. at low redshifts where we may have $z \simeq \sigma_z$, even though the photo-z error is well constrained). Therefore there is an advent of systematics which can also be understood as the result of the significant reduction of the statistical errors. As this is an extremely important issue, bound to be very relevant in extragalactic astronomy for the foreseeable future, we study it in some detail in this Chapter, in a somewhat different manner than what has already been done in the literature so far.

Once photometric redshifts have been estimated the next task is to model the radial selection function dN/dz, which describes how the galaxies were selected according to their true redshifts. This is another major complication that we have to consider and a potential cause of systematic error. It is relatively straight-forward to get a photometric redshift estimate and use this as a proxy of the galaxy radial position or any other redshift derived quantity, but it is nontrivial to quantify the effect that the errors have on the way galaxies were selected. Almost every scientific result which uses photometric redshifts depends crucially on the recovery of the true redshift distribution. Then one has to model the photo-z derived quantities. We do so for the absolute magnitude M_r, as we are interested in studying the luminosity dependence of galaxy clustering. Lastly, as shown in Chapter 5 and in particular Section **??**, one has to be critical of the scientific results per se, as there can be unsuspected and systematic contamination in large imaging galaxy catalogues.

In the next Section we briefly introduce the techniques for obtaining photometric redshifts. In Section 4.3 we describe step by step the process of building a large photometric redshift catalogue, which we are going to use in Chapter 5 for the clustering study of galaxies as functions of our derived observables, photometric redshift and photometric

2005).

[2] `http://www.darkenergysurvey.org`

[3] `http://www.lsst.org`

[4] `http://pan-starrs.ifa.hawaii.edu/public/`

[5] `http://www.pausurvey.org/home-PAU.html`

redshift derived absolute luminosity and colour. We conclude in Section 4.5.

4.2 Overview of photometric redshift methods

4.2.1 Template methods

Template methods were the first photo-z estimators, mainly due to their direct correspondence with a galaxy's spectral energy distribution (SED). Since the direct observable is broadband photometry, the template method relies on the fit of a galaxy's apparent magnitude and SED to a set of template SED's, themselves associated with redshifts (Abdalla et al., 2008). In this scheme, according to the quality of the match with the templates at a given redshift, usually quantified with the χ^2 statistic, the galaxy is assigned a redshift. One advantage of this method is the possibility of extrapolating to galaxy population for which spectroscopic redshifts are not available, with the caveat of being model-dependent. Due to the uniformity of the data that we use in this work, we do not use template fitting in our photo-z estimations, but we rely on empirical techniques instead.

4.2.2 Empirical Techniques

Under the assumption that the redshift of the galaxy is some empirical, necessarily non-injective, function of its photometric properties (usually magnitudes or colours) we can follow some empirical technique to estimate it. The simplest method is to fit a polynomial function which gives the redshift as a function of the observable fluxes (Connolly et al., 1995; Brunner et al., 2000). One then uses that polynomial with the fitted coefficients to derive unknown redshifts. This method can be extended to Artificial Neural Networks (ANN) which use any arbitrary sigmoid functions, chosen by optimization, as shown by Collister and Lahav (2004).

Collister and Lahav (2004) developed the ANNz package to estimate photometric redshifts. It consists of an input layer, a series of hidden layers and the photo-z output. Each hidden layer has a number of nodes and these nodes are connected with each other according to some weighting scheme. The training set then is used to assign the optimal weights for the training set at hand. ANNz is trained using a set of spectroscopic redshifts with similar photometry to the galaxy sample we wish to obtain photo-z's for. Ideally then, the training set is a subset of the initial galaxy sample.

The input layer of ANNz usually consists of galaxy colours or magnitudes. The reason

for that is that they are the direct observables, with usually a well defined photometric noise component, which can be accounted for by the neural network. It is however possible to use more inputs than just magnitudes and colours in ANNz. A sensible choice is Petrosian radii which clearly distinguish between low redshift and high redshift objects because of the angular size on the sky. We find that this addition increases the accuracy of ANNz.

Another empirical method for estimating photometric redshifts that we use in this thesis is the nearest neighbor (NN) method. As its name suggests instead of fitting a function of photometric observables in order to make a redshift estimation, it requires one to calculate the number of neighbours in the multidimensional magnitude space that each object has and use that information to construct redshift probability distributions according to the position of each object in magnitude space. We test the results of this method extensively in Section 4.4.3, where we also describe it in more detail, but we note that it primarily aims for constructing redshift distributions of galaxy samples.

Lastly, it should be noted that template methods and empirical methods, are not necessarily antagonistic. It is possible to use both methods simultaneously for better photo-z estimations (Budavári, 2009) or use their different advantages to choose the best method for a given dataset (Hildebrandt et al., 2010; Roseboom et al., 2012; Christodoulou et al., 2012).

4.3 A new SDSS photometric redshift catalogue

Our goal is to obtain a photometric redshift catalogue going fainter in apparent magnitude than the Sloan Digital Sky Survey (SDSS). SDSS targets spectroscopically all objects identified as galaxies with $r_{\rm petro} < 17.7$ whereas our aim is for a catalogue for galaxies with $r_{\rm petro} < 19.4$. As a training set, which is a set of galaxies with $r_{\rm petro} < 19.4$, similar photometry and spectroscopically confirmed redshifts, we use GAMA Phase I data (Driver et al., 2011). There are several advantages of going deeper in apparent magnitude. Firstly, we increase the effective volume of our survey, since $z_{mean} = 0.1$ increases to $z_{mean} = 0.19$. Secondly, we have larger numbers of different galaxy populations and therefore we can improve the statistics. As a consequence, we can study the clustering of intrinsically faint galaxies, which are often very difficult to target in wide-field spectroscopic surveys. In this thesis (Chapter 5) we present a two point clustering analysis of these underrepresented galaxy populations.

The photometric redshift estimation is done using the ANNz package (Collister and Lahav,

2004). Then we use GAMA to assess the quality of the performance of ANNz on the various quantities that we derive, like photo-z, absolute magnitudes and redshift distributions. We use these quantities to select galaxy samples in order to study low redshift galaxy clustering in Chapter 5. Traditional photo-z tests are carried out for the whole sample. However, in the following we perform those tests for the individual galaxy samples for which we wish to study clustering, as this way we can directly assess the quality of each sample we construct (see also discussion in Section 5.6).

4.3.1 Photometric redshifts

For the clustering measurements presented in this Chapter, all distance information comes from photometric redshifts. Photo-zs are the basis for estimating redshift distributions to be used in equation 2.66 and distance moduli to calculate absolute magnitudes and colours. For this study we have a truly representative subset of SDSS galaxies down to $r_{\rm petro} < 19.4$ and we therefore use the artificial neural network package ANNz developed by Collister and Lahav (2004) to obtain photo-z estimations.

It is important that the training set and the final galaxy sample from SDSS are built using the same selection criteria. We present below the SQL query used to extract the galaxy sample from the SDSS database[6]. We select galaxies which have "clean" photometry according to the instructions given on the SDSS website[7]. The SQL query used to extract our sample from the SDSS DR7 database is the following

```
SELECT
  objid, g.ra, g.dec, flags, petror50_r,
  petror50Err_r, petror90_r, petror90Err_r,
  petroMag_r - extinction_r as petroMagCor_r,
  petroMagErr_r,
  modelMag_u - extinction_u as modelMagCor_u,
  modelMag_g - extinction_g as modelMagCor_g,
  modelMag_r - extinction_r as modelMagCor_r,
  modelMag_i - extinction_i as modelMagCor_i,
  modelMag_z - extinction_z as modelMagCor_z,
  modelMagErr_u, modelMagErr_g, modelMagErr_r,
  modelMagErr_i,
```

[6] http://casjobs.sdss.org/CasJobs/

[7] http://www.sdss.org/dr7/products/catalogs/flags.html

```
  modelMagErr_z
FROM galaxy g
JOIN Frame f on g.fieldID = f.fieldID
WHERE
 zoom = 0 and stripe between 9 and 44
 and psfmag_r - modelmag_r > 0.25 and
 petromag_r - extinction_r < 19.4
   AND ((flags_r & 0x10000000) != 0)
   AND ((flags_r & 0x8100000c00a0) = 0)
   PSF_FLUX_INTERP, SATURATED,
   AND (((flags_r & 0x400000000000) = 0) or
   (psfmagerr_r <= 0.2))
   AND (((flags_r & 0x100000000000) = 0) or
   (flags_r & 0x1000) = 0)
```

Note that we used the GAMA criterion $m_{\mathrm{PSF},r} - m_{\mathrm{model},r} > 0.25$ (equation 3.4) in conjunction with the SDSS criterion for an object from the imaging catalogue to be classified as a galaxy (equation 3.3), although we expect the former to dominate the selection.

The ANNz input parameters are the following: übercal, extinction-corrected model magnitudes in *ugriz* bands, the radii enclosing 50 per cent and 90 per cent of the Petrosian r-band flux of the galaxy, and their respective uncertainties. The architecture of the network is 7:11:11:1, where 7 gives the input parameters described above, 2 the number of hidden layers, 11 is number of nodes for each layler and 1 is the photo-z output. We use a committee of 5 networks to predict the photo-z's and their uncertainties (see Section 4.3.3).

4.3.2 Colour cuts

Before we build our final sample from ANNz, we remove galaxies with outlier $u-g$, $g-r$, $r-i$, $i-z$ colours both in the SDSS imaging sample and in the training set, because photometric redshift estimates are based primarily on these colours. The complete colour and magnitude cuts are given in Table 4.1. Less than 1 per cent of the galaxies are affected by the colour cuts. These colour cuts in principle could affect the mask that we use for correlation function calculations. To estimate the extent of this effect we study the distribution on the sky of the colour outliers as well as their angular correlation function. This exercise reveals that colour outliers have a spurious correlation an order of magnitude

Table 4.1: Colour and apparent magnitude cuts for the optimization of ANNz. All colours use SDSS model magnitudes.

$12.0 < r_{\rm petro} < 19.4$
$-2 < u - g < 7$
$-2 < g - r < 5$
$-2 < r - i < 5$
$-2 < i - z < 5$

larger on all angular scales than the correlation function of our final sample. However, since the number of these objects is almost three orders of magnitude less than the total, they would have a negligible effect on $w(\theta)$ measurements if included.

4.3.3 Photometric redshift errors

Before we proceed with the photo-z derived quantities that we use in this study, we investigate the possible biases and errors that ANNz introduces, using the known redshifts from GAMA. Following standard practice we split our data into three distinct sets: the training set, the validation set and the test set. Half of the objects constitute the test set and the other two quarters the training and validation set. This investigation is insensitive to the exact numbers in these three sets. The training and validation sets are used for training the network, whereas the test set is treated as unknown. Once ANNz predicts photo-zs for the test set, we are in position to quantify the redshift error

$$\delta z \equiv z_{\rm spec} - z_{\rm phot}. \tag{4.1}$$

δz is the primary quantity of interest as far as true redshift errors are concerned. It can depend on apparent magnitude, colours, the output $z_{\rm phot}$, the intrinsic scatter $z_{\rm err}$ of ANNz committees, as well as the position of an object on the sky. We investigate some of these potential sources of error below. The dispersion σ_z, of δz is given by the equation

$$\sigma_z^2 = \left\langle (\delta z)^2 \right\rangle - \langle (\delta z) \rangle^2 \tag{4.2}$$

and is found to be $\sigma_z = 0.039$.

The fidelity of the overall photometric redshift estimation is shown in Fig. 4.1, where we plot the true spectroscopic redshifts for GAMA objects against the ANNz photo-z estimation. In Fig. 4.1 we also plot the mean photo-z and its standard deviation in photo-

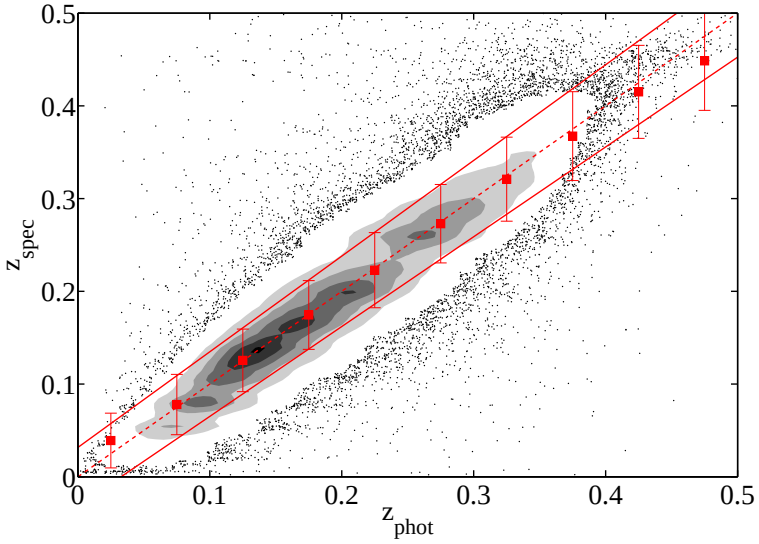

Figure 4.1: Density-scatter plot of the true spectroscopic redshift against the predicted photo-zs from this work. The colour coding is such that the densest area (black contour) is 5 times denser than the white contour. The red squares and error bars represent the mean photo-z values and their respective standard deviation in photo-z bins of width $\Delta z_{\rm phot} = 0.05$. The dashed line shows the diagonal $z_{\rm spec} = z_{\rm phot}$ and the solid lines show the quantity $z_{\rm phot} \pm \sigma_0(1 + z_{\rm phot})$, where σ_0 is defined in equation 4.3.

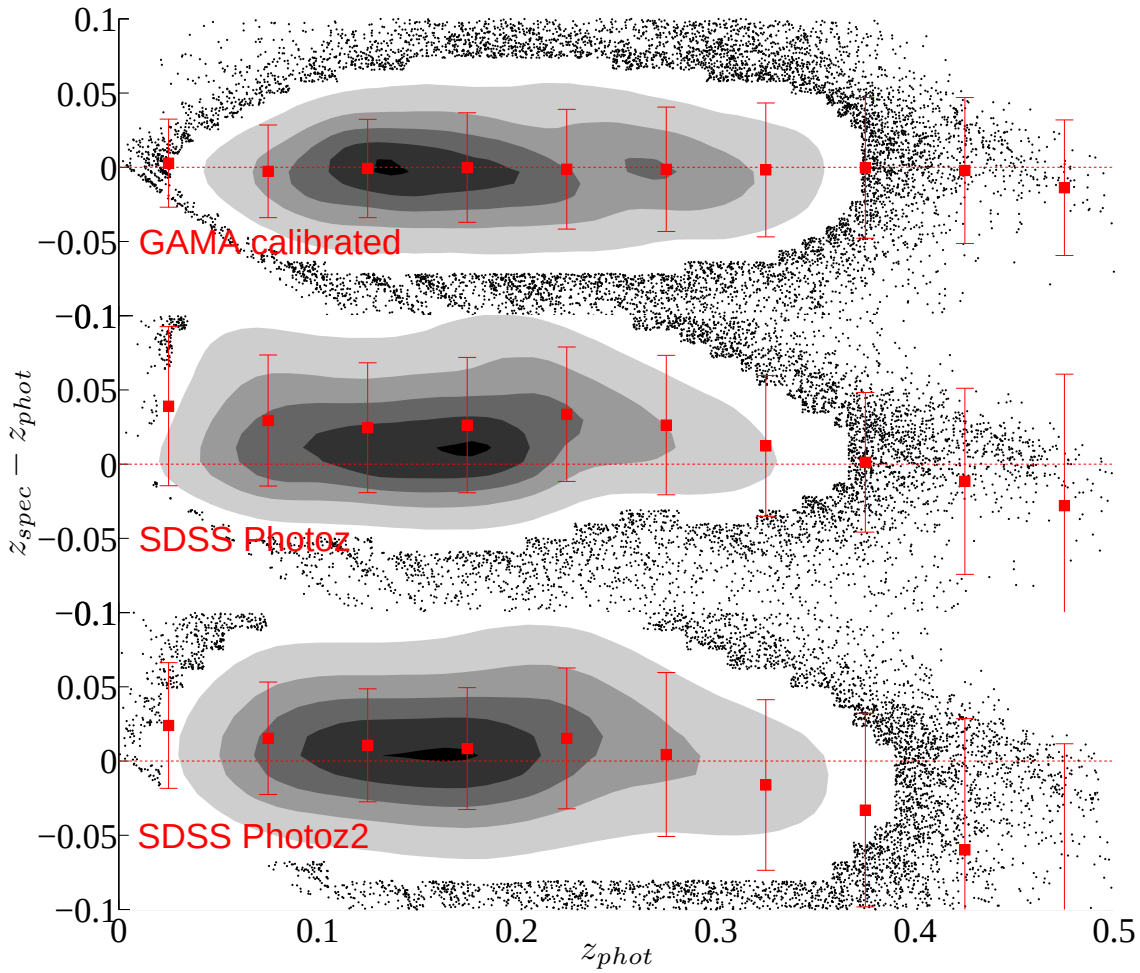

Figure 4.2: Density-scatter plot of the redshift error against the predicted photo-zs from this work (top panel) and SDSS (middle and bottom panel). The colour coding is such that the densest area (black contour) is 5 times denser than the white contour. The red squares and error bars represent the mean redshift error values and their respective standard deviation in photo-z bins of width $\Delta z_{\rm phot} = 0.05$. The improvement of the photometric redshift estimations due to the implementation of the representative GAMA data set is evident.

z bins of $z_{\rm phot} = 0.05$ and along with the redshifted dispersion (equation 4.3). Equation 4.3) is a good description of the photo-z errors.

In Fig. 4.2 we compare our photo-z estimates with the publicly available photo-z's from the SDSS website. For this comparison we plot the redshift error as a function of photo-z. We then calculate the mean and the standard deviation of δz for photo-z bins of width $\Delta z_{\rm phot} = 0.05$. We work in fixed photo-z bins, because all our derived quantities are based on the photo-z estimate and this way it's straightforward to see any biases that the photo-z results have with respect to the true redshifts. We note that if we work in fixed spectroscopic redshift bins, this analysis is unclear, since we cannot use the $\sigma_z(z_{\rm spec})$ information for our imaging data. Our photo-z estimates based on the GAMA training set outperformed the SDSS results: For the redshift range $0.01 < z_{\rm phot} < 0.4$ ANNz gives almost unbiased redshift estimates (relative to the standard deviation). We note however that some SDSS photo-z estimations are designed to perform much better in recovering the total redshift probability distribution function of every galaxy. Since it still not clear how to directly relate a redshift pdf with absolute magnitude and colour for a given galaxy, our approach for the study of luminosity and colour dependent clustering is easier to interpret.

As expected, the error increases with redshift in the range $0.05 < z_{\rm phot} < 0.5$, it is acceptable however for $z_{\rm phot} < 0.4$. In Section 4.4.2 we quantify the photo-z error and possible contamination between redshift bins, in a different way by cross-correlating photo-z bins which are more than $2\sigma_{rms}$ apart. We find that the residual cross-correlation of the different photo-z bins is negligible compared to the respective auto-correlations.

The distribution of the redshift error is in general non-Gaussian and asymmetric (Fig. 4.3). It is also well-known that the redshift error is more significant at low-z and thus a photo-z analysis is more tolerant to redshift errors at high-z. For that reason it is common practice to scale the redshift error with the quantity [8] $1/(1 + z_{\rm phot})$. This is shown in Fig. 4.3 along with a reference Gaussian with width $\sigma = 0.032$. As seen in Fig. 4.3 even after the inclusion of the $1/(1 + z_{\rm phot})$ factor, the error distribution is not strictly a Gaussian, with larger deviations at the tails ($|\delta z| > 0.08$). Besides the width of the distribution of δz, the mode is presents a small bias as it's not exactly zero as we would have expected. Yet, the $1/(1 + z_{\rm phot})$ factor shifts the mode closer to zero. Taking into account the redshift stretch, σ_0 can be defined as

$$\sigma_0^2 = \left\langle \left(\frac{\delta z}{1 + z_{\rm phot}} \right)^2 \right\rangle - \left\langle \left(\frac{\delta z}{1 + z_{\rm phot}} \right) \right\rangle^2 . \tag{4.3}$$

[8] Note that other studies use the scaling $1/(1 + z_{\rm spec})$ (Collister et al., 2007) and therefore confine the applicability of the redshift errors study to the training set only.

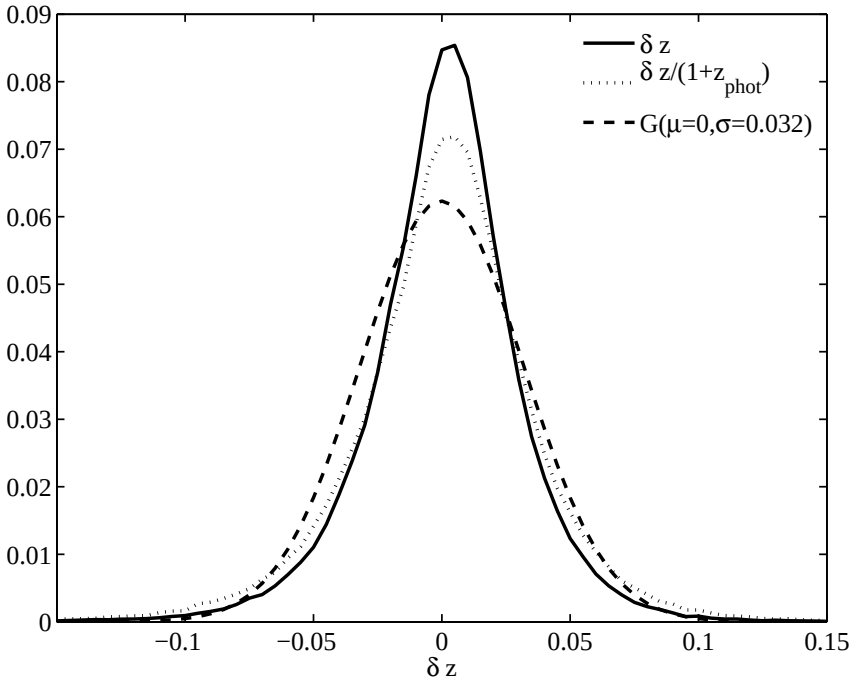

Figure 4.3: The distribution of photometric redshift error δz (solid line, equation 4.1), the scaled $\delta z/(1+z_{\rm phot})$ (dotted line) and a reference Gaussian (dashed line). The distribution of the true redshift errors is not Gaussian, especially on its tails. However, the inclusion of the $1+z_{\rm phot}$ factor renders the redshift distribution "more Gaussian".

Table 4.2: The change in the total number of galaxies as a result of various cuts applied. For the masking see Section 5.3.

Cut description	Number of galaxies remaining
None	$4,914,434$
Colour cuts (Table 4.1)	$4,890,965$
Masking	$4,511,011$
$z_{\rm err}^{\rm (ANNz)} < 0.05$ &	$4,289,223$
$0.002 < z_{\rm phot} < 0.4$	

The redshifted dispersion is $\sigma_0 = 0.032$. ANNz gives $\sigma_{rms} = 0.039$ which in the case of known redshifts is the rms deviation between the network results and the data. Finally, the standard deviation for the redshift range $0 < z_{\rm phot} < 0.4$, which is the redshift range with which we choose to work, is $\sigma_z = 0.035$.

We do not include in our analysis galaxies with $z_{\rm phot} < 0.002$ and $z_{\rm phot} > 0.4$. We also apply a cut on the output parameter $z_{\rm err}$ of ANNz at $z_{\rm err} < 0.05$. These cuts eliminate ~ 4 per cent of the galaxies. Cross-checks show that the correlation function measurements do not change if we use a less strict cut. The final number of galaxies after this cut is 4,289,223. We summarize the changes in the number of galaxies in our sample in Table 4.2. We use Petrosian magnitudes to divide galaxies by luminosity and model magnitudes to calculate galaxy colours.

In the absence of spectroscopic redshifts, ANNz provides a photo-z error estimation $z_{\rm err}$, based on the photometric uncertainty (Collister and Lahav, 2004). Naturally it is expected to correlate with the true redshift error (δz, equation 4.1) as bad photometry obscures photo-z estimation. However, our tests (shown in Fig. 4.4) show that this is an underestimation of the true error. Moreover, in Fig. 4.4 it is evident that the distribution of ANNz's output parameter $z_{\rm err}$ does not follow the distribution of the true redshift error. Therefore, for our Monte-Carlo resamplings, described in Section 4.3.5 we only use the true error δz, obtained from GAMA.

The photo-z work presented here is similar, but not identical, to that of Parkinson (2012). The latter is appropriate for even fainter SDSS magnitudes as it uses, in its training and validation, all GAMA galaxies with $r_{\rm petro} < 19.8$ and fainter zCOSMOS galaxies (Lilly et al., 2007) matched to SDSS DR7 imaging. Minor differences in the two photo-z pipelines, such as the inclusion of different light profile measurements, do

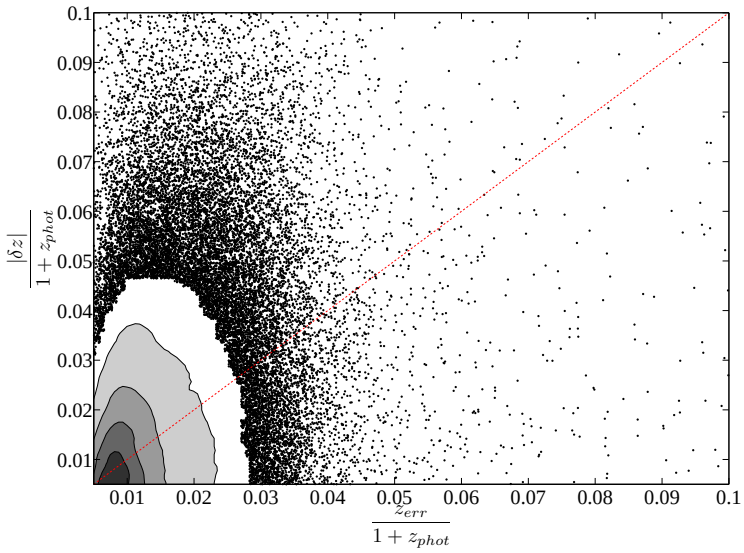

Figure 4.4: Density-scatter plot of the correlation between true redshift errors (equation 4.1) and the output parameter $z_{\rm err}$ from ANNz. Red dashed line shows the desired diagonal correlation. The correlation of the data points is therefore weak and biased towards low $z_{\rm err}$. If ANNz uses only photometric noise to estimate photo-z errors then it generally underestimates them.

not significantly affect the estimated photo-z, which present a similar scatter around the underlying spectroscopic distribution. Our photo-z agree with those of Parkinson (2012) within the estimated errors.

4.3.4 Colour and magnitude selections

Galaxy magnitudes are $k+e$-corrected to $z_{\rm phot} = 0.1$, using `KCORRECT` version 4.1.4 Blanton and Roweis (2007) and the passive evolution parameter $Q = 1.62$ of Blanton et al. (2003). In this simple model, the evolution-corrected absolute magnitude is given by $M_{\rm corr} = M - Q(z - z_0)$, where $z_0 = 0.1$ is the reference redshift. We note that Loveday et al. (2012) using GAMA found $Q = 0.7$, which would change evolution-corrected magnitudes by ≈ 0.3 mag at $z = 0.4$. Approximately equal deviations in absolute magnitude will be induced in our high-z blue galaxy samples, if we use a colour-dependent Q (e.g. Loveday et al., 2012). Assuming a global value for Q however allows for a more direct comparison with the SDSS-based clustering studies of Zehavi et al. (2005, 2011). Galaxy colours, derived from SDSS model magnitudes, are referred to as ${}^{0.1}(g-r)$, while absolute magnitude are derived using the r-band Petrosian magnitude (to match the GAMA redshift survey selection). Fig. 4.5 shows that the r-band absolute magnitude extends to

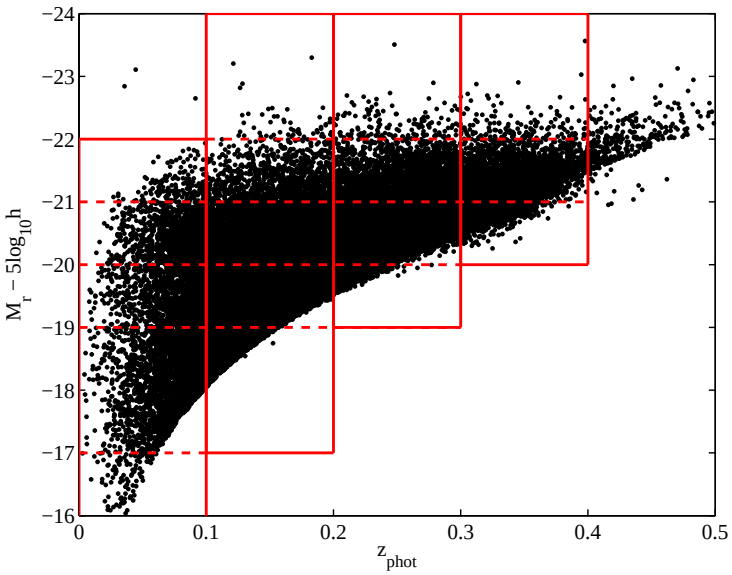

Figure 4.5: r-band absolute magnitude against photo-z in our photometric sample. Solid red lines show the boundaries of our samples in photo-z and absolute magnitude and dashed lines the further split in absolute magnitude bins. Due to the individual $k+e$-corrections the flux cut of the survey is not sharp. Only 1 percent of the galaxies are shown.

$M_r - 5\log h = -16$ mag with a few galaxies reaching as faint as $M_r - 5\log h = -14$ mag.

We split our galaxy sample in photo-z as well as luminosity bins. Our samples are shown in Fig 4.5. Initially we define four photo-z bins in the redshift range $0 < z_{\rm phot} < 0.4$ and then we further split each photo-z defined sample in six absolute magnitude bins in the range $-24 < M_r - 5\log h < -14$. Thus, the photo-z catalogue that we constructed, offers the opportunity for a clustering analysis over the luminosity range of $0.03L^* \lesssim L \lesssim 8L^*$, spanning almost three orders of magnitude in L/L^*.

Fig. 4.6 shows colour-magnitude diagrams for our sample split in photo-z bins. The colour bimodality is evident at ${}^{0.1}(g-r) \simeq 0.8$ for all photo-z bins. As there is some correlation with absolute magnitude, we use the tilted colour cut defined by Loveday et al. (2012),

$$M_r - 5\log h = 5 - 33.3 \times^{0.1} (g-r)_{\rm model}, \tag{4.4}$$

which is a slightly modified version of the colour cut used by Zehavi et al. (2011), also shown in Fig. 4.6.

In Fig. 4.7 we plot the photo-z error against photo-z for galaxies subdivided into sub-samples, where we again have used *photometric* redshifts to estimate galaxy luminosities and colours. It is encouraging that there are no obvious systematic biases of $z_{\rm phot} - z_{\rm spec}$

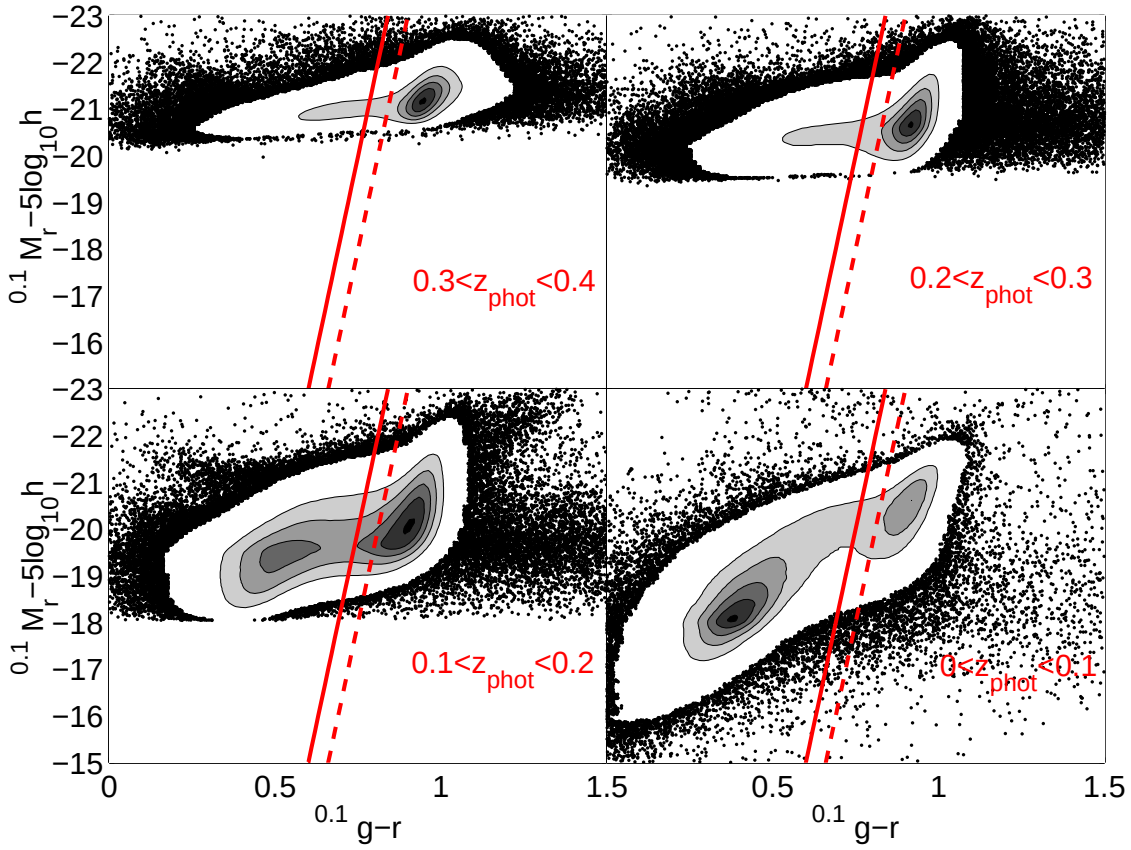

Figure 4.6: r-band absolute magnitude against $^{0.1}(g-r)$ colour (both k-corrected and passively evolved to $z = 0.1$) for galaxies split in photo-z bins. Solid red line shows the colour cut for red and blue populations, suggested by Loveday et al. (2012), used in this work and dashed red line the colour cut used by Zehavi et al. (2011).

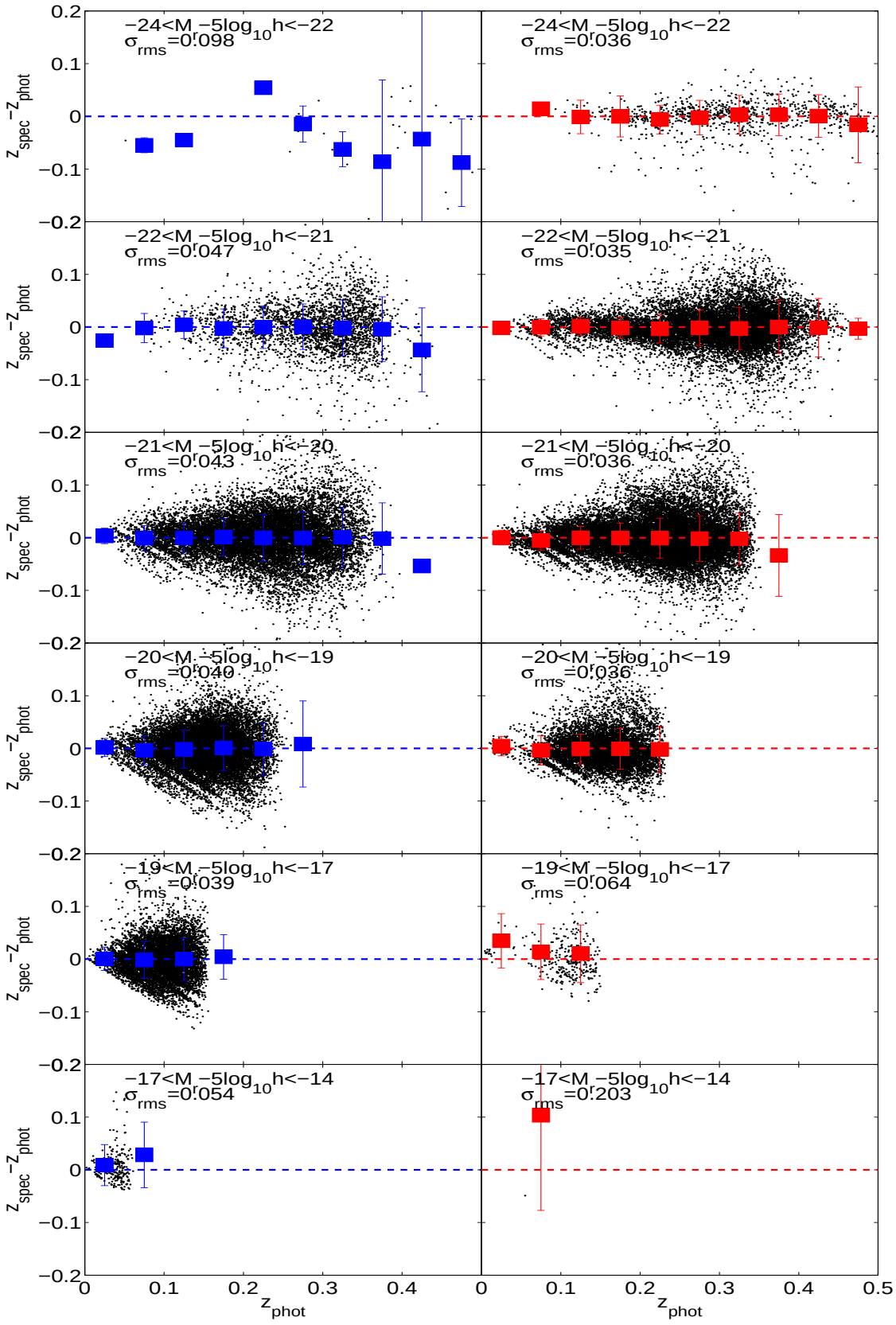

Figure 4.7: δz against photo-z for our luminosity and colour-selected GAMA subsamples. The mean redshift error and standard deviation per photo-z bin for each sample is shown by the coloured data points, while the root mean square standard deviation, σ_{rms}, is listed in each panel.

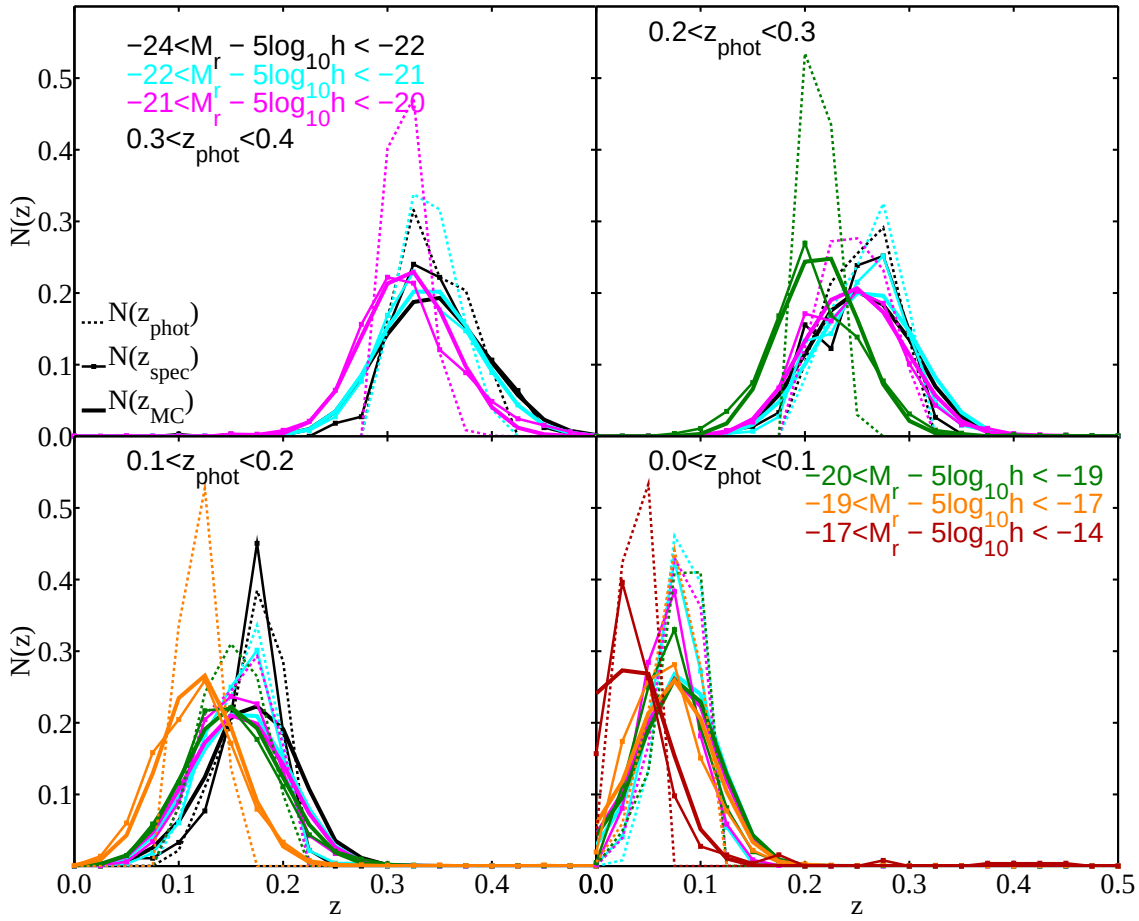

Figure 4.8: Estimates of the underlying redshift distribution as required by the Limber inversion given in equation (2.66) for the luminosity samples that are used in the clustering analysis.

for any of the subsamples, although we do note that the most luminous (faintest) bin contains very few blue (red) galaxies.

The relatively good photo-zs notwithstanding, our analysis does not eliminate completely the main systematic error of neural network derived photo-z, which is the overestimation of low redshifts and the underestimation of high redshifts (see e.g. Fig. 7 of Collister et al., 2007) . As a result a number of faint galaxies have their redshift overestimated and hence appear brighter in our sample. We note that there is a discrepancy between the fraction of faint red objects in the luminosity bin $-19 < M_r - 5\log h < -17$ between this work and Zehavi et al. (2011), which most probably is caused by this systematic shift (see Table 5.1).

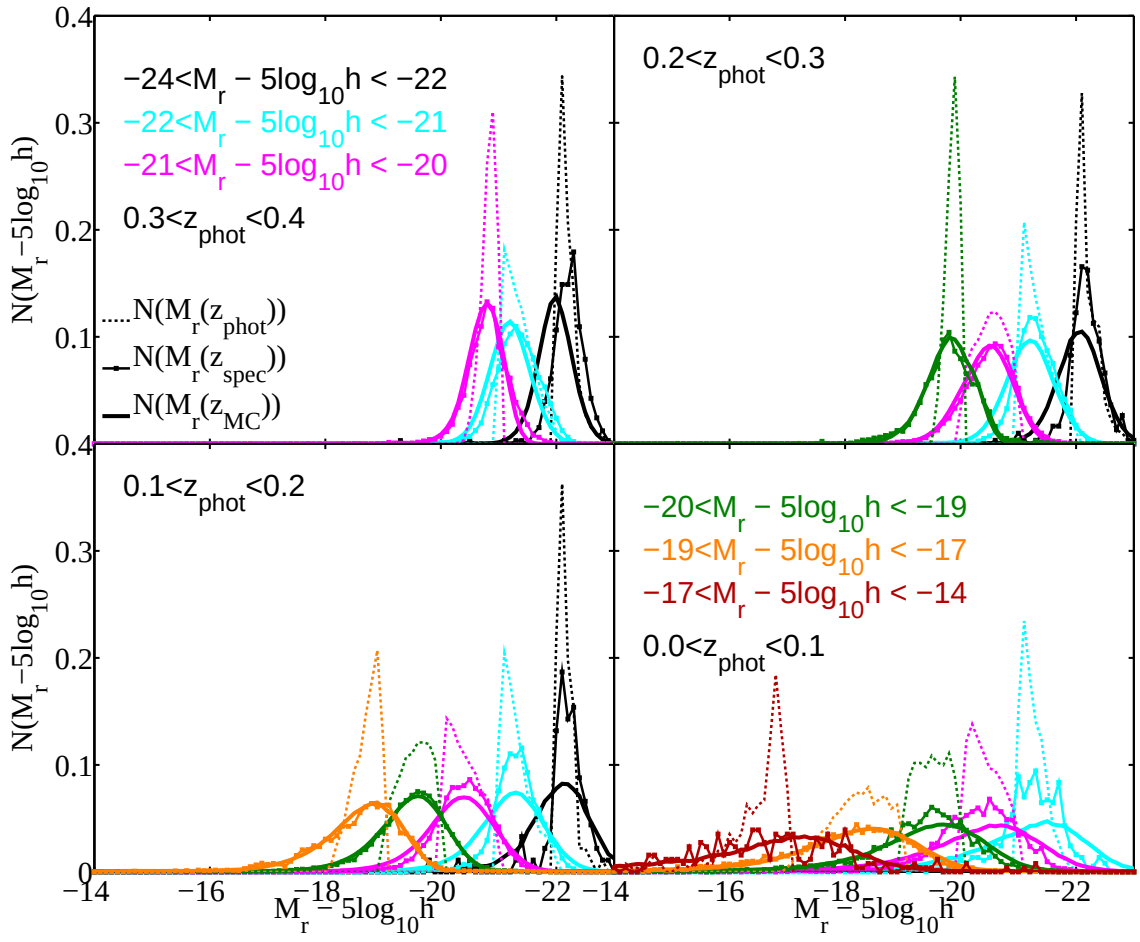

Figure 4.9: The r-band absolute magnitude distribution for GAMA galaxies with $r_{\rm petro} < 19.4$ and split into photo-z and photo-z derived absolute magnitude slices. Dashed lines show the raw photo-z derived $M_r - 5\log_{10}h$ distribution, thin lines the underlying spec-z derived $M_r - 5\log_{10}h$ distribution and the thick lines the Monte-Carlo derived $M_r - \log_{10}h$ distribution respectively. The latter one reproduces rather well the true underlying spec-z inferred $M_r - \log_{10}h$ distribution, however for a few samples there is a discrepancy between the spec-z derived and the Monte-Carlo derived distributions.

4.3.5 Redshift distribution(s)

Despite the fact that ANNz gives fairly accurate and unbiased photo-zs for calculations in broad absolute magnitude bins or photo-z bins, in order to translate the two dimensional clustering signal to the three dimensional one using equation 2.66, the underlying true dN/dz is needed. In this work we loosely follow the approach given in Parkinson (2012) (see also e.g. Driver et al., 2011). The GAMA spectroscopic sample is highly representative and it allows us to calculate the true redshift errors as a function of photo-z for all objects in GAMA with $r_{\rm petro} < 19.4$. Then, under the assumption of a Gaussian photometric error distribution in each photo-z bin, we perform a Monte-Carlo resampling of the ANNz predictions for photo-zs. This is equivalent to replacing each photo-z derived from ANNz with the quantity $z_{\rm MC}$ drawn from a Gaussian distribution, using a photo-z dependent standard deviation, $\sigma(z_{\rm phot}^{\rm (bin)}) = \delta z_{\rm phot}^{\rm (bin)}$:

$$z_{\rm MC} = G[\mu = z_{\rm phot}, \sigma = \sigma_{\rm phot}(1 + z_{\rm phot})]. \tag{4.5}$$

Note that *convolving* the imprecise photo-z with additional scatter improves the $N(z)$ redshift distribution: in other words the photo-z process *deconvolves* the $N(z)$ and makes it artificially narrow.

All our sample selections in Fig. 4.5 have been made using the photo-z derived absolute magnitude $M_r - 5{\rm log}h$. We then use the accurate spectroscopic information from GAMA to assess how well Monte-Carlo resampling compare to the underlying true dN/dz (Fig. 4.8). Since the GAMA area is much smaller than the SDSS area, we are not interested in recovering the exact spectroscopic redshift distribution, merely in matching a smoothed version thereof. Our test shows that MC resampling performs rather well in recovering the true dN/dz. This method performs even better with a larger number of objects, which indicates that we are still dominated by statistical errors and therefore there is room for improvement in future where larger data sets will be available. Nevertheless, as an incorrect redshift distribution can cause a systematic error on r_0, we test in Section 4.4.3 the sensitivity of our results to the assumed dN/dz.

Fig. 4.9 shows, for all samples split by photo-z and photo-z-derived absolute magnitude, the photo-z-derived, the true underlying and the Monte-Carlo inferred absolute magnitude distributions (as dashed, thin and thick solid lines respectively). We note that the photo-z derived absolute magnitude estimates in Fig. 4.9 are obtained from the resampled redshifts and not by resampling the absolute magnitudes per se. We then $k+e$-correct every Monte-Carlo absolute magnitude realization using the procedure described in Section 4.3.4. As expected, the true underlying distribution extends well beyond the photo-z

inferred luminosity bins, but is yet again rather well described by the Monte-Carlo inferred distribution.

It is crucial that we have a good understanding of the true underlying absolute magnitude for all our samples. For galaxy clustering studies with spectroscopic redshifts it is desirable to work with volume-limited samples. Using photometric redshifts, however, one can form only approximately volume-limited samples, since photo-z uncertainties will propagate into absolute magnitude estimates. Essentially, any tophat absolute magnitude distribution, as selected using photo-z, corresponds to a wider true absolute magnitude distribution, as shown in Fig. 4.9. This is rather similar to selecting galaxies from a photometric redshift bin and then convolving the initial tophat distribution with the photo-z error distribution in order to obtain the true $N(z)$. However, using the $w(\theta)$ statistic and an accurate dN/dz for that particular galaxy sample we can extract its respective spatial clustering signal, which would then correspond to the $z_{\rm MC}$ derived absolute magnitude. Direct comparisons with other studies can then be made, modulo the extent of the overlap between the two absolute magnitude distributions.

4.4 Tests for systematics

Clustering studies using photometric redshifts are subject to systematic errors which become more pressing as the statistical errors are significantly decreased. In this Section we study the most relevant sources of systematic errors that can affect our results. A similar study, for a brighter sample of galaxies at higher redshifts ($0.4 < z < 0.7$) was recently presented by Ross et al. (2011a).

Here we present tests that we believe are more likely to affect the results shown in this paper. We start in Section 4.4.1 with the scaling test, which is mostly tests the reliability of the whole sample for clustering studies. In Section 4.4.2 we quantify the possible systematics in the clustering signal, due to spurious cross-correlations of different photometric redshift bins. Lastly, in Section 4.4.3 we test the possible systematics in the spatial correlation function that are being introduced by the redshift distributions that we use in Limber's equation.

4.4.1 Scaling test

With a photometric sample of this size it is prudent to perform a scaling test in order to uncover any dependence of clustering on apparent magnitude. In order to do this we split our sample in apparent magnitude bins and then calculate the angular correlation function.

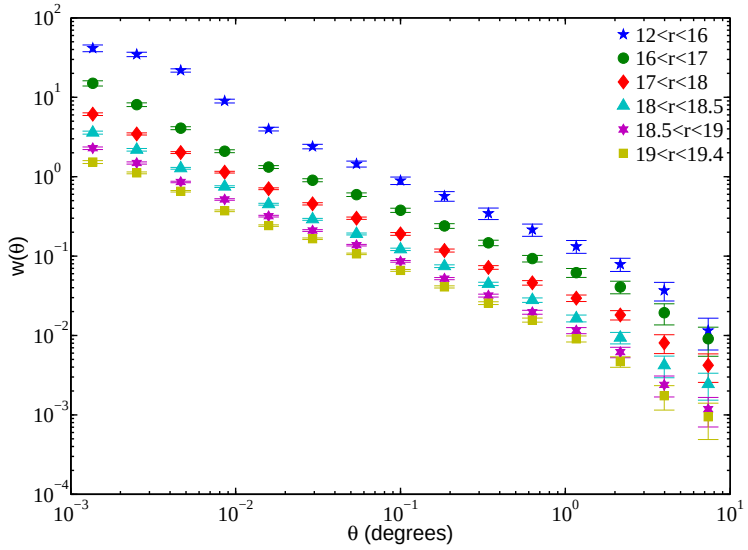

Figure 4.10: Angular correlation functions of the r-band apparent magnitude bins defined in Table 4.3.

Table 4.3: Clustering properties in apparent magnitude bins defined by r-band Petrosian magnitude. Column 1 lists magnitude range, column 2 the number of galaxies, columns 3 and 4 give the values of γ and r_0, defined in equation 2.56. Column 5 lists the quality of the power law fits. Errors were calculated using the full covariance matrix, but we don't include the $N(z)$ uncertainty.

r-bin (mags)	N_g	γ	r_0	χ^2_ν
$12.0 < r < 16.0$	79543	1.81 ± 0.03	5.01 ± 0.48	1.01
$16.0 < r < 17.0$	201805	1.72 ± 0.02	5.76 ± 0.31	3.1
$17.0 < r < 18.0$	671315	1.73 ± 0.01	5.62 ± 0.2	3.38
$18.0 < r < 18.5$	768620	1.74 ± 0.01	5.58 ± 0.17	2.28
$18.5 < r < 19.0$	1336411	1.73 ± 0.01	5.5 ± 0.12	2.55
$19.0 < r < 19.4$	1720930	1.71 ± 0.01	5.2 ± 0.12	3.48

The apparent magnitude ranges are given in Table 4.3. The angular correlation functions are shown in Fig. 4.10. For all apparent magnitude bins the slope is approximately equal, but the amplitude varies as expected, shifting from high to low values as we go fainter. Then we use equation 2.66 to calculate the correlation length for each magnitude range, using a smoothed version of the underlying $N(z_{\rm spec})$. The scales over which we fit are $0.01 < \theta < 2$ degrees ($0.02 < \theta < 1.2$ degrees for the $12 < r < 16$ sample). The correlation length for each magnitude bin is found to be equal within the error bars and in agreement with the earlier study of Budavári et al. (2003). Thus, for all well populated apparent magnitude bins we recover the fiducial power law (Peebles, 1980)

$$\xi(r) \simeq \left(\frac{r}{5\ h^{-1}{\rm Mpc}}\right)^{-1.7}. \tag{4.6}$$

4.4.2 Cross correlation of photometric redshift cells

A crucial consistency check that is necessary for the validation of our results, is the study of the induced cross correlations between redshift shells defined by photo-zs from our sample. Since we have established that $\sigma_z \approx 0.04$ we start from $z_{\rm photo} = 0$ and use five continuous slices with $\Delta z = 0.08$, in order to allow all galaxies with photo-z error of $\lesssim 2\sigma$ to be included in the correct redshift bin. We then cross-correlate slices which are more than one Δz apart.

If a Gaussian with $\sigma = 0.04$ provides good approximation of the error σ_z, then we can estimate what fraction of galaxies should lie outside the width of each photo-z slice. A galaxy which is outside its redshift slice with width $\Delta z = 0.08$ will have an error greater than 2σ. For a Gaussian distribution ~ 5 per cent of all galaxies should lie outside their redshift boundaries. Therefore their residual contribution to the cross correlation should be ~ 10 per cent of their auto-correlation[9]. In Fig. 4.11 we present three auto-correlation functions and their respective cross-correlations. The cross-correlation functions from Fig. 4.11 are not entirely consistent with zero, but on all scales the residual signal is of the expected order of magnitude. Fig. 4.11 demonstrates that ANNz does not produce spurious correlations between physically disjoint galaxies.

4.4.3 Angular clustering in GAMA area

In this Section we perform a study of the angular clustering in the GAMA area. Since we have precise knowledge of the spectroscopic redshift distributions in the GAMA area, we

[9]Assuming that the two auto-correlations are equal and the number of galaxies in each sample is equal as well. For a detailed treatment of these effects see Benjamin et al. (2010).

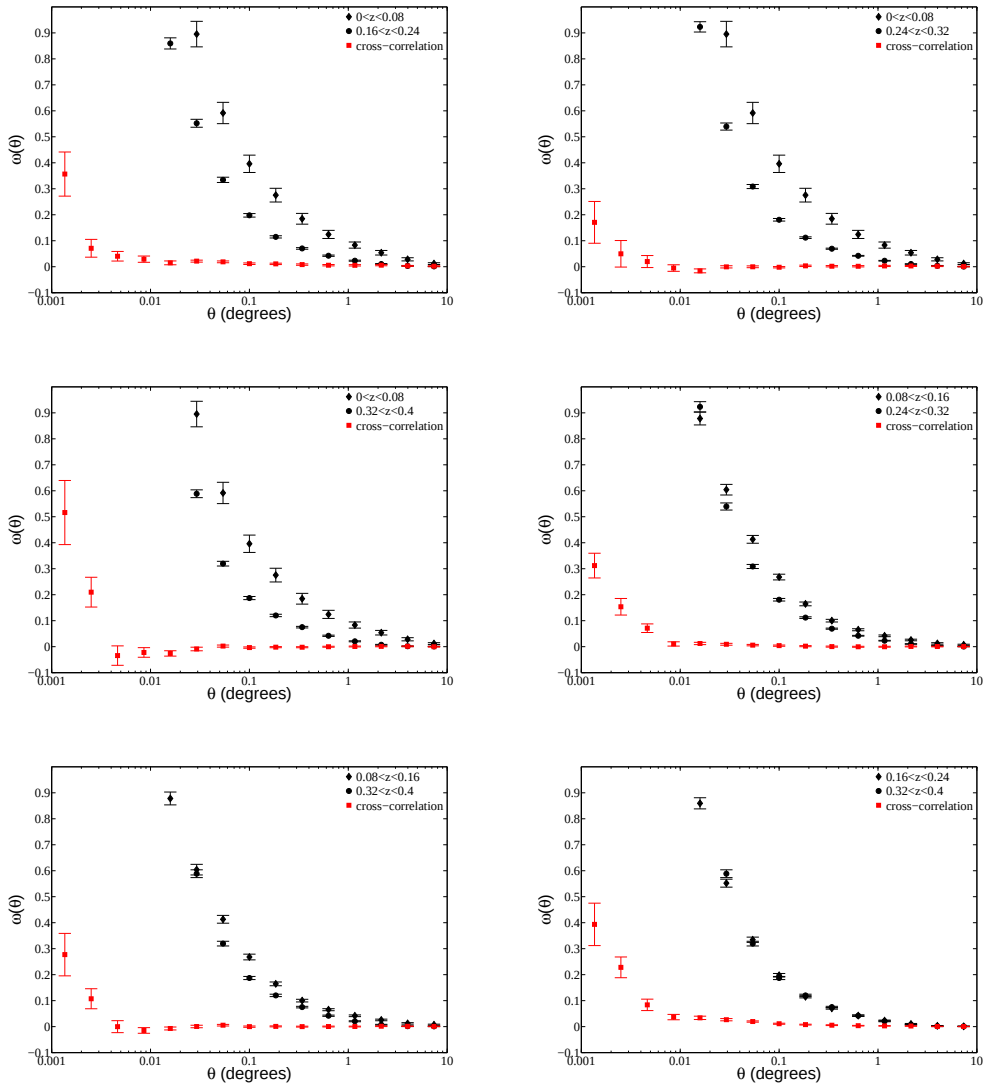

Figure 4.11: Auto-correlation (diamonds and circles) and cross-correlation (squares) functions for photo-z bins. On small angular scales, the cross-correlation signal has negligible magnitude and for angular separations ≥ 0.01 degrees is consistent with zero. The errors are calculated using JK resampling as explained in Section 5.3.

use these angular clustering measurements to test how robust are our spatial clustering results using different methods of recovering dN/dz. The methods that we test against the given GAMA spectroscopic redshift distributions, are the Monte-Carlo realizations of the photo-z distributions, assuming Gaussian errors (equation 4.5), which has been used for all the results in this paper and the weighting method of Cunha et al. (2009) (also known as nearest neighbour method).

The latter method can be summed up in three distinct steps. First, one estimates the distance in apparent magnitude space to the 200th nearest neighbour of each object in the spectroscopic set, using a euclidean metric. The exact ordinal number of the neighbouring object should not change the result significantly. For the GAMA number density N=200 is the best trade-off between smoothing out the large scale structure while at the same time preserving the locality of the photometric information. Second, one calculates the number of objects in the photometric set that are within the hypervolume defined by this distance and then one calculates the weight of each object in the spectroscopic set at the point m_i according to the equation

$$w_i = \frac{1}{N_{\rm phot,tot}} \frac{N(m_i)_{\rm phot}}{N(m_i)_{\rm spec}}, \tag{4.7}$$

where $N(m_i)_{\rm spec} = 200$. In the third step, the already known spectroscopic distribution is weighted to match the distribution of the photometric sample. The weighting is done by summing the weights w_i of each object in the spectroscopic sample for all redshift ranges:

$$N(z)_{\rm wei} = \sum_{i=1}^{N_{\rm spec,tot}} w_i N(z_1 < z_i < z_2)_{\rm spec}. \tag{4.8}$$

Cunha et al. (2009) show that their method is superior in recovering the true dN/dz to other methods using photo-zs, but they do not include the Monte-Carlo resampling in their comparisons.

The comparison of the different methods is depicted in Fig. 4.12, where all the clustering measurements are confined to the GAMA area. The errors for the angular clustering measurements are assumed to be Poisson, which is just a lower bound, and the errors on the redshift distributions are obtained from the scatter of Monte-Carlo simulations. This test is performed for the same luminosity bins as in Section 4.3.4, apart from the brightest and faintest bin which have a very small number of galaxies and hence large statistical errors on $w(\theta)$.

The (a priori required) agreement between the r_0 measurements from the different methods of recovering dN/dz is not perfect. The r_0 measurements are not significantly affected by the differences between the redshift distributions of Fig. 4.8. In conclusion,

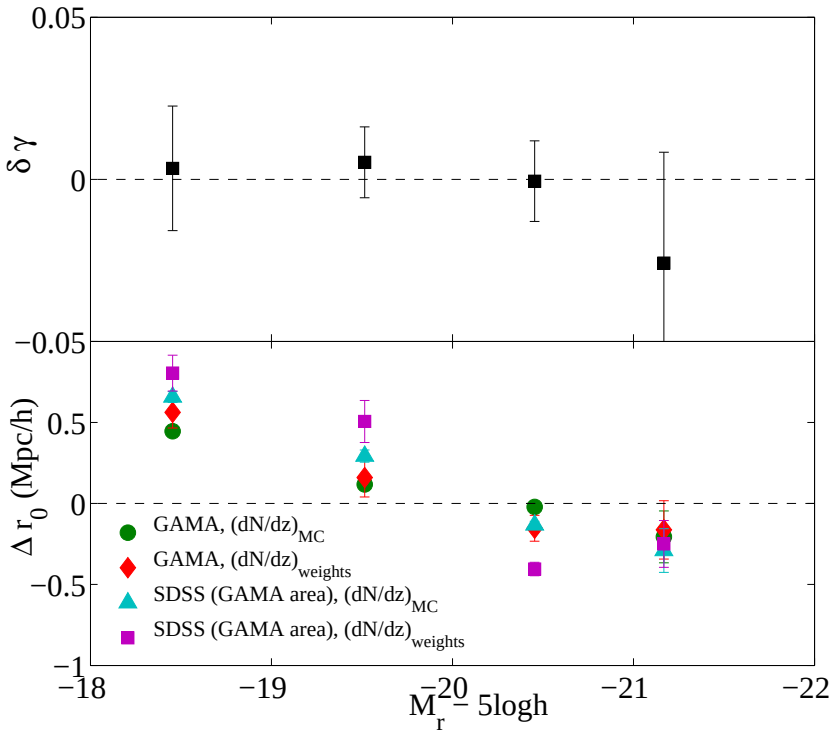

Figure 4.12: Upper panel: Slope residual of the correlation function measurements in the GAMA area, using the measurement of the GAMA sample with spectroscopic redshifts as a reference ($\Delta\gamma = \gamma(\mathrm{SDSS}) - \gamma(\mathrm{GAMA})$). Bottom panel: Comparison of the effect of the various redshift distributions (as shown in Fig. 4.8) on r_0 measurements again using the GAMA sample as a reference ($\Delta r_0 = r_0(\mathrm{i}) - r_0(\mathrm{GAMA})$). Following the discussion in Sec. 5.4 the error bars are the combined effect of the power law fit uncertainties (assumed to be Poisson) which are independent of the underlying dN/dz and the scatter in r_0 due to 100 Monte-Carlo simulations of each dN/dz (only $(dN/dz)_{\mathrm{spec}}$ is known precisely).

Fig. 4.12, for the three intermediate and well populated luminosity bins, implies that the reconstruction of the underlying redshift distribution is not introducing any systematic errors in the r_0 measurements.

This comparison does have limitations. Samples with small numbers of objects are sensitive to number variations due to the different selections of the two surveys (mainly the more conservative star-galaxy separation that we use in this paper). Moreover, it is very difficult to get realistic error bars for samples with a small number of galaxies and for which the survey's angular extent is comparable with the angular scales used for the $w(\theta)$ measurements. The difficulty in getting the exact angular clustering signal is shown in the upper panel of Fig. 4.12 which shows the residuals of the measured slopes for the GAMA and SDSS samples. In spite of these, Monte-Carlo resampling seems to recover the true r_0 slightly better than the weighting method.

4.5 Conclusions

This Chapter set the stage to proceed with the clustering analysis of SDSS imaging data. We described the methodology behind the construction of our photometric redshift catalogue and then we proceeded with a careful error analysis. The training set has been constructed with as much clean photometry as possible, with additional colour cuts, which optimized the performance of ANNz. The derived photo-z's using the GAMA data set, are shown to be less biased, than other available photo-z's from SDSS data repository.

As photometric redshifts come with significant systematic errors we also studied those in some detail. The tests that we performed were the scaling test, the cross-correlation between different photo-z bins and finally the angular clustering in the GAMA area, were true spectroscopic redshift are known. We found that, although systematic errors cannot be ignored, we are in position to account for those in our clustering study that follows.

Subsequently we selected our desired galaxy samples and taking advantage of the complete GAMA spectroscopic data down to our apparent magnitude limit, we calculated their true redshift and absolute magnitude distributions using MC resampling. Our photo-z catalogue is sufficient for extracting redshift-derived galaxy properties like luminosities and colour. Moreover, we stressed the fact that a selection in absolute magnitude derived using photo-z cannot be directly compared with a selection in spec-z derived absolute magnitude. For that reason we constructed the true underlying absolute magnitude distributions, along with the recovered radial selection functions dN/dz, which will help us with comparison with other studies. Another curious point of that method is that

an apparent convolution of the photo-z distribution with a Gaussian results in a deconvolution of the photo-z distribution and the recovery of the true spectroscopic redshift distribution. Thus, we have placed ourselves in a position with galaxy samples that have well-defined selection functions for the clustering analysis as a function of photometric redshift, luminosity and colour to which we now proceed.

Chapter 5

Results for the two-point correlation function

This Chapter have been previously published in Christodoulou et al. (2012)

5.1 Introduction

Galaxy properties are the fundamental observables for inferences and tests of galaxy formation and evolution models. The dependence of galaxy clustering on properties such as morphology, colour, luminosity or spectral type has been established over many decades. Elliptical galaxies or galaxies with red colours, which both trace an old stellar population, are known to be more clustered than spiral galaxies (e.g. Davis and Geller 1976; Dressler 1980; Postman and Geller 1984; Loveday et al. 1995; Guzzo et al. 1997; Goto et al. 2003). Recent large galaxy surveys have allowed the investigation of galaxy clustering as a function of both colour and luminosity (Norberg et al. 2002; Budavári et al. 2003; Zehavi et al. 2005; Wang et al. 2007; McCracken et al. 2008; Zehavi et al. 2011). Among the red population, a strong luminosity dependence has been observed whereby luminous galaxies are more clustered, because they reside in denser environments.

The galaxy luminosity function shows an increasing faint-end density to at least as faint as $M_r - 5\log h = -12$ mag (Blanton et al. 2005a; Loveday et al. 2012), thus intrinsically faint galaxies represent the majority of the galaxies in the universe. These galaxies with luminosity $L \ll L^*$ have low stellar mass and are mostly dwarf galaxies with ongoing star formation. However, because most wide-field spectroscopic surveys can only probe luminous galaxies over large volumes, this population is often under-represented, an effect known as Malmquist bias (Peebles, 1993).

Systematic study of faint red galaxies ($M_r - 5\log h > -19$) from large contemporary galaxy surveys has elevated their status as one of the most intriguing galaxy populations due to their large clustering length (Norberg et al., 2001, 2002; Zehavi et al., 2002; Hogg et al., 2003; Zehavi et al., 2005; Swanson et al., 2008a; Zehavi et al., 2011; Ross et al., 2011b). Nevertheless the case for the clustering of faint res galaxies is not considered closed. Recently, Ross et al. (2011b) compiled from the literature bias measurements for red galaxies over a wide range of luminosities for both spectroscopic and photometric data. They showed that the bias measurements of the faint red population are strongly affected by non-linear effects and thus on the physical scales over which they are measured. They conclude that red galaxies with $M_r > -19$ mag are similarly or less biased than red galaxies of intermediate luminosity.

Small-scale ($r < 0.1\ h^{-1}$Mpc) galaxy clustering provides additional tests of the fundamental problem of how galaxies trace dark matter. Previous studies have used SDSS data and the projected correlation function to study the clustering of galaxies at the smallest scales possible (Masjedi et al., 2006), using however extensive modeling to account for the fibre constraint in SDSS spectroscopic data. The interpretation of these results offers unique tests about how galaxies trace dark matter and the inner structure of dark matter halos (Watson et al., 2012). Motivated by these studies we present measurements of the angular correlation function down to scales of $\theta \approx 0.005$ degrees. We work solely with the angular correlation function and we pay particular attention to systematics errors and the quality of the data.

On the other hand, on sufficiently large scales ($r > 60\ h^{-1}$Mpc), it is expected that the galaxy density field evolves linearly following the evolution of the dark matter density field (Tegmark et al., 2006). However, it is less clear if this assumption holds on smaller scales, where complicated physics of galaxy formation and evolution dominate. In the absence of sufficient spectroscopic data to comprehensively study the evolution of clustering, Ross et al. (2010) used SDSS photometric redshifts to extract a volume-limited sample with $M_r < -21.2$ and $z_{\rm phot} < 0.4$. Their analysis revealed significant deviations from the passive evolution model of Tegmark and Peebles (1998). Here we perform a similar analysis, again using photometric redshifts, for the L^* population.

This Chapter is organised as follows. In Section 5.2 we present our data for this study and the method for estimating the clustering errors. The galaxy clustering results are presented in Sections 5.3, 5.4, 5.5, 5.6 and 5.7 . In Section 5.8 we present bias measurements as functions of colour, luminosity and redshift. Our findings are summarised in Section 5.9.

5.2 Sample Selection

The galaxy samples that we consider for this clustering analysis are those defined in Section 4.3.4. In Section 4.3.5 we presented extensive tests on the robustness of the recovery of the true redshift and absolute magnitude distributions and postulated that this is an important task for galaxy clustering studies which are using photometric redshifts.

To sum up, we divide our sample into four photometric redshift bins $0.3 < z_{\rm phot} < 0.4$, $0.2 < z_{\rm phot} < 0.3$, $0.1 < z_{\rm phot} < 0.2$ and $0 < z_{\rm phot} < 0.1$. We then split each photo-z bin in six absolute magnitude defined as: $M_r - 5\log h = (-24, -22]$, $(-22, -21]$, $(-21, -20]$, $(-20, -19]$, $(-19, -17]$, $(-17, -14]$. For colour, luminosity and photo-z dependent clustering we further split all our samples in colour as defined by equation 4.4. These samples are not volume limited, yet they have well defined radial selection functions and absolute magnitude distributions as shown in Section 4.3.5.

5.3 Masking, pixelisation scheme and jackknife resampling

For the clustering analysis we use the SDSS angular masks, obtained from the NYU Value-Added Catalogue[1]. These masks are provided in the `mangle` format (Hamilton and Tegmark, 2004; Swanson et al., 2008b) a standard astronomical tool for angular masking. We use the file `lss_combmask.dr72.ply` in the NYU Value-Added Catalogue (Blanton et al., 2005b). This mask, gives the exact geometry of the survey, with areas around bright star excluded. Approximately 8% of the galaxies in our sample are affected by masking (see Table 4.2). Once we have constructed the angular mask, we use `mangle` to create random points with the exact geometry of the survey for our clustering estimator (equation. 2.71). The upper panel of Fig. 5.1 shows the boundaries of the final mask for SDSS DR7 that we use for creating random catalogues. Our random catalogues consist of $\sim 10^7$ objects, approximately ten times larger than the number of galaxies in each luminosity and colour bin. Consistency checks have shown that our clustering results are not sensitive to any particular realization of the random catalogue. Note that the scaling test presented in Section 4.4.1 also checks the accuracy of the survey mask, as well as the photometric uniformity of the sample, by studying the angular clustering of our sample as a function of r-band apparent magnitude.

Throughout, we use the Landy and Szalay (1993) estimator for calculating the galaxy correlation function. Due to the size of our samples we pre-grid our data and random points

[1] `http://sdss.physics.nyu.edu/vagc/`

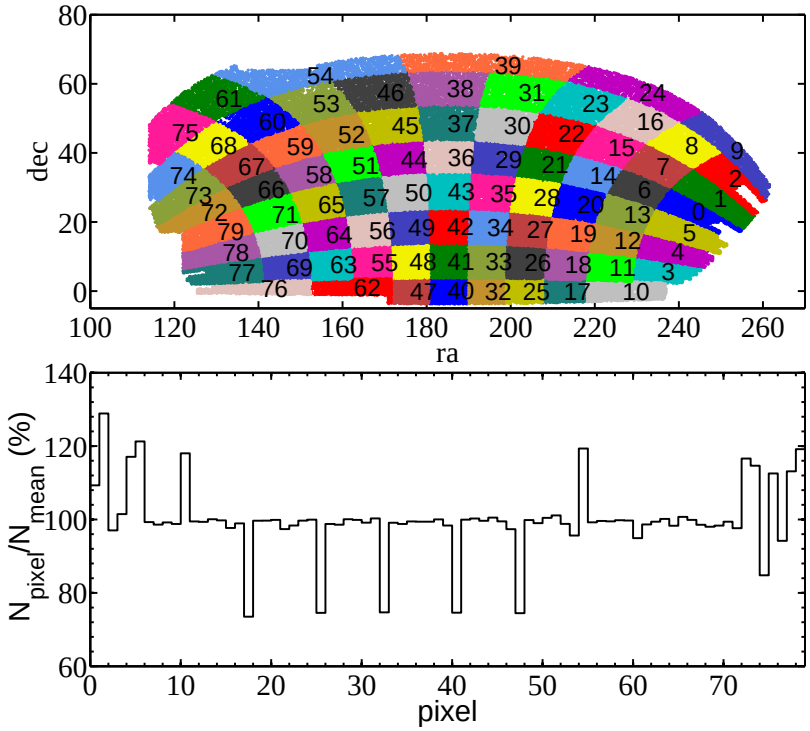

Figure 5.1: The upper panel shows the jackknife regions used for the error estimation of our correlation function measurements. After modifying the SDSSPix scheme, there are 80 jackknife regions which contain approximately equal numbers of random points. The lower panel reports the normalized area of each pixel, based on a random catalogue. The deviations from uniformity show that differences in the areas of the JK regions are limited to ± 30 per cent at most.

so that the code only considers pairs inside the same pixels and pairs in neigbouring pixels only. We take advantage of the SDSSPix[2] routines, which define pixels on the sky using the SDSS η and λ spherical coordinates (see Section 3.2.2 of Stoughton et al., 2002) and then assign a pixel number on all galaxies and randoms. For that exercise we use the "basic" resolution which divides the sky in 468 pixels of size $\sim 9.4 \times 9.4$ deg. Thus, at all times, each pixel and its 8 direct neighbouring pixels include all coherent galaxy and random pairs within angular separations up to 9.4 degrees, the largest angular separation we consider (see Section 5.4).

With a galaxy sample of this size and depth which covers almost a quarter of the sky we aim for a robust way to estimate our statistical uncertainties. We also use this pixelisation scheme to define the Jackknife (JK) regions for the error analysis. In order to minimize the variation in the number of objects in each JK region, some neighbouring pixels that contain the survey boundary are merged in order that they contain a more nearly equal number of random points. This modification of the SDSSPix pixelisation yields 80 JK regions, as shown in the upper panel of Fig. 5.1. The lower panel of Fig. 5.1 presents the relative variation in area of each region, as measured by the relative number of randoms each one contains. Hereafter, errors on $w(\theta)$ are determined from 80 JK resamplings, by calculating $w(\theta)$ omitting each region in turn. We have checked that our results are not significantly affected by using either 104 or 40 Jackknife regions. The elements of the covariance matrix, $\boldsymbol{C}$, are given by:

$$C_{ij} = \frac{N-1}{N} \sum_{k=1}^{N} (\log(w_i^k) - \log(\bar{w}_i))(\log(w_j^k) - \log(\bar{w}_j)), \tag{5.1}$$

where w_i^k is the angular correlation function of the $k^{\rm th}$ JK resampling on scale θ_i, $\bar{w}_i$ the mean angular correlation function and N the total number of JK resamplings. In practice, $\bar{w}_i$ is identical to the angular correlation function measurement from the whole survey area. The $N-1$ factor in the numerator of equation (5.1) accounts for correlations inherent in the jackknife procedure (Miller, 1974).

Jackknife is a method of calculating uncertainties on a quantity that that we measure from the data itself. In wide-field galaxy surveys, more often than not, large superstructures appear to significantly influence clustering measurements. The best known example is the SDSS Great Wall (Gott et al., 2005). The presence of such structures makes it tempting to present the results with and without the JK region that encloses them, as done in the clustering studies of Zehavi et al. (2005, 2011). Better still, Norberg et al.

[2] `http://dls.physics.ucdavis.edu/~scranton/SDSSPix/`

(2011) devise a more objective method to consistently remove outlier JK regions, from the distribution of all JK measurements that one has at hand. We follow that method in the present analysis, and find that for all samples considered, the number of JK regions that are outliers, and therefore removed, is mostly two or three and no more than five.

5.4 Luminosity and redshift dependence

We first calculate the angular correlation function $w(\theta)$ for our samples selected on absolute magnitude and photometric redshift over angular scales from 0.005 to 9.4 degrees, in 15 equally spaced bins in $\log(\theta)$[3]. In a flux-limited survey like SDSS, intrinsically bright galaxies dominate at high redshifts and intrinsically faint objects dominate at low redshifts (see Fig. 4.5). For that reason, we calculate $w(\theta)$ for the 17 well-populated samples given in Table 5.1. Errors are estimated using the jackknife technique, with the covariance matrix given by equation 5.1. Even if the validity of a given error method based on data alone is still widely debated, it is commonly accepted that the jackknife method is adequate for angular clustering studies (see e.g. Cabré et al., 2007), while for 3-D clustering measurements, Norberg et al. (2009) have shown that the jackknife method suffers from some limitations, in particular on small scales.

Our angular correlation function measurements are broad and probe both highly non-linear and quasi-linear scales. Fig. 5.2 presents galaxy angular correlation functions for six photo-z selected absolute magnitude bins. We show the angular scale (lower x-axis), used for the correlation function estimation, and the corresponding comoving scale estimated at the mean redshift of the sample (upper x-axis).

Over the range of angular scales fitted, chosen to correspond to approximately 0.1–20 h^{-1} Mpc comoving separation according to the mean redshift of each sample, the angular correlation function can be reasonably well approximated by a power law, equation 2.65. We perform power law fits, both with the full covariance matrix and with the diagonal elements only. The power law fits for our L^* sample are shown in Fig. 5.2. Dotted lines in Fig. 5.2 show the extension of the power laws beyond the scales over which they were fitted. The resulting correlation lengths, r_0, slopes, γ, and quality of the fits as given by the reduced χ^2, χ^2_ν, for all samples are listed in Table 5.1.

The luminosity dependence of galaxy clustering is present in all photo-z shells: the shape and the amplitude of the angular correlation function differ for galaxies with dif-

[3]Initially our analysis was done down to $\theta = 0.001$ degrees. However, as shown in Section 5.6, the data is not reliable enough on such small scales.

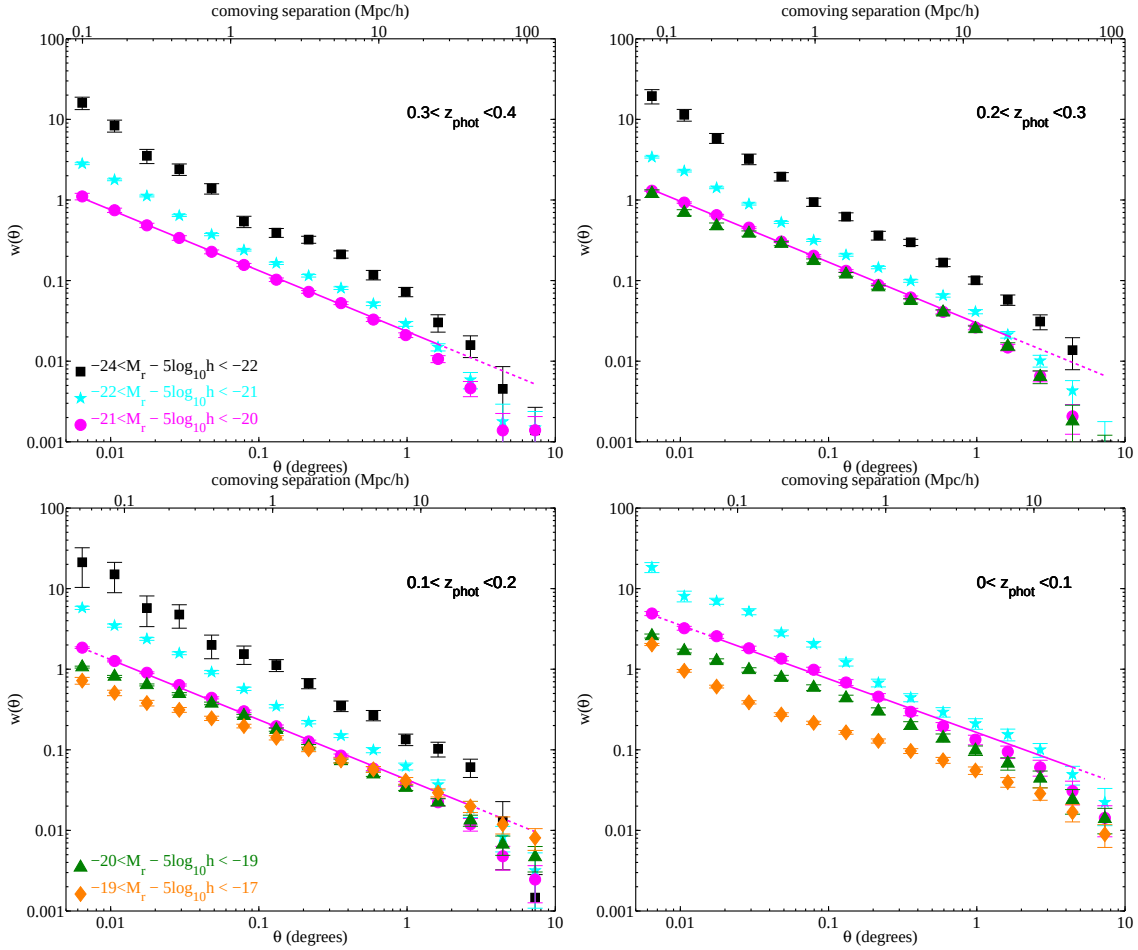

Figure 5.2: Two-point angular correlation functions $w(\theta)$ of our samples split into photo-z bins and six photo-z-inferred absolute magnitude bins, as indicated in each panel, with jackknife errors. The solid lines show power law fits estimated using the full covariance matrix for the L^* sample. Dotted lines show the extension of the power law fits on scales $< 0.1h^{-1}$Mpc and $> 20h^{-1}$Mpc.

Table 5.1: Clustering properties of luminosity-selected samples. Col. 1 lists the photo-z based absolute magnitude ranges, col. 2 the median absolute magnitude and the associated $16^{\rm th}$ and $84^{\rm th}$ percentiles from the Monte-Carlo resampling (Fig. 4.9) and col. 3 the number of galaxies in each sample. Cols. 4, 5 and 6 list respectively the slope, γ, the correlation length, r_0, and the reduced χ^2, χ^2_ν, of the power law fit as defined in Section 2.5.4. Cols. 7, 8 and 9 show the same information but for power law fits using only the diagonal elements of the covariance matrix. All power law fits are approximately over the comoving scales $0.1 < r < 20\ h^{-1}$ Mpc. Finally col. 10 presents the relative bias at $5\ h^{-1}$ Mpc measured using equation 5.2.

Sample	Magnitude$^{(\rm MC)}$	$N_{\rm gal}$	γ	r_0	χ^2_ν	$\gamma^{(d)}$	$r_0^{(d)}$	${\chi^{(d)}}^2_\nu$	b/b^*
$M_r - 5\log h$	$M_r - 5\log h$			$[h^{-1}{\rm Mpc}]$			$[h^{-1}{\rm Mpc}]$		
All colours $0.3 < z_{\rm phot} < 0.4$									
$[-24,-22)$	$-22.0^{-0.2}_{+0.2}$	13257	2.01 ± 0.15	14.08 ± 2.09	3.41	2.02 ± 0.09	13.68 ± 1.22	2.6	2.13 ± 0.30
$[-22,-21)$	$-21.2^{-0.3}_{+0.3}$	339834	1.94 ± 0.11	8.23 ± 1.54	28.08	1.91 ± 0.09	8.46 ± 1.06	13.0	1.22 ± 0.22
$[-21,-20)$	$-20.8^{-0.2}_{+0.2}$	158860	1.75 ± 0.06	6.96 ± 0.56	3.76	1.78 ± 0.05	6.80 ± 0.33	1.8	1.00 ± 0.01
All colours $0.2 < z_{\rm phot} < 0.3$									
$[-24,-22)$	$-22.0^{-0.3}_{+0.3}$	12294	2.02 ± 0.11	13.29 ± 2.01	2.37	2.01 ± 0.07	13.17 ± 1.13	1.7	2.02 ± 0.32
$[-22,-21)$	$-21.2^{-0.4}_{+0.3}$	284969	1.92 ± 0.09	7.92 ± 1.13	10.91	1.90 ± 0.06	8.12 ± 0.70	5.5	1.17 ± 0.17
$[-21,-20)$	$-20.4^{-0.3}_{+0.4}$	930539	1.75 ± 0.05	6.94 ± 0.76	7.96	1.77 ± 0.05	6.74 ± 0.36	3.3	1.00 ± 0.03
$[-20,-19)$	$-19.8^{-0.3}_{+0.3}$	122870	1.75 ± 0.08	5.84 ± 0.57	2.44	1.76 ± 0.06	5.84 ± 0.29	1.5	0.86 ± 0.10
All colours $0.1 < z_{\rm phot} < 0.2$									
$[-24,-22)$	$-22.0^{-0.4}_{+0.3}$	4311	1.96 ± 0.09	12.58 ± 1.35	0.59	1.95 ± 0.08	12.57 ± 1.13	0.4	2.10 ± 0.35
$[-22,-21)$	$-21.2^{-0.4}_{+0.5}$	106728	1.92 ± 0.05	7.31 ± 0.60	3.56	1.92 ± 0.04	7.40 ± 0.32	1.7	1.22 ± 0.18
$[-21,-20)$	$-20.3^{-0.5}_{+0.5}$	604181	1.75 ± 0.05	6.03 ± 0.77	7.16	1.78 ± 0.06	5.85 ± 0.43	3.9	1.00 ± 0.05
$[-20,-19)$	$-19.5^{-0.4}_{+0.5}$	916563	1.63 ± 0.11	6.36 ± 2.42	42.40	1.71 ± 0.10	5.81 ± 0.75	11.7	1.03 ± 0.30
$[-19,-17)$	$-18.6^{-0.4}_{+0.6}$	211336	1.55 ± 0.08	5.17 ± 0.83	4.41	1.58 ± 0.07	4.89 ± 0.34	1.6	0.87 ± 0.16
All colours $0.0 < z_{\rm phot} < 0.1$									
$[-22,-21)$	$-21.1^{-0.7}_{+0.8}$	19218	1.89 ± 0.13	8.21 ± 2.32	6.36	1.88 ± 0.07	8.09 ± 0.80	1.6	1.15 ± 0.43
$[-21,-20)$	$-20.3^{-0.7}_{+0.9}$	122787	1.68 ± 0.09	7.31 ± 1.40	9.00	1.75 ± 0.05	6.84 ± 0.50	2.1	0.99 ± 0.23
$[-20,-19)$	$-19.4^{-0.6}_{+0.8}$	155147	1.60 ± 0.08	6.23 ± 1.06	9.08	1.65 ± 0.08	6.10 ± 0.64	4.5	0.86 ± 0.20
$[-19,-17)$	$-18.1^{-0.8}_{+1.0}$	271389	1.54 ± 0.06	4.33 ± 0.58	6.20	1.58 ± 0.09	3.97 ± 0.24	2.9	0.65 ± 0.18
$[-17,-14)$	$-16.6^{-0.9}_{+1.4}$	14659	2.03 ± 0.25	4.28 ± 1.56	5.82	2.00 ± 0.28	4.41 ± 1.03	2.1	0.62 ± 0.25

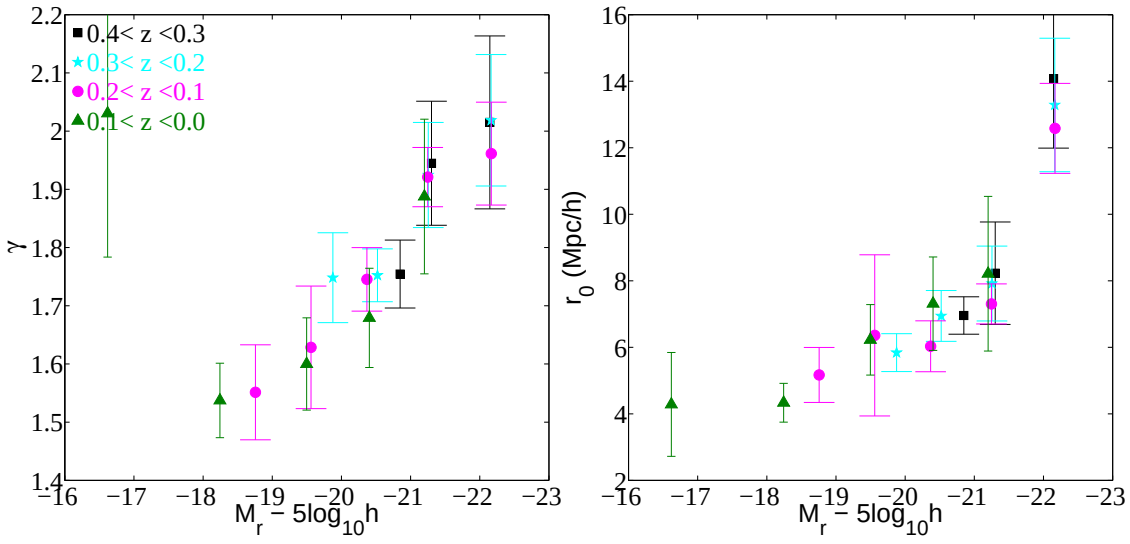

Figure 5.3: Left: Power law slope, γ, as a function of absolute magnitude and redshift. Right: Real space correlation length, r_0, as a function of absolute magnitude and redshift. Absolute magnitude ranges for which r_0 and γ measurements are valid are given in Table 5.1.

ferent luminosity. The amplitude of the angular correlation function decreases as we go from bright to faint galaxies for all photo-z bins. The slope of the correlation function also decreases with decreasing luminosity, very much in line with the change in the fraction of red and blue galaxies. As observed in Section 5.5, red (blue) galaxies dominate the brightest (faintest) luminosity bins, with red galaxies preferentially having a steeper correlation function slope than blue galaxies.

For each sample, we estimate the correlation length r_0 via equation 2.66 using the Monte-Carlo inferred redshift distribution described in Section 4.3.5. The redshift distribution dN/dz is calculated separately for each sample, as shown in Fig 4.8. In Section 4.4.3 we investigated the effects of the assumed dN/dz on the recovered correlation length r_0, and showed that the adopted dN/dz recovery method compares favourably with the true underlying dN/dz, as obtained from the smoothed $dN/dz_{\rm spec}$.

For our luminosity bins in the redshift range $0 < z < 0.1$, the correlation length is found to decrease as we go to fainter absolute magnitudes, from 8.21 ± 2.32 h^{-1}Mpc ($-22 < M_r - 5\log h < -21$) to 4.28 ± 1.56 h^{-1}Mpc ($-19 < M_r - 5\log h < -17$). This is very much in line with the recent results of Zehavi et al. (2011). Moreover, we do not observe strong evolution with redshift for samples of fixed luminosity. All r_0 and γ measurements are shown in Fig. 5.3.

Table 5.2: Clustering properties of luminosity-selected red galaxies. Columns are the same as in Table 5.1.

Sample	Magnitude$^{(\mathrm{MC})}$	N_{gal}	γ	r_0	χ^2_ν	$\gamma^{(d)}$	$r_0^{(d)}$	${\chi^{(d)}}^2_\nu$	b/b^*
$M_r - 5\log h$	$M_r - 5\log h$			$[h^{-1}\mathrm{Mpc}]$			$[h^{-1}\mathrm{Mpc}]$		
			Red	$0.3 < z_{\mathrm{phot}} < 0.4$					
$[-24,-22)$	$-22.0^{-0.2}_{+0.2}$	13095	2.02 ± 0.15	13.91 ± 2.22	3.01	2.03 ± 0.11	13.65 ± 1.86	2.4	1.78 ± 0.26
$[-22,-21)$	$-21.2^{-0.3}_{+0.3}$	287622	1.98 ± 0.10	8.40 ± 1.64	24.60	1.94 ± 0.10	8.71 ± 1.17	13.7	1.06 ± 0.20
$[-21,-20)$	$-20.7^{-0.2}_{+0.2}$	79073	1.86 ± 0.05	8.19 ± 0.54	1.33	1.88 ± 0.05	8.08 ± 0.40	1.2	1.00 ± 0.01
			Red	$0.2 < z_{\mathrm{phot}} < 0.3$					
$[-24,-22)$	$-22.0^{-0.3}_{+0.3}$	12200	2.02 ± 0.11	13.33 ± 1.95	1.89	2.01 ± 0.07	13.24 ± 1.11	1.8	1.73 ± 0.41
$[-22,-21)$	$-21.2^{-0.4}_{+0.3}$	242452	1.95 ± 0.10	8.26 ± 1.31	11.23	1.92 ± 0.06	8.41 ± 0.72	6.0	1.05 ± 0.25
$[-21,-20)$	$-20.5^{-0.3}_{+0.4}$	597678	1.81 ± 0.06	8.01 ± 1.20	17.10	1.84 ± 0.06	7.69 ± 0.52	6.5	0.98 ± 0.04
$[-20,-19)$	$-19.8^{-0.3}_{+0.3}$	44588	1.95 ± 0.09	8.53 ± 1.30	5.59	1.91 ± 0.08	8.57 ± 0.43	2.8	1.07 ± 0.21
			Red	$0.1 < z_{\mathrm{phot}} < 0.2$					
$[-24,-22)$	$-22.0^{-0.4}_{+0.3}$	4271	1.96 ± 0.08	12.61 ± 1.26	0.47	1.95 ± 0.08	12.57 ± 1.13	0.4	1.87 ± 0.48
$[-22,-21)$	$-21.2^{-0.4}_{+0.5}$	93975	1.94 ± 0.05	7.56 ± 0.71	2.52	1.93 ± 0.04	7.65 ± 0.36	1.6	1.13 ± 0.28
$[-21,-20)$	$-20.3^{-0.5}_{+0.5}$	393344	1.78 ± 0.11	7.07 ± 1.81	17.30	1.84 ± 0.08	6.68 ± 0.64	6.3	1.03 ± 0.10
$[-20,-19)$	$-19.5^{-0.4}_{+0.5}$	344815	1.71 ± 0.20	9.69 ± 5.98	82.81	1.85 ± 0.12	8.19 ± 1.26	16.9	1.33 ± 0.66
$[-19,-17)$	$-18.7^{-0.4}_{+0.5}$	12942	1.86 ± 0.18	17.86 ± 4.26	9.69	1.84 ± 0.14	17.72 ± 2.88	4.6	2.46 ± 0.83
			Red	$0.0 < z_{\mathrm{phot}} < 0.1$					
$[-22,-21)$	$-21.1^{-0.7}_{+0.9}$	18631	1.90 ± 0.14	8.20 ± 2.62	5.97	1.88 ± 0.07	8.14 ± 0.78	1.7	0.96 ± 0.47
$[-21,-20)$	$-20.4^{-0.7}_{+0.9}$	83541	1.71 ± 0.11	8.82 ± 2.34	10.98	1.79 ± 0.07	7.90 ± 0.76	3.2	0.97 ± 0.29
$[-20,-19)$	$-19.5^{-0.6}_{+0.8}$	45541	1.77 ± 0.16	10.41 ± 3.89	19.29	1.85 ± 0.14	10.39 ± 1.66	8.1	1.15 ± 0.46
$[-19,-17)$	$-18.7^{-0.5}_{+0.7}$	6690	1.88 ± 0.13	11.59 ± 2.82	2.65	1.90 ± 0.09	11.77 ± 1.32	1.0	1.43 ± 0.51

There are two main sources of error in the r_0 estimates: (a) the correlated uncertainties on the power law parameters γ and A_w which propagate through equation 2.66 to r_0; (b) statistical and systematic uncertainties in the modelling of the underlying redshift distribution. The $w(\theta)$ uncertainties and the induced error on r_0 and γ are obtained using the standard deviation from the distribution of JK resampling estimates (Section 5.3). As in the case of the covariance matrix, these uncertainties are multiplied by a factor of $N-1$ (Norberg et al., 2009). The dN/dz uncertainties are investigated in great detail in Section 4.4.3, where we show that the Monte-Carlo inferred dN/dz performs best, while still returning a residual systematic uncertainty of $\pm 0.2\ h^{-1}$Mpc on r_0 that depends on the sample considered. We find that both sources of uncertainty have a comparable contribution to the errors. In Table 5.1 we quote the total error on the correlation length after adding the two (independent) errors in quadrature.

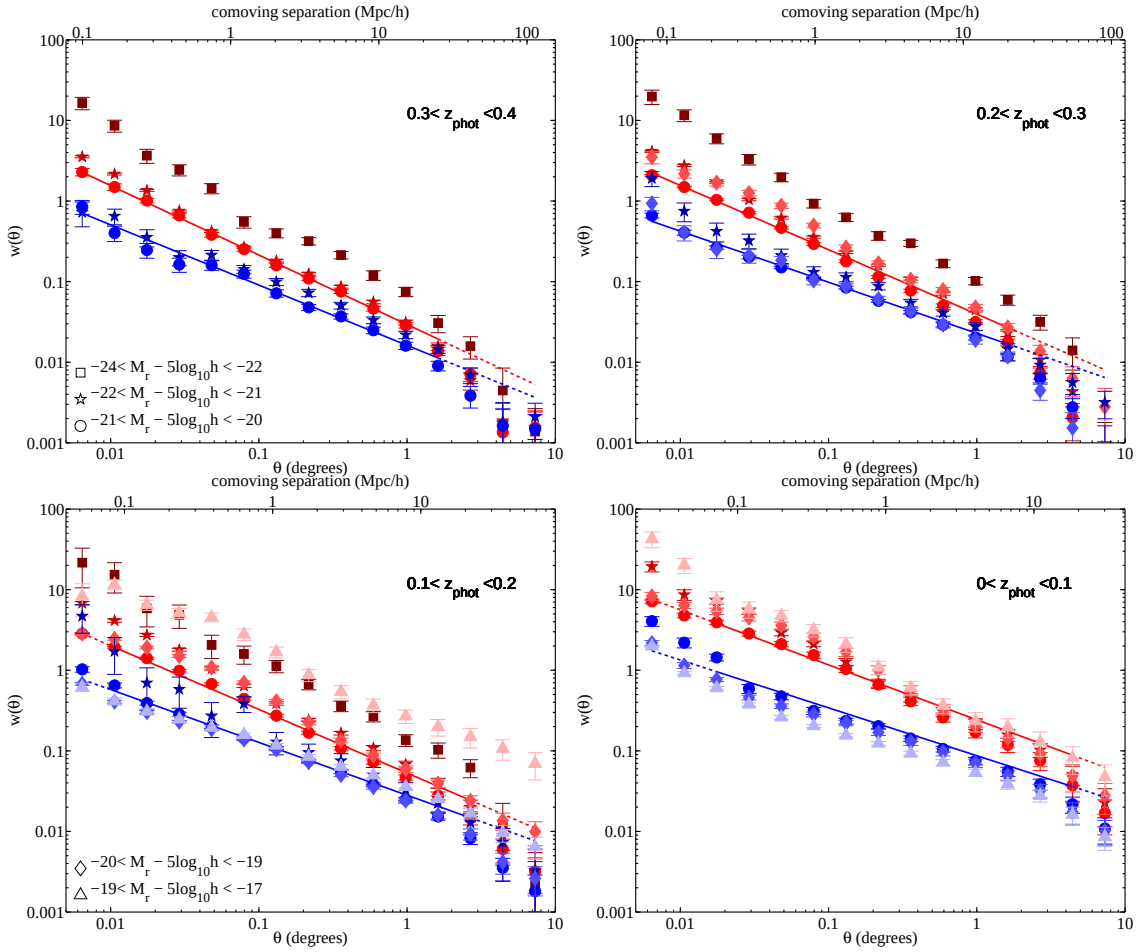

Figure 5.4: Two-point angular correlation functions $w(\theta)$ split by absolute magnitude and colour, with red circles (blue squares) showing the red (blue) sample. Colour gradients indicate the transition from bright (darker shade) to faint (lighter shade) luminosities. Lines are as in Fig. 5.2. The faintest (brightest) sample does not contain enough red (blue) galaxies to robustly estimate $w(\theta)$.

Table 5.3: Clustering properties of luminosity-selected blue galaxies. Columns are the same as in Table 5.1.

Sample	Magnitude$^{(\mathrm{MC})}$	N_{gal}	γ	r_0	χ^2_ν	$\gamma^{(d)}$	$r_0^{(d)}$	${\chi^{(d)}}^2_\nu$	b/b^*
$M_r - 5\log h$	$M_r - 5\log h$			$[h^{-1}\mathrm{Mpc}]$			$[h^{-1}\mathrm{Mpc}]$		
			Blue	$0.3 < z_{\mathrm{phot}} < 0.4$					
$[-22,-21)$	$-21.2^{-0.3}_{+0.3}$	52212	1.71 ± 0.07	6.88 ± 0.47	0.78	1.72 ± 0.07	6.87 ± 0.38	0.6	1.14 ± 0.12
$[-21,-20)$	$-20.8^{-0.2}_{+0.3}$	79787	1.75 ± 0.06	5.86 ± 0.49	1.52	1.75 ± 0.10	5.83 ± 0.44	1.3	1.00 ± 0.01
			Blue	$0.2 < z_{\mathrm{phot}} < 0.3$					
$[-22,-21)$	$-21.2^{-0.3}_{+0.3}$	42517	1.74 ± 0.11	6.42 ± 0.81	3.05	1.75 ± 0.12	6.46 ± 0.57	1.5	1.17 ± 0.14
$[-21,-20)$	$-20.4^{-0.4}_{+0.4}$	332861	1.63 ± 0.06	5.35 ± 0.48	4.08	1.66 ± 0.05	5.23 ± 0.23	2.6	0.99 ± 0.01
$[-20,-19)$	$-19.8^{-0.3}_{+0.3}$	78282	1.72 ± 0.09	5.08 ± 0.47	1.69	1.72 ± 0.09	4.88 ± 0.34	1.2	0.95 ± 0.11
			Blue	$0.1 < z_{\mathrm{phot}} < 0.2$					
$[-22,-21)$	$-21.1^{-0.4}_{+0.4}$	12753	1.85 ± 0.13	5.70 ± 0.83	0.86	1.85 ± 0.16	5.67 ± 0.64	0.6	1.22 ± 0.17
$[-21,-20)$	$-20.3^{-0.5}_{+0.5}$	210837	1.67 ± 0.07	4.43 ± 0.32	3.54	1.70 ± 0.06	4.44 ± 0.25	2.6	0.98 ± 0.35
$[-20,-19)$	$-19.4^{-0.5}_{+0.5}$	571748	1.57 ± 0.08	4.75 ± 0.73	11.72	1.62 ± 0.09	4.45 ± 0.42	6.9	1.04 ± 0.14
$[-19,-17)$	$-18.6^{-0.4}_{+0.6}$	198394	1.53 ± 0.06	4.50 ± 0.49	2.26	1.56 ± 0.06	4.31 ± 0.23	1.2	1.00 ± 0.10
			Blue	$0.0 < z_{\mathrm{phot}} < 0.1$					
$[-21,-20)$	$-20.3^{-0.7}_{+0.9}$	39246	1.61 ± 0.14	4.84 ± 0.82	6.52	1.65 ± 0.13	4.66 ± 0.31	3.2	0.97 ± 0.10
$[-20,-19)$	$-19.3^{-0.7}_{+0.9}$	109606	1.53 ± 0.06	4.63 ± 0.45	2.42	1.57 ± 0.07	4.45 ± 0.40	2.4	0.94 ± 0.21
$[-19,-17)$	$-18.1^{-0.8}_{+1.0}$	264699	1.54 ± 0.08	4.16 ± 0.63	7.29	1.58 ± 0.11	3.85 ± 0.30	4.4	0.86 ± 0.22
$[-17,-14)$	$-16.6^{-0.9}_{+1.3}$	14305	2.02 ± 0.23	4.17 ± 1.41	5.05	1.99 ± 0.28	4.34 ± 1.00	2.1	0.82 ± 0.33

5.5 Luminosity, redshift and colour dependence

We repeat the clustering analysis splitting the samples into red and blue colour using equation 4.4. For each new sample we re-estimate the underlying redshift distribution used in the inversion of Limbers equation. The corresponding 50^{th}, 16^{th} and 84^{th} percentiles of the underlying absolute magnitude distributions are given in Tables 5.2 and 5.3. We also repeat the procedure outlined in Section 5.4 for the error estimation.

In Fig. 5.4 we present the angular correlation functions in each luminosity and photo-z bin, for red and blue galaxies. The power law fits over approximately fixed comoving scales, their corresponding errors as well as the quality of the fits and the correlation length are estimated as in Section 5.4 and summarized in Tables 5.2 and 5.3. As noted earlier, the power law fits describe the clustering measurements quite well in a qualitative sense, although certainly not well enough in a quantitative sense, with most samples presenting a typically too large reduced χ^2 (see Tables 5.2 and 5.3).

For all absolute magnitude ranges, the red population displays a steeper correlation function slope than the blue one. Blue galaxies have a much shallower slope which gradually decreases with luminosity until a sudden increase in the slope for the faintest luminosity range probed (Table 5.3).

The correlation length of red galaxies for all redshift bins presents a minimum value around M^*, with increasing values both faintwards and brightwards (Table 5.2). We note however, that this result comes with large uncertainties. For red galaxies the correlation lengths of the brightest and faintest bin are comparable and faint red objects are more strongly clustered than red objects with intermediate luminosities. For the blue population r_0 behaves more regularly (like the overall population), gradually decreasing with luminosity and redshift. Blue galaxies generally have smaller uncertainties as well. Our measurement of the correlation length for the faintest luminosity bin ($r_0 = 4.17 \pm 1.41\ h^{-1}$Mpc) indicates that these galaxies are similarly clustered to blue galaxies of intermediate luminosity. The robustness of this result and some caveats are discussed in Section 5.6.

Due to the complicated way that the slope and the correlation length, as well as their respective uncertainties, change between colour selected samples, we chose to study more quantitatively the clustering of these samples using the relative bias, i.e. their clustering with respect to the L^* sample. Our relative bias results for all samples, selected by photometric redshift, absolute luminosity and colour, are presented in Section 5.8.1.

5.6 Clustering of faint blue galaxies

One of the aims of this Chapter is to study the clustering of intrinsically faint galaxies for which only photometric redshifts are available in sufficient numbers to reliably calculate $w(\theta)$. The GAMA depth and the extensive SDSS sky coverage allow us to measure the auto-correlation function of the faintest optically selected galaxies, i.e. with photo-z estimated absolute magnitudes in the $-17 < M_r - 5\log h < -14$ range and $z_{\rm phot} < 0.08$. This faint sample contains a total of 14,659 galaxies, which are mostly star-forming (as evident by their colours). From the subset with spectroscopic redshifts, the 68-central percentile of the actual absolute magnitude distribution covers the range $-18 < M_r - 5\log h < -12.7$.

The correlation function of the faintest sample $[-17,\ -14)$ exhibits a seemingly unnatural clustering amplitude at small scales. This increase in the clustering signal is not hinted in the $-19 < M_r - 5\log h < -17$ luminosity bin and for that reason we firstly investigate whether there is some sort of contamination in the data set.

We randomly select ~ 10 per cent of the objects in the faintest luminosity bin and we visually inspect them to see if they are genuine galaxies. The fraction of spurious objects is shown in the left panel of Fig. 5.5 and we observe that it is significant at the very faint end, where the actual number of galaxies is low (red line in the same figure), and ~ 40 per cent at the bright end of that luminosity bin. From our visual inspection most spurious

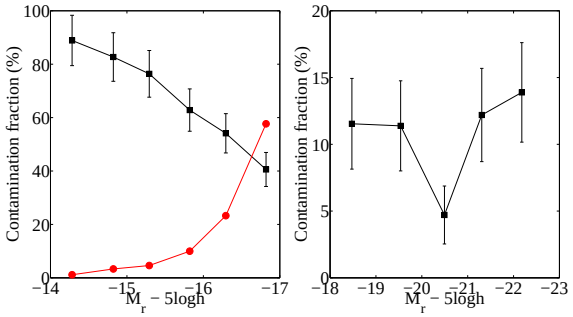

Figure 5.5: Left panel: Black symbols show the fraction of spurious objects for the faintest luminosity bin as a function of absolute luminosity. These fractions are estimated by visually inspecting ~ 10 per cent of the total number of objects in that bin. Red symbols show the overall distribution of objects as a function of absolute magnitude. Right panel: Fraction of spurious objects as function of absolute luminosity, obtained by visually inspecting a small subset of all objects in the luminosity bins defined in Sec. 4.3.4. In both panels the error bars are obtained assuming Poisson errors.

objects are local, deblended, spiral galaxies and a few of them are merging systems or just sky noise. Evidently as we go fainter the contamination level is increasing and this presents a serious drawback for clustering studies using photo-zs.

The right panel of Fig. 5.5 shows the fraction of spurious objects in the other five absolute magnitude bins. We visually inspected ~ 100 objects from each of those bins and we found that the contamination level is much lower, with an expected increase for the brightest bin. The detailed study of the correlation function of the faintest bin showed that it is not affected by contamination on the scales of primary interest, something which we expect to hold true for the brightest bin, which has a significantly smaller fraction of spurious objects.

The contamination in the $-17 < M_r - 5\log h < -14$ luminosity bin is going to affect the two point correlation function differently at different angular scales. We address this issue by counting the number of pairs of genuine galaxies in the visually inspected subset. The results are shown in Fig. 5.6. Due to the fact that the subset has a weakened signal at very small scales we can only draw conclusions for angular scales > 0.1 degrees. From Fig. 5.6 we see that at these scales the contamination doesn't affect the correlation function and the γ and r_0 measurements. For this reason we present our results, limited to the angular scales that we trust, using the total sample, which has smaller statistical errors.

We also tried repeating the calculations and masking out the areas in the sky covered

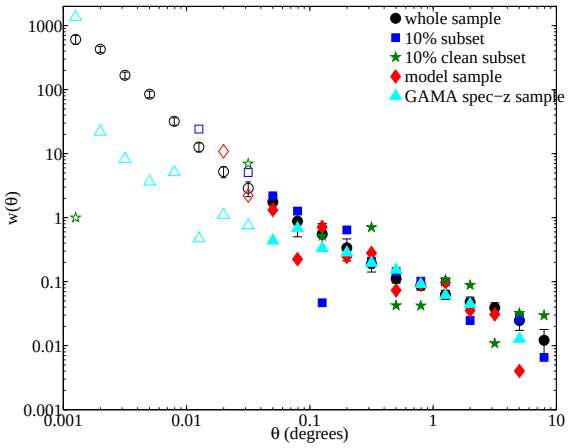

Figure 5.6: Two point correlation function of the faintest luminosity bin. Black circles show the total correlation function, blue squares show the correlation function of the ~ 10 per cent subset of objects that we visually inspected, green stars show the correlation function of the "clean" part of the previously mentioned subset and finally red diamonds show the total correlation function corrected to account for the spurious pairs on scales $\gtrsim 0.1$ degrees. Errors bars for the total sample are calculated using the JK method.

by RC3 galaxies (de Vaucouleurs et al., 1991; Corwin et al., 1994) to test whether that would decrease the contamination level. We did not observe any qualitative differences in the power law parameters estimated and more importantly, the amplitude of $w(\theta)$ at small scales did not reduce, indicating that the RC3 catalogue doesn't capture all over-deblended galaxies in the SDSS galaxy catalogue.

The upper panel of Fig. 5.7 shows the correlation functions of all galaxies in our sample with $z_{\rm phot} < 0.08$ split into finer luminosity bins than used previously. There exists a seemingly artificial steepening of $w(\theta)$ on scales $\theta < 0.1^\circ$ for galaxies with $M_r - 5\log h > -17$. In the bottom panel of Fig. 5.7, we further split the $-17.9 < M_r - 5\log h < -14$ range into two finer luminosity bins, and again we find that for fainter samples, source contamination affects larger angular scales.

Having established the angular scales over which we trust our $w(\theta)$ measurements, we proceed to the clustering analysis. Using only the diagonal elements of the covariance matrix[4], we note that a power law describes the clustering signal rather well, even though there is a hint of an increase in the clustering strength at $\sim 1\ h^{-1}$Mpc. It is possible that this increase is due to blue galaxies that are satellites in small dark matter halos.

[4]Use of diagonal covariance elements only is appropriate for this faint sample, as it covers a rather small volume for which JK resampling is unable to provide an accurate description of the full covariance matrix.

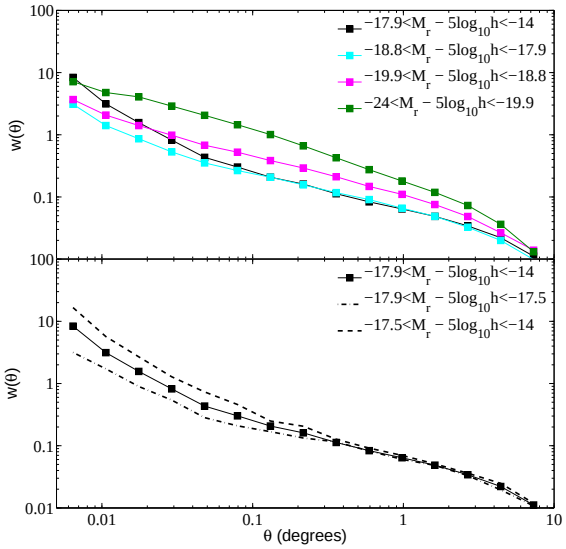

Figure 5.7: Angular correlation functions for the low redshift galaxies in our sample split in luminosity bins. The finer luminosity binning allows one to track the scales where contamination effects are significant. Error bars have been omitted for clarity.

These halos should not be dense enough to stop star formation and thus we observe only blue galaxies in this luminosity range (Eminian, 2008). A recent detailed study of the star formation history of Hα -selected faint blue galaxies in GAMA can be found in Brough et al. (2011).

In conclusion, the angular clustering for the faintest sample has a spurious amplitude at small angular scales, unless one takes into account the sample contamination. We visually inspected ~ 10 per cent of the objects in this sample and we found that a significant fraction of them are spurious, mainly due to poorly deblended sources. We quantified the effect of this contamination for all luminosity bins. This investigation revealed that the angular clustering results on scales $\lesssim 0.1$ degrees are not trustworthy enough to be considered reliable. We note that the power law fits are performed on larger scales, which we show are unaffected by this contamination. However, much more detailed investigation of the data is required to robustly confirm the observed increase in the slope of the correlation function. Finally, we note that we have repeated the analysis presented in this Section for objects selected from the most recent SDSS release, DR8 (Aihara et al., 2011), and we observe no differences in the results. The contamination from over-deblended spiral galaxies is still present in DR8 for the low luminosity bin.

Finally, it is important to note (and caution) that the source contamination due to over-deblending only became apparent when interpreting Figs. 5.7. Had we completely

trusted the results of the scaling test (Section 4.4.1) or only used the data point around L^* in Fig. 5.5 (since that population dominates), we would have significantly underestimated the number of spurious objects.

5.7 Quality of fits and the HOD formalism

The power law fits presented in Table 5.1 are not all satisfactory in a quantitative sense. The angular correlation function is only to first order well-described by a power law. The rather high reduced χ^2 for some samples are either due to underestimated errors or due to the power law model being inadequate in describing the angular correlation function over a large range of scales. From the test of Section 5.3, we conclude that the JK method gives consistent errors irrespective of the way we define the jackknife regions, and therefore it is most likely that the large reduced χ^2 values are more due to a limitation in the power law model rather than in the error estimates themselves.

A more sophisticated model, like the halo occupation distribution (HOD) model (for a review see Cooray and Sheth, 2002), would provide a more physically motivated description of the full correlation function shape, both as a function of colour and luminosity (Zehavi et al., 2004; Zheng et al., 2005; Zehavi et al., 2005, 2011). In this formalism galaxies are formed in virialized dark matter halos and thus the mass and the angular momentum of the parent halo determine the properties of the galaxies residing in the halo. Red galaxies which have their star formation suppressed tend to reside in dense environments, whereas blue galaxies with ongoing star formation live in less dense environments. In a similar manner, luminous galaxies reside in more massive halos and thus are more strongly clustered, since massive halos are more strongly clustered themselves. The HOD framework, as shown by Zehavi et al. (2005), explains the increase of clustering in the faint red population. Bright red galaxies are central galaxies in massive halos, whereas faint red galaxies are satellite galaxies in massive halos. The issue of actual fraction of faint red galaxies that are satellites in their halos has not yet been conclusively settled (Zehavi et al., 2011; Ross et al., 2011b). Our measurements suggest that both bright and faint red galaxies are more strongly clustered than red galaxies with intermediate luminosity. We also observe a bump in the angular correlation function of red galaxies at separations $\sim 1\ h^{-1}$Mpc which signals the transition (change in slope) between the one-halo and two-halo term in the correlation function. On the contrary, such a change in slope is not evident for the blue population, hence they have a smaller χ^2_ν. This is also in agreement with HOD predictions, which predict a simple power law for blue galaxies

with luminosities $M_r - 5\log h < -21$ (Zehavi et al., 2005). A complete HOD modelling of these angular clustering results with photometric redshifts is beyond the scope of the present work, as this would require photo-z dedicated HOD tools to be developed as the standard threshold samples cannot be defined.

5.8 Bias measurements

5.8.1 Relative bias and comparison with previous studies

In this Chapter we parametrize the real space correlation function with a power law, and infer $\xi(r)$ from angular clustering measurements via a Limber inversion. To ease comparison with samples using similar, but not identical, selection, we follow Norberg et al. (2002) and define the relative bias of a class of galaxies i with respect to our L^* ($-21 < M_r - 5\log h < -20$) sample as

$$\frac{b_i}{b^*}(r) = \sqrt{\frac{(r_0^i)^{\gamma_i}}{r_0^{\gamma}} r^{\gamma-\gamma_i}}. \tag{5.2}$$

Equation 5.2 preserves any scale dependence for samples with different slopes and we choose here to estimate the relative bias at $r = 5\ h^{-1}$ Mpc. The advantage of using this definition of relative bias instead of the raw correlation length to compare with other studies is twofold. First, the former uses the slope as well as the correlation length, which as we know from equation 2.66 are strongly correlated. Second, if the sample selections are just slightly different, the relative bias is a much more robust way of comparing them as it measures deviations from a series of appropriate reference samples. In this study this is particularly important, as photo-z inferred properties are not straightforwardly related to the underlying ones, as shown in Section 4.3.5. Our results are shown in Fig. 5.8.

Previous studies from both 2dFGRS (Norberg et al., 2001, 2002) and SDSS (Zehavi et al., 2002, 2005, 2011) have established that the relative bias, b/b^*, as a function of relative luminosity, L/L^*, is well described by an affine relation. We compare our results with these studies in Fig. 5.8. For all luminosity bins given in Table 5.1 we fit the equation

$$b/b^* = a_0 + a_1 L/L^*, \tag{5.3}$$

where a_0 and a_1 are free parameters. Our best fit values for samples selected on luminosity, colour and photo-z, using the corresponding L^* for each sample, are given in Table 5.4. The high redshift bin only provides three data points and thus we do not include it in this exercise (black squares in Fig. 5.8). In this Table we also compare with the bias relation of Norberg et al. (2001) who found $(a_0, a_1) = (0.85, 0.15)$. The $\Delta\chi^2$ between our best fit and

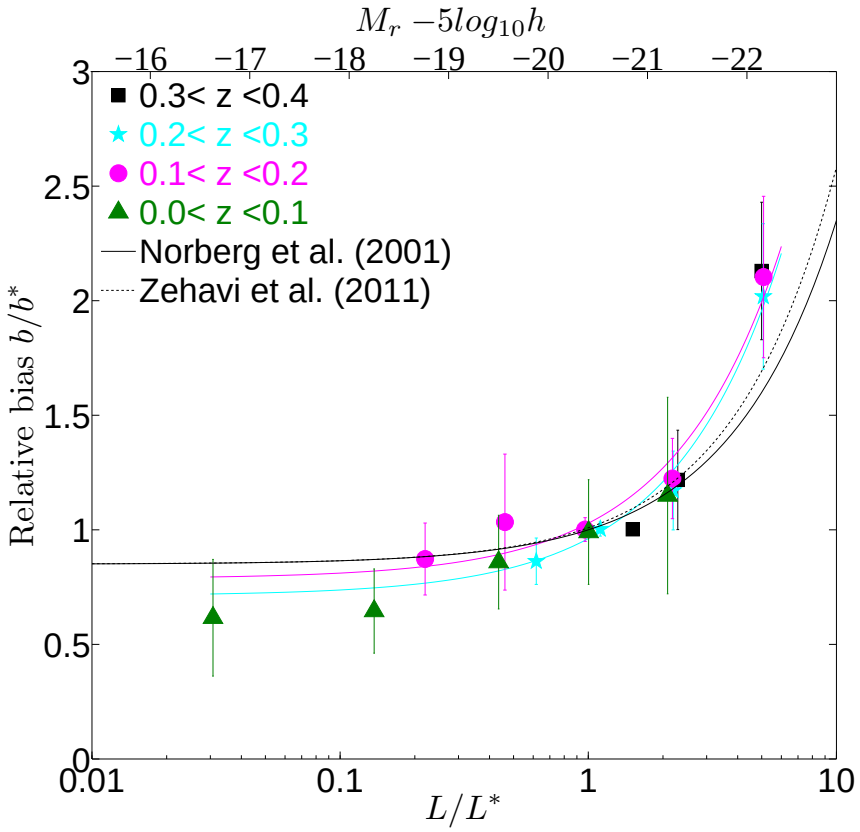

Figure 5.8: The relative bias, defined in equation 5.2, at separations $r = 5\ h^{-1}$ Mpc, of all the absolute magnitude selected samples used in this study. Data points show the mean and errors of b/b^* obtained from the distribution of 80 JK measurements (Sec. 5.3) appropriately scaled to account for the jackknife correlations. Cyan and magenta lines show our fits over the redshift ranges $0.2 < z_{\rm phot} < 0.3$ and $0.1 < z_{\rm phot} < 0.2$ respectively. The solid black line shows the fit of Norberg et al. (2001) and the dotted line the fit of Zehavi et al. (2011).

that of Norberg et al. is 1.2 to 2.3, which makes the fits statistically compatible, as the 68% confidence interval for 2 degrees of freedom corresponds to $\Delta\chi^2 = 2.31$ (Press et al., 1992). Zehavi et al. (2011) measured the bias relative to dark matter, and in Fig. 5.8 we rescale their relation with respect to L^*. They also observed a steeper rise in relative bias at high luminosities. Including a power of (L/L^*) in our fit, we also obtain a steeper slope whilst χ^2 remains unchanged, despite the additional degree of freedom.

For samples selected by colour as well as luminosity, it is more difficult to fit equation 5.3 in each redshift bin. For most photo-z bins we have four or fewer data points. Moreover, using finer luminosity bins would worsen the statistical errors on $N(z)$ and $N(M_r)$ and thus make any fit more difficult to interpret. Fig. 5.9 shows that the blue population follows a similar trend to the full sample but the relative bias changes more smoothly as a function of luminosity. Table 5.4 gives the values of a_0 and a_1 for the colour selected samples. We fit the same linear relation for red galaxies as well, despite the fact that a quadratic function would seem more appropriate. χ^2 values for the linear fit are also shown in Tabel 5.4 and from a purely statistical point of view, a linear relation between b/b^* and L/L^* is still acceptable. Fig. 5.9 shows that the statistical uncertainty for the two faint red samples is quite large. This is due to the small number of objects in the $-19 < M_r - 5\log h < -17$ sample and due to the poor quality of fit for the $-20 < M_r - 5\log h < -19$ sample.

5.8.2 The evolution of absolute bias for L^* galaxies

In Section 5.8.1 we calculated the relative galaxy bias using the L^* sample ($-21 < M_r - 5\log h < -20$) as our reference sample. In this Section we calculate the absolute bias of the L^* population defined as the mean ratio of the observed galaxy correlation function, parametrized with a *power law*, over the non-linear dark matter theoretical correlation function

$$b^*(r) = \sqrt{\frac{\xi_{GG}(r)}{\xi_{DM}(r)}} = \sqrt{\frac{(r_0^*)^{\gamma^*}}{r^{\gamma^*}\xi_{DM}(r)}}, \tag{5.4}$$

where 5 h^{-1}Mpc $< r <$ 20 h^{-1}Mpc. The theoretical power spectrum $P(k)$, was obtained using `CAMB` (Lewis et al., 2000) and the halo correction recipe of Smith et al. (2003). We then Fourier transform the non-linear $P(k)$ to obtain the real space $\xi_{DM}(r)$ using the `FFTLog` package provided by Hamilton (2000).

Since we have correlation function measurements of the L^* population for a range of redshifts we can answer the question of whether the evolution of the bias can be described

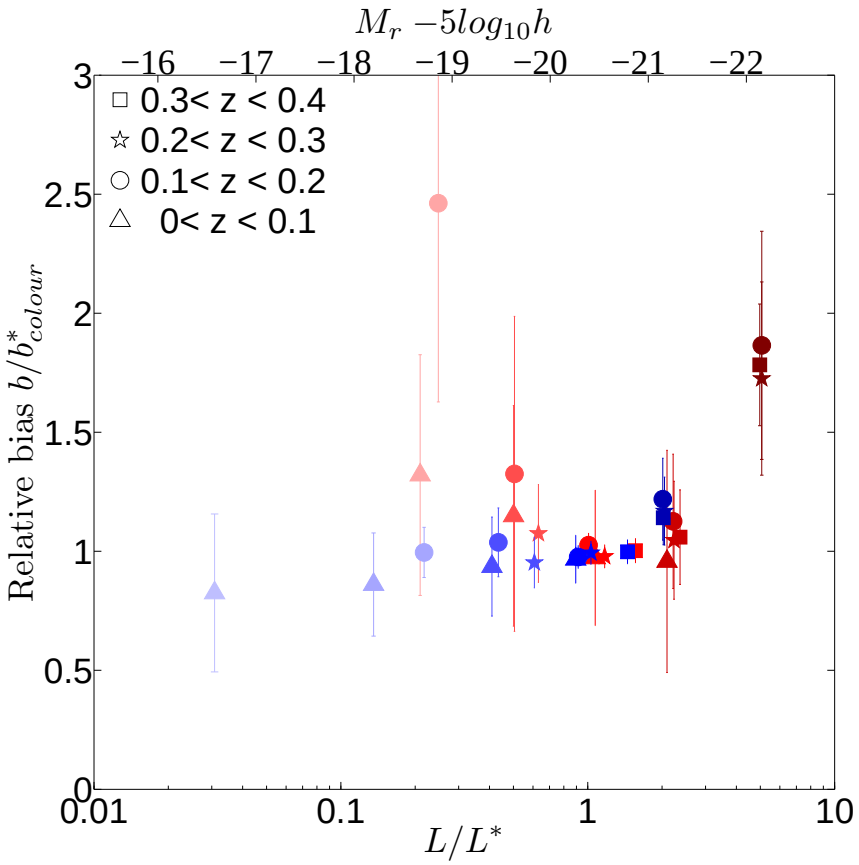

Figure 5.9: The relative bias, defined in equation 5.2, at separations $r = 5\ h^{-1}$ Mpc, of all the samples used in this study split by colour (equation 4.4). Data points show the mean and errors of b/b^* obtained from the distribution of 80 jackknife measurements (Sec. 5.3) appropriately scaled to account for the jackknife correlations. Colour coding is as in Fig. 5.4.

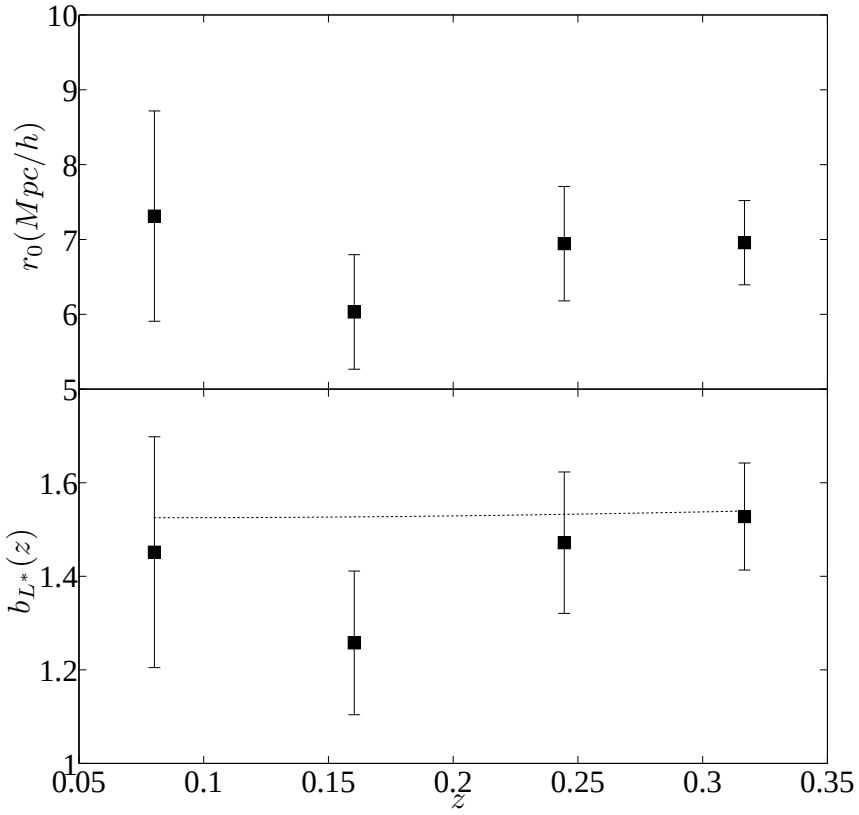

Figure 5.10: The evolution of clustering of L^* galaxies in the local universe: Upper panel shows the correlation length r_0; lower panel shows the bias $b_{L^*}(z)$, as a function of redshift. The dashed line in the lower panel shows the linear theory prediction from equation 5.5. Across the redshift range $0.07 < z < 0.32$ the bias of L^* galaxies agrees rather well with the linear theory model.

Table 5.4: Fitted values of a_0 and a_1 in the bias–luminosity relation (equation 5.3) in three photo-z ranges. Column 1 lists the redshift bin limits, columns 2, 3 and 4 the fitted values and the quality of fit (reduced χ^2) and column 5 lists $\Delta\chi^2$ between our best fit values and the fit by Norberg et al. (2001).

Redshift range	a_0	a_1	χ^2_ν	$\Delta\chi^2$
All colours				
$0.2 < z_{\rm phot} < 0.3$	0.71 ± 0.04	0.25 ± 0.02	1.10	2.32
$0.1 < z_{\rm phot} < 0.2$	0.82 ± 0.06	0.24 ± 0.03	0.14	1.79
$0.0 < z_{\rm phot} < 0.1$	0.65 ± 0.05	0.27 ± 0.06	0.12	1.18
Red				
$0.2 < z_{\rm phot} < 0.3$	0.92 ± 0.17	0.12 ± 0.07	0.36	0.29
$0.1 < z_{\rm phot} < 0.2$	1.28 ± 0.43	0.03 ± 0.17	2.33	1.76
Blue				
$0.2 < z_{\rm phot} < 0.3$	0.84 ± 0.08	0.15 ± 0.06	0.29	0.77
$0.1 < z_{\rm phot} < 0.2$	0.98 ± 0.07	0.08 ± 0.06	0.23	4.22
$0.0 < z_{\rm phot} < 0.1$	0.86 ± 0.02	0.08 ± 0.02	0.07	0.02

by the passive evolution model introduced by Tegmark and Peebles (1998):

$$[b(z_1) - 1]D(z_1) = [b(z_2) - 1]D(z_2), \tag{5.5}$$

where D is the growth of structure (Peebles, 1980) which we calculate accurately using the `growl` package by Hamilton (2001), which includes corrections to $D(z)$ due to the presence of the cosmological constant. The model described by equation 5.5 assumes that the galaxy density field linearly traces the dark matter density field and all clustering evolution comes from the growth of structure in the linear regime, i.e. no merging. It is believed that L^* galaxies have undergone very little merging since $z \approx 1$ (Conselice et al., 2009; Lotz et al., 2011).

In the upper panel of Fig. 5.10 we plot the correlation length as a function of redshift. r_0 is observed to change very little since $z \approx 0.32$. The lowest redshift point has larger errors due to the limited volume sampled. For comparisons with theory, it is more lucid to use the bias instead of the correlation length. In the lower panel of Fig. 5.10 we plot the evolution of the absolute bias, as defined in equation 5.4, along with the theoretical prediction of Tegmark and Peebles (1998) for passive clustering evolution (dashed line). In practice, we fix the high-z value of $b(z)$ and then solve equation 5.5 over the redshift

range $0.07 < z < 0.32$. We find that the evolution of clustering of L^* galaxies is consistent with the model of Tegmark and Peebles (1998).

This agreement between the clustering of L^* galaxies and the passive evolution model was not observed by Ross et al. (2010) who used SDSS photo-z's. The sample selection and the modeling of $w(\theta)$ and bias between this study and the one by Ross et al. (2010) are very different, as we use GAMA calibrated photo-z and model the correlation function with a power law, whereas they used SDSS calibrated photo-z down to $r < 21$ and use halo modelling for the correlation function. Ideally one would expect that the two studies should give consistent results, but it might be that the aforementioned differences in the theoretical modelling and the sample selection influence the results significantly.

5.9 Discussion and conclusions

Despite their inherent limitations, photometric redshifts offer the opportunity to study the clustering of various galaxy populations using large numbers of objects over a wide range of angular scales with improved statistics, with the caveat that their systematic uncertainties are significantly more complex to deal with. In this section we summarize and discuss the main implications of our results.

Using GAMA spectroscopic redshifts as a training set, we have compiled a photometric redshift catalogue for the SDSS DR7 imaging catalogue with $r_{\rm petro} < 19.4$. We carried out extensive tests to check the robustness of the photo-z estimates and use them for calculating r-band absolute luminosities. We split our sample of 4,289,223 galaxies into samples selected on photometric redshift, colour and luminosity and estimate their two point angular correlation functions. Redshift distributions for the Limber inversion are calculated using Monte-Carlo resampling, which we show are very reliable.

Our clustering results are in agreement with other clustering studies such as Norberg et al. (2002) and Zehavi et al. (2011) who used spectroscopic redshifts. We extend the analysis to faint galaxies where photo-zs allow us to obtain representative numbers for clustering statistics. We find that the correlation length decreases almost monotonically toward fainter absolute magnitudes and that the linear relation between b/b^* and L/L^* holds down to luminosities $L \sim 0.03L^*$. For the L^* population we observe a bias evolution consistent with the passive evolution model proposed by Tegmark and Peebles (1998).

As shown by others (Norberg et al., 2002; Hogg et al., 2003; Zehavi et al., 2005; Swanson et al., 2008a; Zehavi et al., 2011) and confirmed here, the colour dependence is more intriguing because faint red galaxies exhibit a larger correlation length than red galaxies at inter-

mediate luminosities. This trend is explained by HOD models, as shown by Zehavi et al. (2005). Clustering for blue galaxies depends much more weakly on luminosity. We find that at faint magnitudes the SDSS imaging catalogue is badly contaminated by shreds of over-deblended spiral galaxies, which makes the interpretation of the clustering measurements difficult. We determine an angular scale beyond which our results are not affected by this contamination, and test this by modelling the scale-dependance of the contamination as well as studying its luminosity dependence.

The use of photometric redshifts is likely to dominate galaxy clustering studies in the future. A number of assumptions made in this work might need to be reviewed when we have even better imaging data and training sets. In particular, for cosmology, the non-Gaussianity of photo-z and robust reconstruction of redshift distributions will become a very pressing issue. For galaxy evolution studies, it is essential to study the mapping between a photo-z derived luminosity range and the true underlying one, as HOD modelling of the galaxy two point correlation function relies heavily on the luminosity range considered. In this Chapter, we report only qualitative agreement and leave any HOD study using these photometric redshift inferred clustering results to future work.

Chapter 6

Galaxy clustering and redshift space distortions from GAMA

6.1 Introduction

Galaxy clustering analysis in redshift space provides a complementary set of observables to the projected two-point functions for galaxy evolution and cosmology. These observables are dynamical quantities that depend on key properties of the cosmological model like the matter budget of the Universe, the density fluctuation amplitude and the growth rate of structure. They also depend on fundamental properties of cold dark matter, like the structure of bound objects (dark matter haloes) and galaxy evolution, like the velocity distribution of objects within their parental haloes. A systematic way to study these effects is to use the peculiar motions of galaxies outside the Hubble flow. Redshift-based distance measurements contaminated by peculiar velocities break the clustering isotropy observed in real space, a phenomenon known as redshift-space distortions (rsd). Thus redshift space observations require a special interpretation, as it appears that we cannot observe directly the background universe as described by the perturbed FLRW metric, yet despite of these complications, redshift-based distance measurements come with some advantages as well. Redshift-space distortions are caused directly by the inhomogeneities in the universe and studying them it is possible to extract information about the growth rate of structure and the dynamics that affect the small scale clustering of dark matter haloes.

The linear theory of rsd is been known for some time (Kaiser, 1987; see Hamilton, 1998, for a an extensive review) and its main aspects are introduced in Section 6.2. It's also well-known that peculiar velocities are a direct probe of the mass density of the

universe (Peebles, 1980, § 76; Hamilton, 1992). Indeed, some of the first indications for a low-Ω_m universe, which were in apparent disagreement with the $\Omega_{\rm tot} = 1$ prediction from inflation, came from direct peculiar velocity estimations (Davis and Peebles, 1983) and the study of the anisotropic correlation function, through the linear redshift distortion parameter β (see Section 8 of Hamilton, 1998 and references therein). In this picture matter tracers (galaxies) that exist as correlated pairs coherently fall into large structures, or equivalently empty voids. Since redshift determines the line of sight distance (LOS) of objects, clustering in redshift space along the LOS is seemingly enhanced by these motions. Thus, the isotropy of the real space correlation function is broken.

Small scale anisotropic galaxy clustering is more difficult to model, due to the combination of the non-linearities in the real space correlation function and the added redshift distortion. A phenomenological approach that has proven very useful is adding a dispersion function on small scales, where galaxies are bound by gravity. Formally, the Kaiser (1987) model cannot describe rsd effects of these pairs of galaxies as it is only valid in the linear regime. However, adding a dispersion function with a small coherent length to the Kaiser (1987) model disentangles the different contributing scales to rsd. In this simple model there exist two free parameters. The growth rate of structure, which dominates on large scales and the pairwise velocity dispersion which dominates on small scales. Both quantities are important dynamical observables which can constrain cosmological models and provide insights into galaxy formation and evolution (Berlind and Weinberg, 2002; Zheng and Weinberg, 2007; Tinker et al., 2007).

In this Chapter we are interested in studying the small to intermediate scales of the anisotropic galaxy correlation function. In particular, we measure the pairwise velocity dispersion (pvd) as a function of scale and luminosity using the GAMA data described in Section 3.2. A handfull of studies have exploited recent large galaxy surveys (2dFGRS and SDSS) to explicitly study the galaxy pvd (Hawkins et al., 2003; Jing and Börner, 2004; Li et al., 2006, 2007; Tinker et al., 2007; Cabré and Gaztañaga, 2009b). Albeit the growth rate is extremely interesting to study as well, GAMA data is superior for constraining the pvd and here we aim to produce measurements of pvd as a function of luminosity and redshift.

Observations in redshift-space introduce a preferable direction along the line of sight and therefore two-point statistics are now functions of the angle that separates the two objects, as well as their distance. From the binning of pairs as function of distances s and angles μ one can recover the multipoles of the correlation function by integrating over all

angular separations

$$\xi_\ell(s) = \frac{2\ell+1}{2}\int_{-1}^{+1}\xi(s,\mu)P_\ell(\mu)d\mu, \tag{6.1}$$

where $P_\ell(\mu)$ are the well-known Legendre polynomials. Setting $\ell = 0$ one can recover the monopole of the spatial correlation function, also known as the angle-averaged redshift-space correlation function, which is the quantity closer to the real space correlation function (see Scoccimarro, 2004, for theoretical limitations of this method). Switching from angles to distances only one can define the parallel ($s_\parallel$) and transverse ($s_\perp$) separations as follows (Fisher et al., 1994)

$$s_\parallel = \frac{\mathbf{s}\cdot\mathbf{r}_{\rm los}}{|\mathbf{r}_{\rm los}|} \tag{6.2}$$

$$s_\perp = \sqrt{\mathbf{s}\cdot\mathbf{s} - {s_\parallel}^2}, \tag{6.3}$$

where $\mathbf{r}_{\rm los} = 1/2(\mathbf{s}_1 + \mathbf{s}_2)$ and $\mathbf{s} = \mathbf{s}_1 - \mathbf{s}_2$. Note that equation 6.1 holds on for the anisotropic power spectrum as well, where $\mu = k_z/k$.

In contrast with the wide application of the monopole of the correlation function in cosmology and extragalactic astronomy via well-established techniques, there is no blueprint for the use of the full two-dimensional correlation function in the literature. Two main issues cause some concern here. Firstly, which model one should fit to the data and whether that is exact or just the first few terms of the multipole expansion. As we discuss in Section 6.2.2 relative to the published studies exploiting available data, there is a disproportional amount of theoretical models most of which are strictly valid on large-scales. Secondly is the issue of the covariance matrix of the two-dimensional correlation function. The correlation function is strongly covariant and uncertainties are significantly underestimated if covariance is not taken into account. However, the construction of covariance matrix is almost prohibited for practical reasons: The number of data points used in the fits are usually ~ 1000 and therefore one has to construct a covariance matrix with $\sim 1000 \times 1000$ elements. Yet, this covariance matrix must be inverted thousands of times during the parameter fitting and such a procedure is clearly not practical. On the other hand, working solely with one-dimension observables such as the second multipole of the redshift-space correlation function, requires extra manipulation of the (usually noisy) data (since one has to integrate over angular separations to compute it; see equation 6.1). Furthermore it is not clear up to which multipole one should expand in order to to capture fully non-linear effects like the pairwise velocity dispersion.

In spite of the modeling and uncertainty issues a consensus also does not appear to exist on the fitting method of the anisotropic correlation function $\xi(s_\perp, s_\parallel)$, specifically on the

scales over which one is fitting and in the chosen parameters to fit. E.g. Guzzo et al. (2008) used the whole $\xi(s_\perp, s_\parallel)$ plane to fit simultaneously the pairwise velocity dispersion and the β parameter, whereas Cabré and Gaztañaga (2009a) fitted β on sufficiently large scales and away from small line of sight angles and then used that measurement to conditionalize their fits on σ_{12} (Cabré and Gaztañaga, 2009b). Blake et al. (2011) used only large scales ($k_{\rm max} = \sqrt{k_\perp^2 + k_\parallel^2} < 0.3\, h{\rm Mpc}^{-1}$) but fitted both parameters simultaneously. In principle, a correct model on all scales for the rsd and the underlying $\xi(r)$ should be immune to such manipulations, but a lacking a definitive model and noisy data, compromise parameter estimations using rsd.

To overcome these obstacles, we use a tailored approach for the specific problem of measuring the pvd and its uncertainties as a function of luminosity. We make use of the information from the real space correlation function in order to avoid biases in the fitting process and we take advantage of the first set of GAMA mocks to study the covariances of the measurements. We fit over the entire $\xi(s_\perp, s_\parallel)$ plane and carry tests on mock catalogues to establish the optimized scales for unbiased parameter fitting. We then interpret our results in the context of halo models of large-scale structure.

In Section 6.2 the theory behind rsd is introduced as well as the motivation for implementing a specific model. Measurements of the anisotropic correlation function from GAMA are presented in Section 6.3 and the analysis of the results in Section 6.4. In Section 6.5 we compare our results with previous studies and finally in Section 6.6 we present a discussion and our conclusions.

In this Chapter cosmological parameters primarily enter through the construction of the real space correlation function. h parametrizes the effect of the Hubble constant uncertainty on distances, distance moduli and comoving volumes through the relation $H_0 = 100h$ km s^{-1}Mpc^{-1}. In accord with the standard model we use flat ΛCDM cosmology with $\Omega_{m,0} = 0.25$, $\Omega_{b,0} = 0.05$, $n_s = 0.96$ and $\sigma_8 = 0.8$. Unless otherwise stated we use the term pairwise velocity dispersion as the width of an exponential function.

6.2 Theory

6.2.1 Real to redshift-space mapping

In the previous Chapters we have assumed that the redshift for an object that we use for deriving its comoving distance is purely cosmological, yet generically this is not the case. Instead, all observed redshifts have a non-cosmological component which arises from the

peculiar motions induced by structure formation in the universe. Thus, measuring the correlation function from galaxy redshift surveys removes its isotropy, since redshift space observations yield a modified comoving distance $\mathbf{s}$, due to peculiar velocity v, contaminating the cosmological redshift

$$\mathbf{s} = \mathbf{r} + \frac{v(\mathbf{r})}{H(z)}(1+z), \tag{6.4}$$

where z is the cosmological redshift and $\mathbf{r}$ is its respective comoving distance[1]. A direct consequence of rsd is to modify the observed cosmological density fields (Hamilton, 1998) and the therefore the frequently used two point clustering statistics. Mass is necessarily conserved and therefore

$$[1+\delta(\mathbf{s})]d\mathbf{s} = [1+\delta(\mathbf{r})]d\mathbf{r}. \tag{6.5}$$

Two-point statistics are affected by the modified distance that separates each galaxy pair and therefore one should estimate the distribution of the pairwise velocities $v_{12}(\mathbf{r}) = v_2(\mathbf{r}) - v_1(\mathbf{r})$ from equation 6.4. The distribution of $v_{12}(\mathbf{r})$ gives the pairwise velocity PDF $\mathcal{P}(v, \mathbf{r})$, which maps pairs from real to redshift space (Scoccimarro, 2004)

$$1 + \xi(s_\perp, s_\parallel) = \int [1+\xi(\mathbf{r})]\mathcal{P}(r_\parallel - s_\parallel, \mathbf{r})dr_\parallel. \tag{6.6}$$

Equation 6.6 utilizes the flat sky approximation, where all rsd effects operate only on the radial direction.

In reality, $\mathcal{P}(v, \mathbf{r})$ is a complicated function resulting from the convolution of the pairwise PDF's of all pair configurations $\mathbf{r}$ (Scoccimarro, 2004). Scoccimarro (2004) shows how the Kaiser (1987) infall ("squashing") term emerges through the pairwise velocity PDF, but also criticizes the implicit assumption of a symmetric pairwise velocity PDF.

6.2.2 A model for two point statistics in redshift-space

In linear perturbation theory the observed redshift space power spectrum $P(k)$ has the following form (Kaiser, 1987)

$$P(k, \mu) = (1 + f_g \mu^2)^2 P(k), \tag{6.7}$$

where μ is the cosine of the angle between k and $k_\parallel$ and f_g is the logarithmic derivative of the growth of structure (Peebles, 1980). A useful approximation for f_g is given by Peebles (1980) (see also Linder, 2005.)

$$f_g = \Omega_m^{0.55}. \tag{6.8}$$

[1]Hereafter, r will denote distances based on the true redshift and s distances based on the observed, contaminated with peculiar velocities redshift.

The growth rate of structure primarily depends on the matter budget of the universe and has only a weak dependence on dark energy (Carroll et al., 1992; Amendola and Tsujikawa, 2010).

These results can be translated in the configuration space either by directly Fourier transforming in 2 dimensions equation 6.7 (using the cylindrical symmetry of the problem, see equation 6 of Jing and Börner, 2004) or by expanding $P(k,\mu)$ in Legendre polynomials (in linear theory only the first 4 even Legendre polynomials are non-zero) and then Fourier transforming each term (Hamilton, 1992). The multipole expansion of $\xi(s_\perp, s_\parallel)$ has the following form

$$\xi(s_\perp, s_\parallel) = \xi_0(s)P_0(\mu) + \xi_2(s)P_2(\mu) + \xi_4(s)P_4(\mu), \tag{6.9}$$

where P_ℓ's are Legendre polynomials and $\xi_\ell(s)$ the first multipoles of $\xi(s_\perp, s_\parallel)$ given by the following relations

$$\xi_0(s) = \left(1 + \frac{2f_g}{3} + \frac{f_g^2}{5}\right)\xi(r) \tag{6.10}$$

$$\xi_2(s) = \left(\frac{4f_g}{3} + \frac{4f_g^2}{7}\right)\left[\xi(r) - \bar{\xi}(r)\right] \tag{6.11}$$

$$\xi_4(s) = \frac{8}{35}f_g^2\left[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)\right], \tag{6.12}$$

where

$$\bar{\xi}(r) = \frac{3}{r^3}\int_0^r \xi(\tilde{r})\tilde{r}^2 d\tilde{r} \tag{6.13}$$

$$\bar{\bar{\xi}}(r) = \frac{5}{r^5}\int_0^r \xi(\tilde{r})\tilde{r}^4 d\tilde{r}. \tag{6.14}$$

At small non-linear scales we expect part of the clustering signal to be wiped out due to the random motions of galaxies inside the halos. This indicates that the model $P(k,\mu)$ has to be convoluted with a damping function. It was first found by Davis and Peebles (1983) and then confirmed by subsequent observational studies (Landy et al., 1998; Landy, 2002; Hawkins et al., 2003; Jing and Börner, 2004; Li et al., 2006; Guzzo et al., 2008; Cabré and Gaztañaga, 2009b; Blake et al., 2011), as well as theoretical considerations (Sheth, 1996; Sheth and Diaferio, 2001; Cooray and Sheth, 2002) that the form of the damping function is Lorentzian, so a model of the redshift space power spectrum on all scales is given by

$$P(k,\mu) = (1 + f_g\mu^2)^2 P(k)\frac{1}{1 + k^2\sigma_{12}^2\mu^2/2} \tag{6.15}$$

and σ_{12} is the pairwise velocity dispersion. In configuration space the convolution of

equation 6.15 from a multiplication in Fourier space, becomes an integration

$$\xi(s_\perp, s_\parallel) = \frac{1}{\sqrt{2}\sigma_{12}} \int_{-\infty}^{\infty} \xi\left(s_\perp, s_\parallel - \frac{v}{(1+z)H(z)}\right) e^{-\frac{\sqrt{2}|v|}{\sigma_{12}}} dv, \tag{6.16}$$

where we took into account equation 6.4 and $\xi(s_\perp, s_\parallel)$ comes from equation 6.9. Despite the fact the integral is symmetric and the correlation function is defined only for positive separations, for numerical reasons it is better to perform it along its whole domain of integration. Eqs 6.6 and 6.16 are similar, but they differ with respect to the quantity that is convolved. In equation 6.6 the term $1 + \xi$ is convolved with a pairwise velocity PDF that applies on all scales (and the total number of pairs is proportional to $1 + \xi$), whereas equation 6.16 applies only on strictly non-linear scales that have already undergone gravitational collapse.

Physically the two effects (Kaiser and damping) act to alter the shape of the two-dimensional correlation function as follows. On large scales and for all angles galaxies fall into groups and clusters and thus dense regions are enhanced and voids are becoming more underdense. As a consequence of this, correlated pairs appear to be closer in redshift space and thus decreasing the amplitude of the correlation function along the line of sight compared to the transverse direction. On small scales the opposite effect occurs. Random, virial motions of galaxies inside the halo make the pairs to be seemingly separated by larger distances and thus increasing the amplitude of the line of sight component of the correlation function with respect to the transverse direction which remains unaffected at all times. This effect is known as finger-of-god (FOG) due to the fact that the anisotropic correlation function appears elongated towards the origin.

Despite its success over the years the "Kaiser+damping" model suffers from some shortcomings (Scoccimarro, 2004; Tinker, 2007). The most important are the lack of a direct physical interpretation of the pairwise velocity dispersion and the absence of modeling of the quasi-linear regime to describe the transition between the linear squashing of $\xi(s_\perp, s_\parallel)$ and FOG. Nevertheless, the "Kaiser+damping" model is the simplest possible one that can be applied over a wide range of scales. Moreover, a significant improvement for fitting data is achieved if one uses the non-linear corrections (Smith et al., 2003) of the matter power spectrum as an input in equation 6.15 (Blake et al., 2011).

Another shortcoming of the "Kaiser+damping" model is its inability to take explicitly into account non-linearities. As shown mathematically by Scoccimarro (2004) the very mapping from real space to redshift space is non-linear on all scales and therefore it cannot be described using linear theory, even on scales $k < 0.1h$ Mpc^{-1}. Physically these problems arise because pairs from linear scales get shifted to seemingly non-linear scales

in redshift space as shown by equations 6.6 and 6.16.

There exist a wealth of other proposed models for the redshift space $P(k, \mu)$ in the literature, either based on analytical considerations (Scoccimarro, 2004; Percival and White, 2009; Taruya et al., 2010; Reid and White, 2011) or calibrated from simulations (Jennings et al., 2011; Kwan et al., 2012) or combinations thereof (Tinker et al., 2006; Tinker, 2007). All these approaches have provided valuable insights into a complex problem. However, they all start from the linear theory power spectrum in real space (sometimes including a linear bias term) and then try to predict the evolved redshift space result and inevitably these models have a limited regime of validity and are possibly cosmology dependent via the Alcock and Paczynski (1979) effect.

Nonetheless, one can avoid modeling directly redshift space quantities and note that the real space galaxy correlation function is a direct observable and one can take advantage of it. The first step to achieve this is to integrate $\xi(s_\perp, s_\parallel)$ along the line of sight, in order to negate the redshift distortion effects on projected galaxy pairs

$$w(r_\perp) = 2\int_0^\infty \xi(r_\perp, s_\parallel) ds_\parallel = 2\int_{r_\perp}^\infty \frac{r\xi(r)}{\sqrt{r^2 - r_\perp^2}} dr, \tag{6.17}$$

where the second step is a change of integration variable using the relation $r^2 = r_\perp^2 + r_\parallel^2$. The second relation in equation 6.17 gives the real space correlation function (Lilje and Efstathiou, 1988; Saunders et al., 1992)

$$\xi(r) = -\frac{1}{\pi}\int_r^\infty \frac{w(r_\perp)}{dr_\perp}\left(r_\perp^2 - r^2\right)^{-1/2} dr_\perp. \tag{6.18}$$

Therefore the galaxy $\xi(r)$ is an observable and one can "distort" it in order to induce the rsd effects (Tinker, 2007; Guzzo et al., 2008; de la Torre and Guzzo, 2012, Peacock et al. 2012, in Prep.). The rsd model given in equation 6.16 has two free, purely dynamical parameters, namely f_g and σ_{12}.

There is, however, one last obstacle. The inversion of equation 6.18 is not perfect on large scales as shown by Norberg et al. (2009). Realistically, we cannot expect to integrate equations 6.17 and 6.18 to infinity and therefore recovering $\xi(r)$ will be biased to some extent and valid only on limited scales.

6.2.3 Observing galaxies and the scale-dependent bias

The discussion so far is valid for any pressureless density field in the universe. Nonetheless, we observe galaxies which are generically biased tracers of the dark matter density field, which drives the dynamical matter effects such as rsd. Galaxies are formed inside virialized

dark matter haloes and complicated astrophysical effects bias the relation between galaxies and dark matter (Cooray and Sheth, 2002). At large scales all the pair contributions to the galaxy correlation function come from galaxies residing in different haloes which trace the dark matter field up to a constant, scale-indepedent bias factor.

In this work we rely heavily on having the correct model $\xi_{gg}(r)$ on small scales. We thus modify the theoretical form of the dark matter correlation function $\xi_{\delta\delta}(r)$ by allowing bias to have both a scale-independent and scale-dependent component in order to obtain a realistic model for the galaxy correlation function (de la Torre and Guzzo, 2012)

$$\xi_{gg}(r) = b^2_{\rm eff}(r)\xi_{\delta\delta}(r) = b^2_L b^2_{NL}(r)\xi_{\delta\delta}(r). \tag{6.19}$$

In equation 6.19, b_L modifies the overall amplitude of $\xi_{\delta\delta}(r)$ and scales $\propto \sigma_8$, whereas b_{NL} modifies the shape of $\xi_{\delta\delta}(r)$. In Section 6.4.1 we test the effect of the scale-dependent term on our parameter fits. The model of $\xi_{gg}(r)$ given in equation 6.19 will be used as input to equation 6.16 for the construction of the model $\xi(s_\perp, s_\parallel)$.

6.3 Measuring the galaxy anisotropic correlation function

6.3.1 Correlation function estimator and binning issues

We calculate the 2-dimensional correlation function using the the Landy and Szalay (1993) estimator generalized for galaxy separations along the line of sight (LOS) $s_\parallel$ and transverse to the line of sight $s_\perp$,

$$\xi(s_\perp, s_\parallel) = \frac{DD(s_\perp, s_\parallel) - 2DR(s_\perp, s_\parallel) + RR(s_\perp, s_\parallel)}{RR(s_\perp, s_\parallel)}, \tag{6.20}$$

where DD is the normalized number of data pairs separated by a given $(s_\perp, s_\parallel)$ bin, DR is the respective normalized number of data-random pairs and RR is the respective normalized number of random-random pairs. For this calculation we use the "flat sky" approximation which assumes that galaxy separations can be disentangled as purely radial (difference in comoving distance only) and purely transverse. For wide field surveys this approximation clearly breaks down (Matsubara, 2000; Raccanelli et al., 2010), but for a relatively narrow survey like GAMA (due to the three distinct GAMA regions maximum angular separation is ≈ 12 degrees and most correlated pairs belong in the same region), it is valid to a first order approximation. GAMA is a moderately deep survey with $z_{\rm median} \approx 0.2$ and therefore the LOS component dominates pair separations.

We use equation 6.20 to calculate $\xi(s_\perp, s_\parallel)$ for linearly spaced $(s_\perp, s_\parallel)$ bins as well as logarithmically spaced. Linearly spaced $\xi(s_\perp, s_\parallel)$ is used for the parameter fitting of model

anisotropic $\xi(s_\perp, s_\parallel)$, whereas logarithmically spaced $\xi(s_\perp, s_\parallel)$ is used for the calculation of the projected correlation function (equation 6.17). Large $(s_\perp, s_\parallel)$ separations yield much more noisy measurements of $\xi(s_\perp, s_\parallel)$ and therefore one might be tempted to use bins of varying size, which in principle would remove the noise from large scales. We tested this proposition against the mocks and we found that the results on the fitted parameters were significantly biased. We conclude that for the data at hand 2-dimensional fitting to linearly spaced $(s_\perp, s_\parallel)$ grid is optimal, whereas for the reconstruction of the real space correlation function using the Saunders et al. (1992) technique we use logarithmic spacing instead (see Section 6.4.1).

6.3.2 Anisotropic correlation function for luminosity bins

We measure the redshift space correlation function $\xi(s_\perp, s_\parallel)$ in linear bins in $s_\perp$ and $s_\parallel$ of 0.5 h^{-1}Mpc and up to separations of 20 h^{-1}Mpc for six volume limited galaxy samples split in luminosity bins and redshift. Initially, we define four bins with the following limits in $M_r - 5\log h$: $(-22, -21), (-21, -20), (-20, -19), (-19 - 17)$. We then further split the two bright luminosity bins in redshift so that the two redshift bins have approximately equal number of galaxies. These volume limited samples are shown in Fig. 6.1. All absolute magnitudes have been $k + e$-corrected at $z_0 = 0$ using `KCORRECT` (Blanton and Roweis, 2007) and using the passive evolution parameter $Q = 0.7$ (described in Section 4.3.4) of Loveday et al. (2012).

For this analysis we need to account for the GAMA angular and the radial selection function. The GAMA angular selection function is described in Section 5.1 of Driver et al. (2011). For the random catalogues we first use the `MANGLE` software (Hamilton and Tegmark, 2004; Swanson et al., 2008b) to create random objects in the GAMA area (without the flat sky approximation) and then apply the GAMA imaging mask. GAMA is approximately 98% spectroscopically complete, however we do weight each data and random point according to its actual completeness at the part of the sky where it belongs (see Section 5.2 of Driver et al., 2011).

Fig. 6.2 depicts the quantity $\ln[1 + \xi(s_\perp, s_\parallel)]$ for our $L \gtrsim L^*$ samples, in order to capture the features of $\xi(s_\perp, s_\parallel)$ on all scales. Note that anisotropic correlation function measurements are by definition confined in the first quadrant, however we follow the usual practice to mirror $\xi(s_\perp, s_\parallel)$ in the other three quadrants in order to easily reveal the deviations from spherical symmetry. The main characteristics of the redshift space correlation function, the large-scale squashing along the LOS and the FOG are clearly

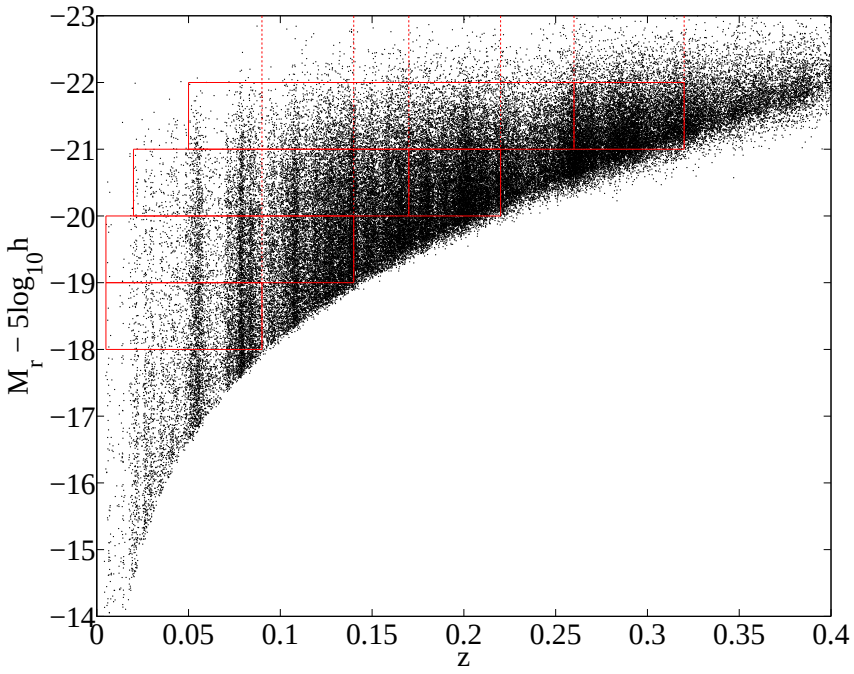

Figure 6.1: Absolute magnitude-redshift plot for GAMA data with $r < 19.4$. Overplotted are the volume limited samples we use in this study. Samples are defined in magnitude bins (solid lines) and magnitude thresholds (dashed lines). GAMA depth allows to split the two brightest samples in redshift and construct volume limited samples of faint galaxies over a substantial redshift range.

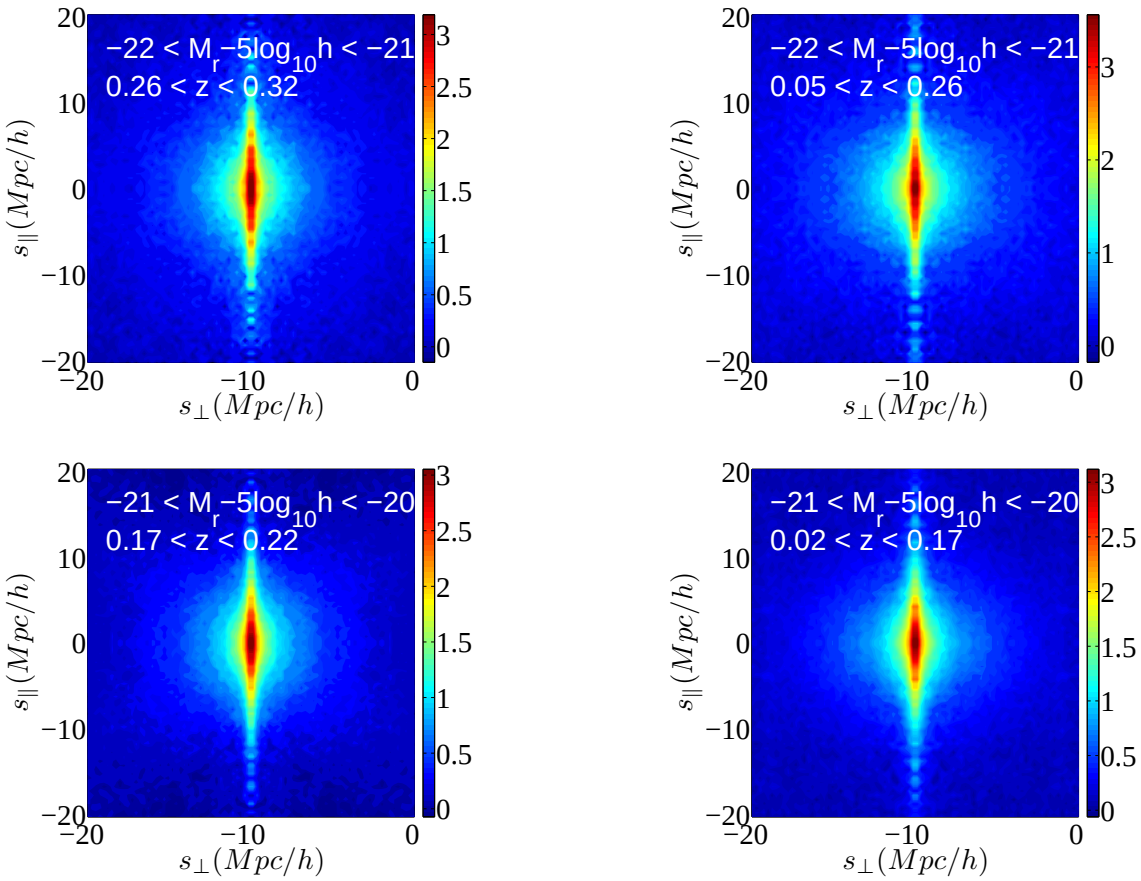

Figure 6.2: The redshift space correlation function $\xi(s_\perp, s_\parallel)$ for galaxy samples with $L \gtrsim L^*$. Due to the dynamical range of $\xi(s_\perp, s_\parallel)$ the colour coding of the plots show the quantity $\ln[1 + \xi(s_\perp, s_\parallel)]$, which we also use for the parameter fitting. GAMA's high completeness provides a very high signal to noise measurement of the anisotropic correlation function on small scales where the finger of god effect dominates. Whilst distinctively present, coherent infall on large scales has smaller signal to noise making it difficult to use these samples for robust measurements of the growth rate of structure.

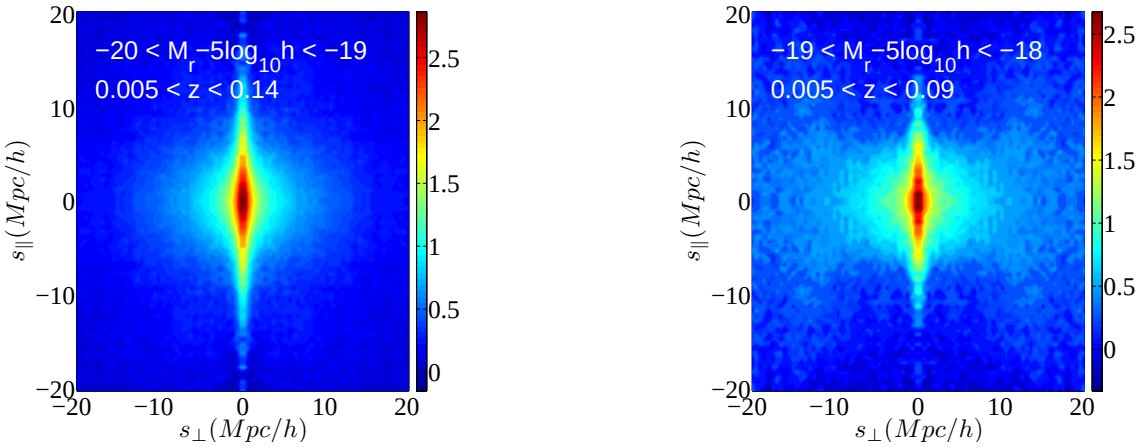

Figure 6.3: The redshift space correlation function $\xi(s_\perp, s_\parallel)$ for low redshift galaxy samples with $L < L^*$. As in Fig. 6.2 the quantity plotted is $\ln[1 + \xi(s_\perp, s_\parallel)]$. The FOG effect is present with a high signal to noise, however coherent large scale infall appears more noisy, especially for the faintest sample (right panel).

present in all samples. Moreover, there are clear differences between the shape of $\xi(s_\perp, s_\parallel)$ on large scales for the two luminosity bins, but there almost no visible differences between same luminosity bins in different redshift ranges. The shape of $\xi(s_\perp, s_\parallel)$ is affected by the bias and therefore is expected to differ between galaxy samples with different bias.

Fig. 6.3 depicts $\ln[1 + \xi(s_\perp, s_\parallel)]$ for our two $L < L^*$ samples. These samples have fewer galaxies and are not split in redshift. In contrast with the $L \gtrsim L^*$ samples, the signal of $\xi(s_\perp, s_\parallel)$ appears to be more noisy on large scales. This is due to fact that the vast majority of these galaxies are satellites and therefore poor tracers of the large scale matter distribution. However, once again the large-scale squashing along the LOS (albeit noisy) and the FOG are present.

For the uncertainty estimation we repeat the analysis for the 9 mock samples described in Section 3.3. Despite the fact that there exist deviations between the clustering observed in GAMA and in the mocks, the overall characteristics described above are also evident in the mocks. We stress that the relative uncertainty (i.e. the one obtained from the mean of the mocks) is sufficient for the parameter fitting of the correlation function, with the caveat of the neglecting of covariances. We address this deficiency in Section 6.4.2, where we test the recovery of the parameters of the model using the mocks. Strictly speaking, one should use the fluctuations in the number counts and then estimate the uncertainty on $\xi(s_\perp, s_\parallel)$, however we found no difference in the results if we simply use scatter of $\xi(s_\perp, s_\parallel)$ directly.

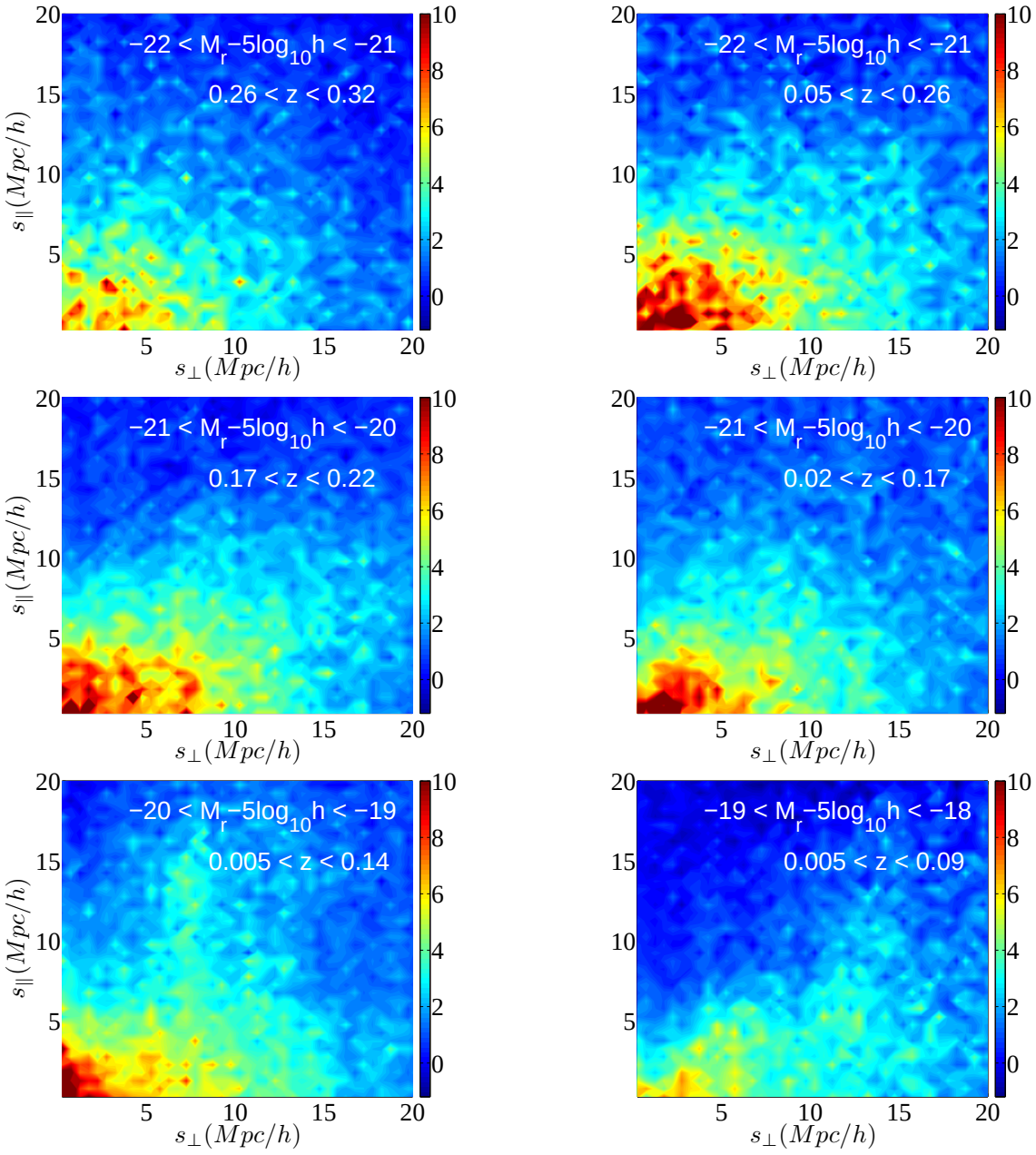

Figure 6.4: Signal-to-noise ratio for all our luminosity bin volume limited samples. At each $(s_\perp, s_\parallel)$ pixel the colour coding shows the value of $\xi(s_\perp, s_\parallel)$ divided by its corresponding uncertainty obtained from the mocks. Dark red pixels show data points with S/N ≥ 10.

Fig. 6.4 shows the signal-to-noise ratio (S/N) of the $\xi(s_\perp, s_\parallel)$, where for the noise estimate we use the standard deviation, obtained from the 9 GAMA mocks, for each $(s_\perp, s_\parallel)$ bin, i.e. the diagonal elements of the covariance matrix. These plots have the advantage of presenting the clustering information with its respective uncertainty. Fig. 6.4 confirms that the $L < L^*$ samples have significantly lower signal-to-noise ratio, especially on large scales. Moreover, the low-z $-22 < M_r - 5\log h < -21$ sample has higher S/N than its high-z counterpart. For all samples in Fig. 6.4, the S/N value is less than 1 on scales $\sim 20\ h^{-1}$Mpc. This is the largest scale for which we can use our $\xi(s_\perp, s_\parallel)$ measurements for parameter estimation. We further test this conjecture in Section 6.4.2.

6.3.3 Anisotropic correlation function for luminosity thresholds

We repeat the analysis with volume limited samples defined with luminosity thresholds. Fig. 6.5 depicts the S/N ratio of anisotropic correlation function measurements for six samples with the same redshift limits as before and with absolute magnitude limits $M_r - 5\log h < -21, -20, -19, -18$, where the two brightest ones are also split in redshift. Due to the fact that they include more galaxies than the luminosity bins samples these measurements are less noisy. This is shown in Fig. 6.5 where for the respective scales of Fig. 6.4, the S/N is systematically higher.

The luminosity threshold samples are complimentary to the luminosity bin samples presented above. Luminosity threshold samples have a more transparent interpretation in the context of halo modeling. For similar low-z samples a complete clustering analysis has been performed recently by Zehavi et al. (2011), which we use for the interpretation of our results. In order to do this we perform a consistency test. We calculate the projected correlation function (equation 6.17) for our low-z $M_r - 5\log h < -20$ sample and check that it is in agreement with the tabulated measurements of the same galaxy sample provided by Zehavi et al. (2011). This is an important consistency test, despite the 1.8 mag difference in the flux cut (which extends our volume limited $M_r - 5\log h < -20$ sample to a higher redshift limit), the significantly smaller number of galaxies in GAMA and the different cosmological and $k + e$-correction parameters we adopted in this work.

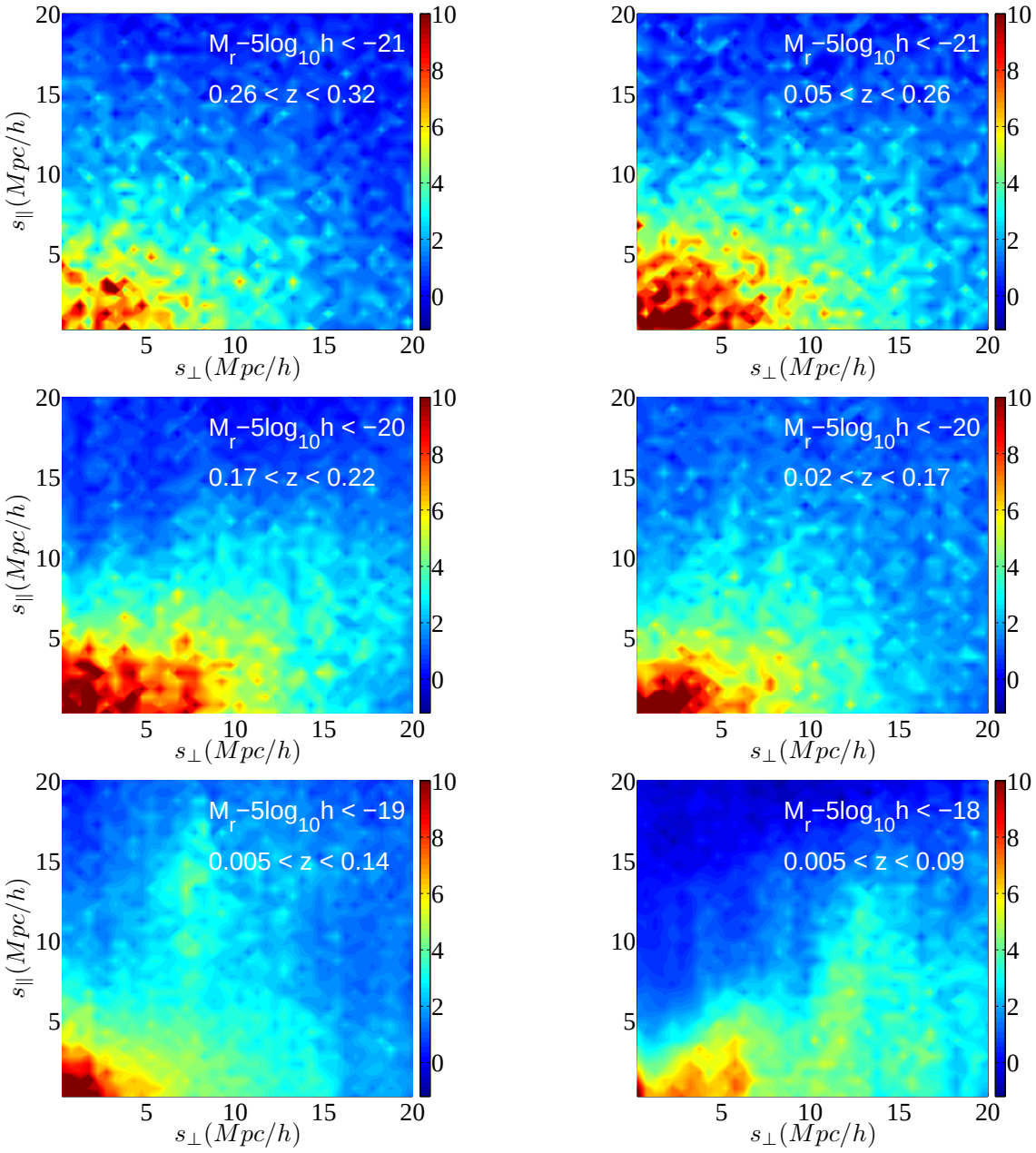

Figure 6.5: Signal-to-noise ratio for all our luminosity threshold volume limited samples. At each $(s_\perp, s_\parallel)$ pixel the colour coding shows the value of $\xi(s_\perp, s_\parallel)$ divided by its corresponding uncertainty obtained from the mocks. Dark red pixels show data points with S/N $\geq$ 10.

6.4 Analysis

6.4.1 Underlying $\xi(r)$ and scale-dependent bias

Although $\xi(s_\parallel, s_\perp)$ is the most direct measurement from the data, in order to use it in the simple model of equation 6.16 we need a model for the underlying $\xi(r)$, i.e. the quantity that is seemingly distorted and loses its isotropy in redshift space. The most direct solution is to use the non-parametric inverted $\xi(r)$ following the interpolation method of Saunders et al. (1992). The advantage of the direct reconstruction of $\xi(r)$ from the data is that all uncertainties on small scales from the modeling of the real space correlation function (scale-dependent bias) are minimized. However, direct reconstruction of $\xi(r)$ has its own implicit assumptions (Norberg et al., 2009) and is only valid on limited scales even when one is using relatively large galaxy samples (Cabré and Gaztañaga, 2009a,b). Furthermore, the reconstructed $\xi(r)$ from eqs 6.17 and 6.18 is an experimental quantity and it is difficult to accurately calculate the numerical integrals by interpolating noisy data. Thus using only the reconstructed $\xi(r)$ is insufficient because we also need a model of $\xi(r)$ on large scales as mentioned in Section 2.

On the other hand, linear theory, assisted by simulations, provides another model (Smith et al., 2003) for the real space two point correlation function, valid on scales $r \gtrsim 5\ h^{-1}$Mpc, that has been shown to stand the comparisons with data to a certain extend. A number of studies used this model, with the addition of a linear bias term, either in Fourier space (equation 6.7, Blake et al., 2011) or in configuration space (equation 6.16, Cabré and Gaztañaga, 2009a) and found it an acceptable fit for the data.

Since galaxies are believed to follow the theoretical power spectrum on large scales up to some multiplication factor, we use the information from the *data*-inferred $\xi(r)$ to get a non-parametric model of the scale-dependent bias which we then use to modify the shape of the theoretical $\xi(r)$ on small scales. This approach avoids any unnecessary assumptions about the form of the bias, apart from the scale for which bias becomes scale-independent. In Fig. 6.6 we show the reconstruction of the smoothed scale-dependent bias for all galaxy samples. Due to the fact that we cannot reliably estimate the form of the bias on separations $r > 5\ h^{-1}$Mpc, we assume that at these scales the bias is constant.

In practice, the reconstruction of $b_{NL}(r)$ for each sample has the following schema.

- We alculate the real space $\xi_{gg}(r)$ from the observed $\xi(s_\perp, s_\parallel)$ using equations 6.17 and 6.18.

- We get an estimate of $P_{\delta\delta}(k)$ from `CAMB` (Lewis et al., 2000) and Fourier transform

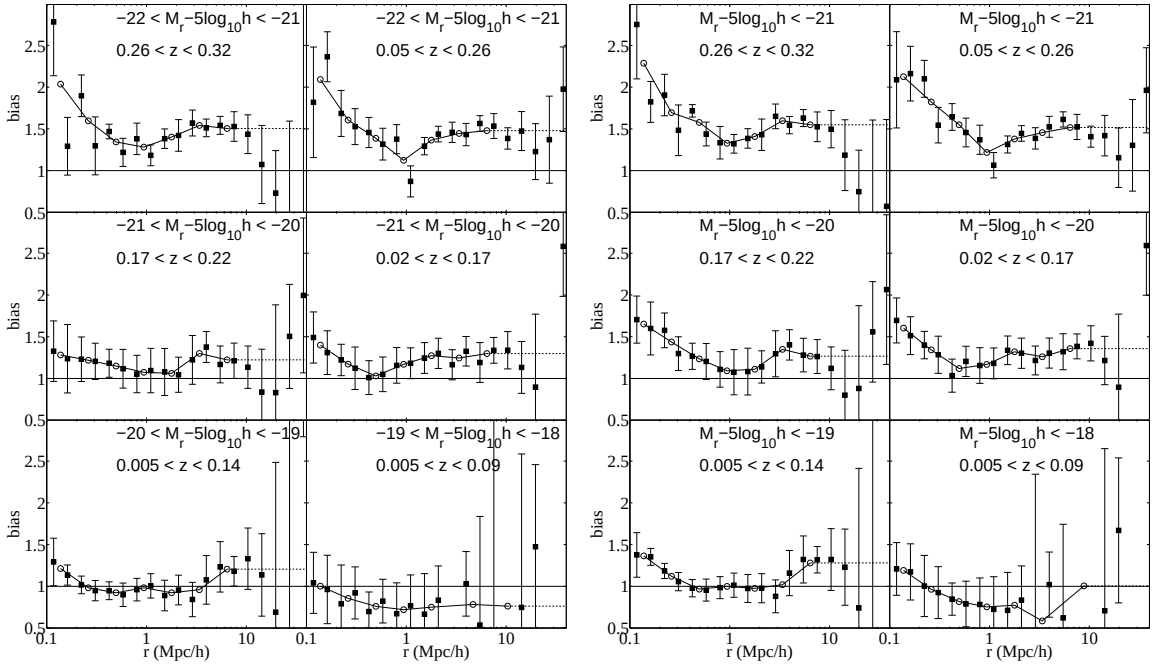

Figure 6.6: Non-parametric estimates of the galaxy bias as function of scale (see equation 6.21). Filled squares show the square root of the ration of the observed $\xi_{gg}(r)$ over the theoretical $\xi_{\delta\delta}(r)$. Open circles show the smooth estimate of $b_{NL}(r)$. Notably the bias for separations $5 \lesssim r \lesssim 10\ h^{-1}$Mpc is approximately constant.

it to $\xi_{\delta\delta}(r)$ using the code `FFTLog` (Hamilton, 2000).

- We calculate the average bias given by

$$b(r) = \sqrt{\xi_{gg}(r)/\xi_{\delta\delta}(r)}, \tag{6.21}$$

for approximately every two $\xi_{gg}(r)$ data points, in order to get a smooth estimate of $b(r)$, over separations $0.5 \lesssim r \lesssim 5\ h^{-1}$Mpc.

- We get an estimate of the linear bias b_N, which we assume to be the average value of bias $b(r)$, over separations $5 \lesssim r \lesssim 10\ h^{-1}$Mpc.

- We divide $b(r)$ by our linear bias estimation in order to get b_{NL}, which gives the bias value as a function of r, scaled with $b_L(r)$.

As the linear bias becomes a free parameter for the samplings in Section 6.4.2 it is not necessary to know its exact value. However, the choice of scales for the onset of the linear bias is very important. Here we chose that scale to be at roughly $5\ h^{-1}$Mpc merely because up to these scales the $b_{NL}(r)$ can be estimated robustly from our data. Still, from Fig. 6.6 we note that there is plateau on scales $5 \lesssim r \lesssim 10\ h^{-1}$Mpc, which is expected as the

halo bias becomes constant at these scales. At scales $r > 10\ h^{-1}$Mpc the inversion of the correlation function becomes noisy and these data points are not reliable anymore for bias estimations.

Equipped with the observed $b(r)$ we modify the Smith et al. (2003) correlation function accordingly for each sample. The crucial advantage of this procedure is that the motivation behind the final form of $\xi(r)$ is not as important as matching it accurately on all scales. The reason for this is that the rsd observables are degenerate with the bias (and σ_8) and therefore any inaccuracies on the bias modeling (e.g. using halo bias) will cause biases in our results.

6.4.2 Optimal parameter space for σ_{12} and its likelihood surface

Equation 6.16 gives our final model to fit the $\xi(s_\perp, s_\parallel)$ data and obtain measurements of the pairwise velocity dispersion. We chose to fit on the entire $\xi(s_\perp, s_\parallel)$ plane because the pairwise velocity disperion is highly a non-linear effect and thus cannot be captured only by the first few multipole expansions of the correlation function.

We sample the parameter space of equation 6.16 which comprises the set

$$\mathbf{p}_1 \equiv (f_g, \sigma_8, b, \sigma_{12}). \tag{6.22}$$

For this analysis we assume that the cosmological parameters and the scale-dependent bias of equation 6.21 are known with perfect accuracy. Under these assumptions the growth rate is technically known for a ΛCDM Universe (Linder, 2005), but here we allowed it to vary, so that we consistently fit all dynamical quantities simultaneously. We also assume no large scales velocity bias so that large scale coherent galaxy infalls follow exactly the respective dark matter one. Other studies have used similar approaches. Guzzo et al. (2008) used a limited two-parameter set comprised only by ($\beta \equiv f_g/b, \sigma_{12}$) and they further assumed that the posterior is a bivariate Gaussian, with widths given from the scatter of the parameters from simulations. One can also obtain fits on the linear bias parameter as this is an independent overall multiplication factor (Cabré and Gaztañaga, 2009a). However, it was pointed by Percival and White (2009) that both f_g and b are exactly degenerate with σ_8 and therefore in the linear regime one can constrain only combinations of this three-parameter set. Our sampling tests, shown in Fig. 6.7 for the low-z L^* sample confirmed this. Fig. 6.7 shows that the probability distributions for the joint fits of $f_g, \sigma_8, b, \sigma_{12}$ are banana shaped and degenerate. On the other hand the distribution for the combinations of $f_g\sigma_8$ and $b\sigma_8$ and σ_{12} are Gaussian-like and well constrained. We tried the same test of

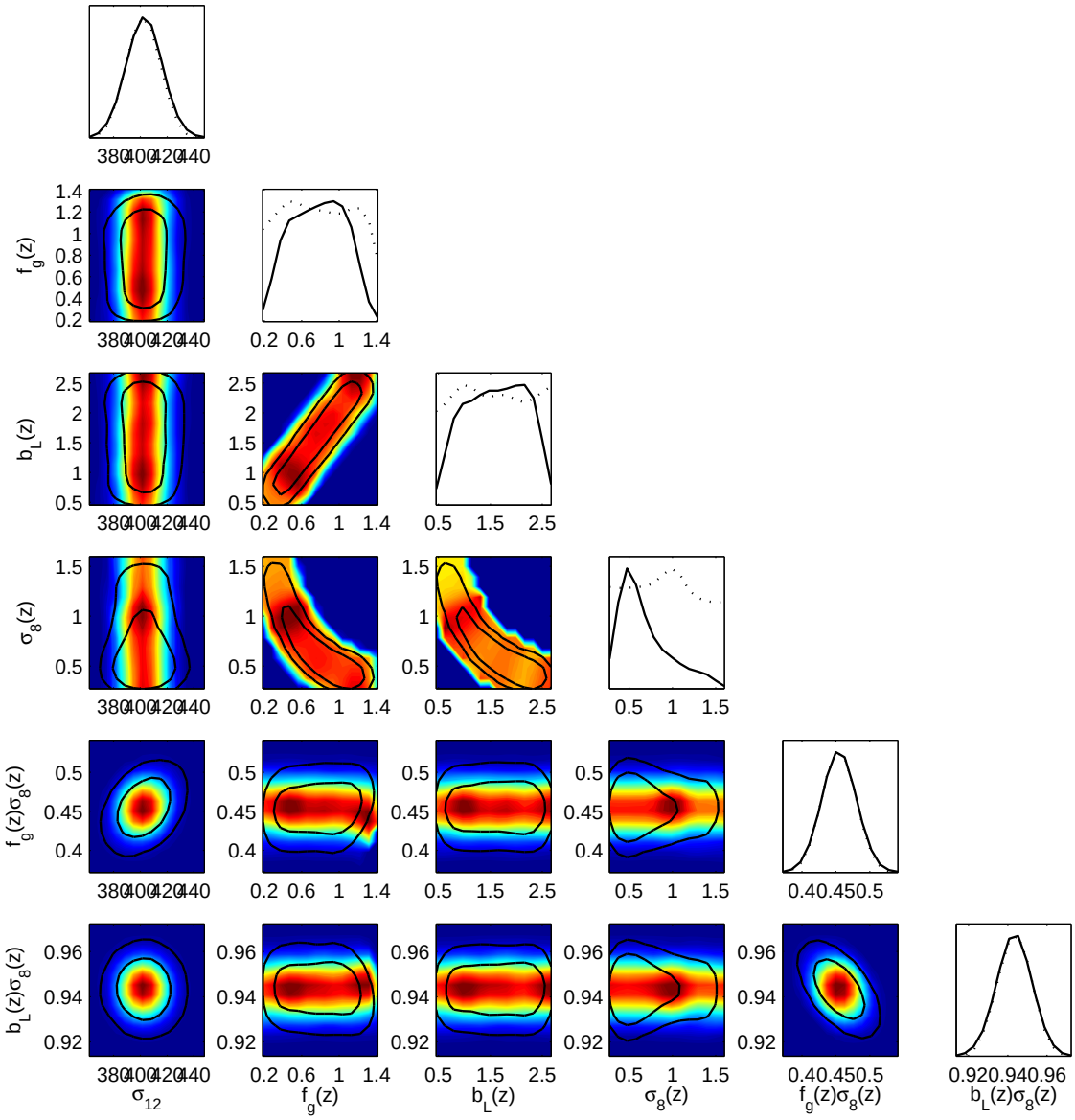

Figure 6.7: The parameter space used for studying the pairwise velocity dispersion. The figure shows the constraints on the parameters of the "Kaiser plus Lorentzian" model for redshift space distortions (equation 6.22) for the low-z, $-21 < M_r - 5\log_{10} < -20$ sample (our L^* sample). Colour contours and dotted lines show the likelihood from the sampling and solid lines the marginalized posterior distributions. Due to the fact that the linear-regime parameters (f_g, b, σ_8) are degenerate their joint constraints have a banana shape. On the other hand the parameter combinations $f_g\sigma_8$ and $b\sigma_8$ are well-described by the data with the caveat that the width of their respective posterior distributions is much narrower that the one would expect from a survey like GAMA (See text for further discussion). The posterior distribution of the parameter set $(f_g\sigma_8, b\sigma_8, \sigma_{12})$ can be approximated by a trivariate gaussian.

all our samples and the results are identical. Consequently we use the parameter set

$$\mathbf{p}_2 \equiv (f_g\sigma_8, b\sigma_8, \sigma_{12}) \tag{6.23}$$

for our inferences. Since we are primarily interested in the pairwise velocity dispersion we don't have to break the degeneracy using external datasets, instead we marginalize over the combinations $f_g\sigma_8$ and $b\sigma_8$ instead. A similar parameter set was used by Tojeiro et al. (2012) who then broke the $\sigma_8 - f_g$ degeneracy using the linear prediction for the evolution of the linear, scale-indepent bias.

We perform our samplings using the publicly available package `Multinest` (Feroz and Hobson, 2008; Feroz et al., 2009)[2]. `Multinest` produces MCMC-type outputs (albeit being a nested sampling algorithm) which can be analyzed using the `GetDist` routine of the `CosmoMC` package Lewis and Bridle (2002) [3]. Our $\xi(s_\perp, s_\parallel)$ uncertainties come from the scatter of the 9 GAMA mocks. When we run `Multinest`, we minimize the quantity (Hawkins et al., 2003; Guzzo et al., 2008)

$$-2\ln\mathcal{L} = \chi^2 = \sum_i \sum_j \frac{\ln(1+\xi_{ij}) - \ln(1+\bar{\xi}_{ij})}{\sigma_{ij}^2}, \tag{6.24}$$

where $\sigma_{ij} = \sigma_{\xi_{ij}}/(1+\bar{\xi}_{ij})$. Only pixels which satisfy the criterion $\sqrt{s_i^2 + s_j^2} < 20\ h^{-1}$Mpc are included in the fitting; we fit on a quarter of a disk of radius s on the $\xi(s_\perp, s_\parallel)$ plane (see Figs 6.2, 6.3, 6.4 and 6.5). We discuss the choice of 20 h^{-1}Mpc as our upper limit for the fits below (see also Fig. 6.8). One issue of utmost importance to address for our parameter fitting is the covariance matrix of $\xi(s_\perp, s_\parallel)$ (Norberg et al., 2009). It's now established that uncertainty covariance matrices are important for clustering measurements since correlation function measurements are strongly correlated (see e.g. Christodoulou et al., 2012) and recent studies have raised the issue again for rsd measurements (Bianchi et al., 2012). However, the small number of available mocks doesn't allow the construction of stable covariance matrices for our samples due to the number of data points that we have (the fits are done using $\sim 40 \times 40$ bins). We can test the assumption of zero or very small covariances using the mocks. If the uncertainties are not underestimated then the scatter between the measurements from the different mocks should agree with the width of the posterior distributions from the nested samplings. We used the maximum likelihood measurements of $f_g\sigma_8$ for this exercise, as it's an easier parameter to interpret and we know its exact value in the simulations a priori. We found that the parameter uncertainties are severely underestimated. In Fig. 6.8 the different $f_g\sigma_8$ fits from the 9 GAMA mocks

[2] `http://ccpforge.cse.rl.ac.uk/gf/project/multinest/`

[3] `http://cosmologist.info/cosmomc/`

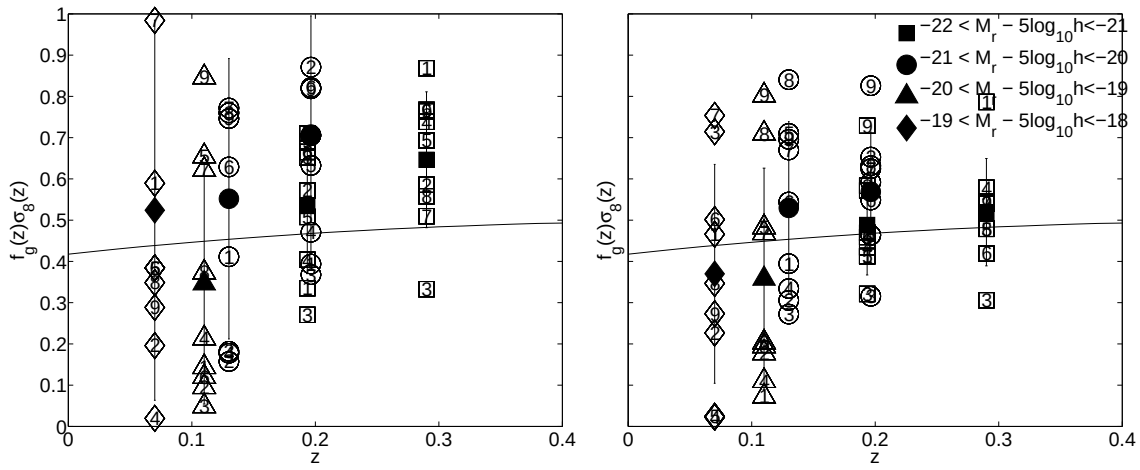

Figure 6.8: Test for the recovery of the growth rate parameter f_g from the 9 GAMA mocks, using the "Kaiser plus Lorentzian" model for redshift space distortions (equation 6.16). Open data points show the 6 volume limited samples (Fig. 6.1) for each of the 9 mocks described in Section 3.3, whereas filled data points show their mean and standard deviation. Solid line shows the fiducial ΛCDM + GR prediction, $f_g = \Omega_m^{0.55}$, multiplied by $\sigma_8 = 0.9$, the corresponding value used for the mocks. Right panel shows the results using the scales $0 < s < 20\ h^{-1}$Mpc for the fit; Left panel the results using the scales $0 < s < 40\ h^{-1}$Mpc for the fit. Limiting our fits to scales $s < 20\ h^{-1}$Mpc improves the estimates of f_g, as well as the distribution of the results from the different mocks around their mean. However in both panels the scatter from the mocks is much wider than what the marginalized distribution of $f_g\sigma_8$ suggests in Fig. 6.7.

are shown along with their mean and standard deviation fit over two different scales: $0 < s < 40\ h^{-1}$Mpc (left panel) and $0 < s < 20\ h^{-1}$Mpc (right panel). The comparison between the width of the posterior distribution of Fig. 6.7 and the spread of the fits around their mean, clearly shows that uncertainties are underestimated.

In order to obtain more realistic uncertainties we use the covariances of the parameters. Using the 9 different estimations of the parameter set $\mathbf{p}_2$ for each mock galaxy volume limited sample, we construct their respective covariance matrix and then approximate their likelihood surface with a trivariate Gaussian, whose widths come from the scatter of the parameter fits of our mocks. Our investigation of the likelihood surface of $\mathbf{p}_2$ (Fig. 6.7) showed that in general a trivariate Gaussian is a reasonable approximation. We then obtain the uncertainties of each parameter in the set $\mathbf{p}_2$ after we marginalize over the remaining two parameters and their correlations (Sivia and Skilling, 2006).

Fig. 6.8 also tests the the scales over which the model $\xi(s_\perp, s_\parallel)$ is fit. The left panel of

Fig. 6.8 shows the fits over the scales $0 < s < 40\ h^{-1}$Mpc and the right panel the fits over scales $0 < s < 20\ h^{-1}$Mpc. Choosing $s_{\max} < 20\ h^{-1}$Mpc improves the *distribution* of the different $f_g\sigma_8$ measurements which are spread more uniformly around their mean, whereas the means themselves are less biased with respect to the fiducial value $f_g(z) = \Omega_m^{0.55}$ (Peebles, 1980). Thus, including only the most appropriate scales of $\xi(s_\perp, s_\parallel)$ improves the Gaussianity of the distribution of the fitted parameters as well as the parameter fits. This corroborates with our approximation of the posterior distribution with a trivariate Gaussian. We conclude that our $\xi(s_\perp, s_\parallel)$ measurements are noise dominated on scales $s > 20\ h^{-1}$Mpc. Fig. 6.8 also shows that as we go to fainter samples the $f_g\sigma_8$ estimations cluster in a biased fashion from their respective mean. This is clearly obvious from the low-z $-21 < M_r - 5\log h < -20$ sample, as well as the $-20 < M_r - 5\log h < -19$ sample. This is in accord with studies using dark matter simulations with HOD galaxy modeling which also showed that the "Kaiser plus Lorentzian" model is not accurate for $L < L^*$ samples (de la Torre and Guzzo, 2012). On the other hand scale-dependent bias, doesn't affect measurements of $f_g\sigma_8$ significantly, but since studies using simulations showed that it does affect measurements of σ_{12} (Cabré and Gaztañaga, 2009b; de la Torre and Guzzo, 2012) we include it in the calculations.

Finally we note that more sophistigated models for $\xi(s_\perp, s_\parallel)$, which could be used to fit to our data (e.g. the models of Scoccimarro, 2004; Taruya et al., 2010; Jennings et al., 2011), are aiming to model much larger pair separations than the ones we consider here. These studies have as their primary aim is to overcome biases on f_g measurements, whereas in this Chapter we are mostly interested in the non-linear regime and the pvd. Yet, in the recent comprehensive test of a manifold of rsd models by the WiggleZ team (Blake et al., 2011), the model of equation 6.15 was found to perform almost as well as the more sophisticated fitting functions from simulations. Clearly the available data at the moment is not good enough to distinguish between different estimators of f_g. For the analysis presented here as long as the bias on our f_g estimations is below our statistical uncertainty then this should not significantly affect our measurements of the pairwise velocity dispersion since we marginalize over it.

6.4.3 Results for the "Kaiser plus Lorentzian" model

We now sample the parameter space $\mathbf{p}_1$ (equation 6.22) for all our galaxy samples. Following the discussion in Section 6.4.2 we approximate the posterior distribution of the parameter set $\mathbf{p}_2$ (equation 6.23) with a trivariate Gaussian and we quote the maximum

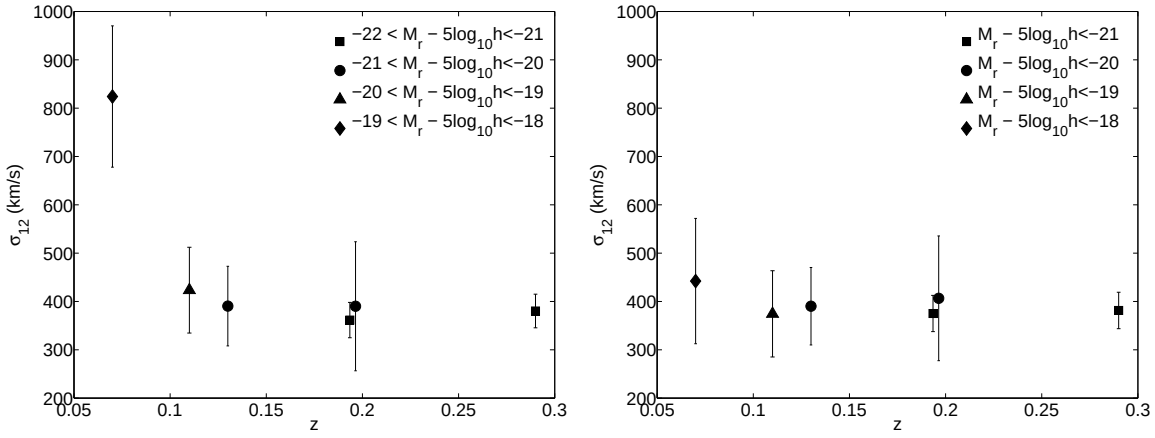

Figure 6.9: Pairwise velocity dispersion as function of redshift and luminosity (symbols) for volume limited magnitude bins (left panel) and volume limited magnitude threshold (right panel) samples. Error bars are 1σ derived from marginalizing a trivariate Gaussian distribution of the parameter set $\mathbf{p}_2$ (equation 6.23). The pvd is roughly constant ($\approx$ 400 kms^{-1}) for all samples apart from the faintest luminosity bin sample.

likelihood parameter values. The uncertainties are 1σ and they come from the marginalized Gaussian distributions obtained from the mocks.

The left panel of Fig. 6.9 shows the pairwise velocity dispersion as a function of redshift and absolute magnitude in four absolute magnitude bins. The pvd is roughly constant for the first three magnitude bins and is found to be $\sigma_{12} \approx 400$ kms^{-1}. However it increases steeply for the faintest one reaching an almost double value $\sigma_{12} = 824 \pm 146$ kms^{-1}. The rapid increase of σ_{12} for faint galaxies shows the strong effect of satellite galaxies on pvd. In general virial motions inside a halo dominate halo-halo motions (Cooray and Sheth, 2002) and since the vast majority of galaxies with magnitude $-19 < M_r - 5\log h < -18$ are expected to be satellites in large haloes (Zehavi et al., 2011) their pvd is much greater than brighter galaxies. Contrary to the absolute magnitude bin samples, the absolute magnitude threshold samples have approximately the same pvd for all magnitude thresholds. This is something to be expected since even for the faintest absolute magnitude range ($M_r - 5\log h < -18$) only $\sim 32\%$ of the galaxies are satellites (Zehavi et al., 2011).

In Fig. 6.9 we also test possible evolution of pvd of the two brightest samples, which are split in redshift. Neither for the absolute magnitude bins, nor for the absolute magnitude thresholds, we observe any hints of redshift evolution of the pvd. All our σ_{12} measurements are given in Tables 6.1 and 6.2.

Although our samples are not designed for cosmology, we can test the consistency of our growth rate values with the fiducial ΛCDM prediction (Linder, 2005). In Fig. 6.10

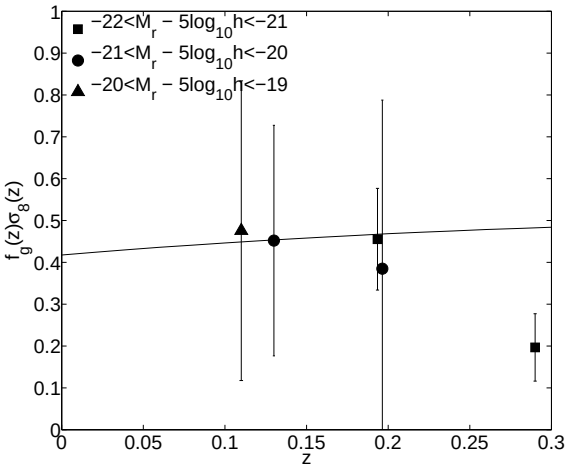

Figure 6.10: Growth rate of structure as a function of redshift. Data points show the measurements from the five brightest volume limited samples corresponding to the luminosity thresholds shown in the legend. Solid line shows the ΛCDM + GR prediction, $\Omega_m^{0.55}(z)\sigma_8(z)$, where $\Omega_m(z=0)=0.25$ and $\sigma_8(z=0)=0.8$.

Table 6.1: Redshift space distortion measurements of the dynamical parameters $\sigma_{12}, f_g\sigma_8, b\sigma_8$ for volume limited samples defined in absolute magnitude bins. See Section 6.4.2 for a description of the inferred parameters. Uncertainties are assumed to be Gaussian 1σ and are obtained from a marginalized trivariate Gaussian. M_r should be understood as $M_r - 5\log h$.

Redshift range	$z_{\rm median}$	Magnitude limit	N_g	$f_g\sigma_8$	$b\sigma_8$	σ_{12} (km s^{-1})
$0.26 < z < 0.32$	0.29	$-22 < M_r < -21$	10582	0.20 ± 0.08	1.05 ± 0.02	380 ± 34
$0.05 < z < 0.26$	0.19	$-22 < M_r < -21$	10323	0.46 ± 0.12	1.11 ± 0.04	361 ± 36
$0.17 < z < 0.22$	0.20	$-21 < M_r < -20$	11006	0.38 ± 0.40	0.95 ± 0.11	390 ± 133
$0.02 < z < 0.17$	0.13	$-21 < M_r < -20$	10525	0.45 ± 0.28	0.94 ± 0.09	390 ± 82
$0.005 < z < 0.14$	0.11	$-20 < M_r < -19$	9594	0.48 ± 0.36	1.00 ± 0.08	423 ± 88
$0.005 < z < 0.09$	0.07	$-19 < M_r < -18$	3202	1.69 ± 0.71	0.57 ± 0.12	824 ± 146

Table 6.2: Redshift space distortion measurements of the dynamical parameters $\sigma_{12}, f_g\sigma_8, b\sigma_8$ for volume limited samples defined in absolute magnitude thresholds. See Section 6.4.2 for a description of the inferred parameters. Uncertainties are assumed to be Gaussian 1σ and are obtained from a marginalized trivariate Gaussian. M_r should be understood as $M_r - 5\log h$.

Redshift range	$z_{\rm median}$	Magnitude limit	N_g	$f_g\sigma_8$	$b\sigma_8$	σ_{12} (km s^{-1})
$0.26 < z < 0.32$	0.29	$M_r < -21$	11642	0.20 ± 0.11	1.09 ± 0.02	381 ± 37
$0.05 < z < 0.26$	0.19	$M_r < -21$	11118	0.52 ± 0.12	1.11 ± 0.04	375 ± 37
$0.17 < z < 0.22$	0.20	$M_r < -20$	14880	0.42 ± 0.41	0.98 ± 0.11	406 ± 129
$0.02 < z < 0.17$	0.13	$M_r < -20$	13873	0.42 ± 0.27	1.00 ± 0.09	390 ± 80
$0.005 < z < 0.14$	0.11	$M_r < -19$	17447	0.42 ± 0.41	1.05 ± 0.09	374 ± 89
$0.005 < z < 0.09$	0.07	$M_r < -18$	7547	1.38 ± 0.68	0.76 ± 0.11	442 ± 129

we plot our $f_g\sigma_8$ measurements for the 5 brightest luminosity threshold samples, with the standard ΛCDM + GR, $\Omega_m^{0.55}\sigma_8(z)$ prediction. Due to the size of GAMA and the way our samples were constructed, the error bars in Fig. 6.10 are big and from that figure we can only conclude that our data is generally consistent with ΛCDM + GR. Recent studies using much larger samples of galaxies showed that ΛCDM + GR is a good description of the data (Reid et al., 2012; Samushia et al., 2012).

6.4.4 Scale dependent pairwise velocity dispersion

We extend the analysis by splitting the $\xi(s_\perp, s_\parallel)$ plane along the transverse direction in order to study the scale dependence of the pairwise velocity dispersion. The methodology is the same as the one we followed in Section 6.4.3, with the exception of the scales on which we perform the fits. Observationally the scale dependence of the pvd can be expressed as a function of the transverse separation, which is unaffected by peculiar velocities. We therefore perform the parameter fits in stripes of constant $s_\perp$ intervals. We also impose the upper limit $s_\parallel < 20\ h^{-1}$Mpc.

As expected from the smaller area of the $\xi(s_\perp, s_\parallel)$ plane that we use for these fits, the uncertainties on σ_{12} are much larger for all samples. For this exercise, we have not investigated the likelihood surface. Furthermore, contrary to the methodology of Section 6.4.2, we choose not to approximate the likelihood surface of each parameter set with a

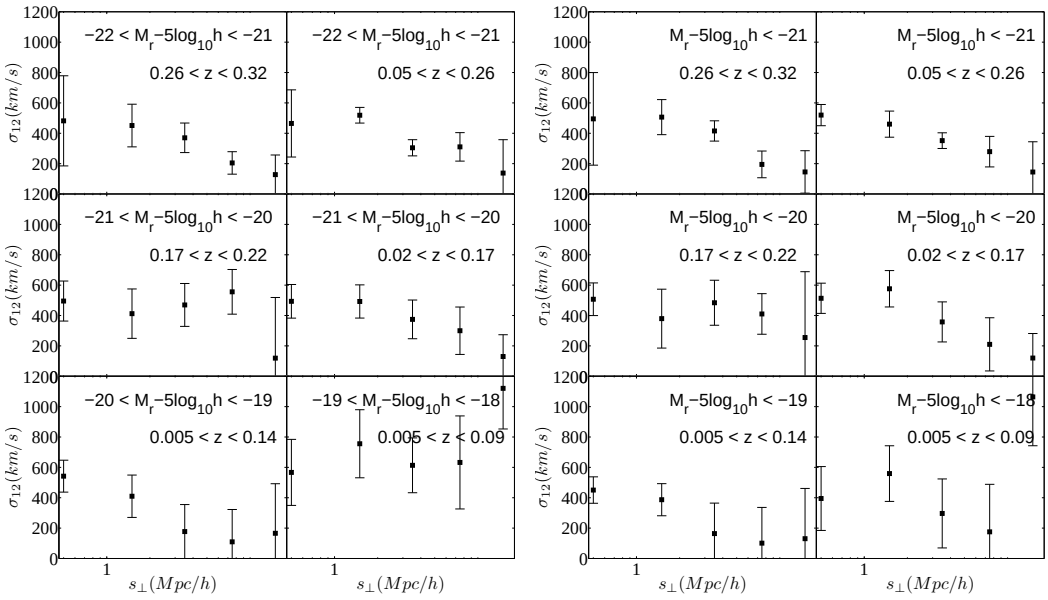

Figure 6.11: Pairwise velocity dispersion as a function of the projected separation $s_\perp$ in absolute magnitude bins (left panel) and absolute magnitude thresholds (right panel). The model given by equation 6.16 has been used for these measurements and error bars come from the scatter of the parameter fits from the 9 mocks.

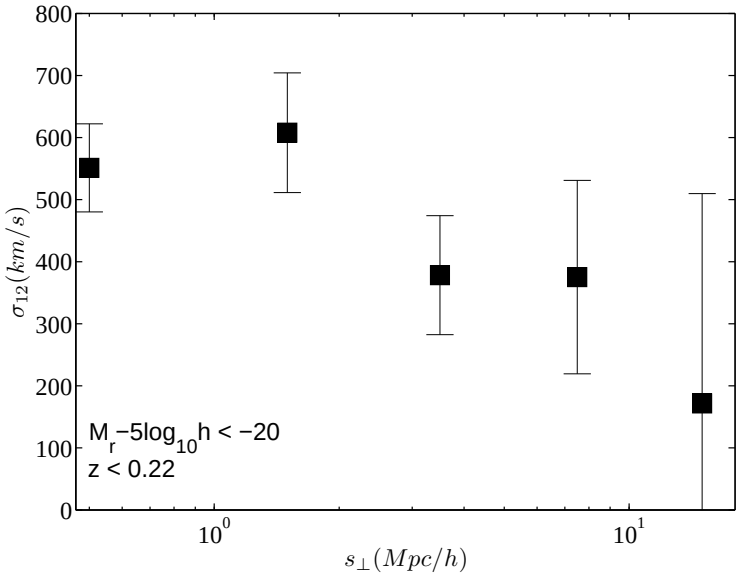

Figure 6.12: Pairwise velocity dispersion as a function of the projected separation $s_\perp$ for a larger volume limited sample defined as $z < 0.22$ and $M_r - 5\log h < -20$. The model given by equation 6.16 has been used for these measurements and error bars come from the scatter of the parameter fits from the 9 mocks.

Gaussian. Thus we report only the uncertainties from the scatter of parameter fits of the mocks, without marginalization. We note that this is most likely an underestimation of the true uncertainties. For the particular case for transverse separations $s_\perp \approx 1.5\ h^{-1}$Mpc, where S/N is larger, we report our measurements in Tables 6.3 and 6.4.

The pvd as a function of separation for volume limited luminosity bins and luminosity thresholds is shown in Fig 6.11. The pvd presents a scale dependence and it decreases with increasing $s_\perp$ for all samples. For the faintest samples $-19 < M_r - 5\log h < -18$ and $M_r - 5\log h < -18$ the pvd shows a rapid increase at separation $s_\perp > 10\ h^{-1}$Mpc, however it is very likely that these values are biased due to the lack of sufficient $\xi(s_\perp, s_\parallel)$ signal at these scales (see Fig. 6.4 and Fig. 6.5). Note that the model of equation 6.16 assumes a constant coherent infall component of peculiar velocities, which is an assumption that might not hold for the samples that we analyze here, due to their size. On scales $s_\perp < 0.5\ h^{-1}$Mpc our $\xi(s_\perp, s_\parallel)$ measurements are limited by shot noise and therefore at these scales we cannot provide reliable measurements of the pvd.

Theoretical predictions for the scale-dependence of the pvd suggest that it should exhibit a maximum at separations of $s \approx 1\ h^{-1}$Mpc, which defines the regime of the largest collapsed structures (Sheth et al., 2001). Although we do observe hints of this effect (e.g. the $M_r - 5\log h < -20$ sample in the right panel of Fig. 6.11), the error bars are still quite large. One more complication arises due the fact that the pvd is model-dependent with respect to the particular form of the damping function that one uses (Zehavi et al., 2002), presumably a result of the weak $\xi(s_\perp, s_\parallel)$ signal at large transverse separations. Moreover, at very large scales the measured pairwise velocity dispersion should match the respective quantity from linear theory

$$\sigma_v^2 = \frac{1}{3} \int \frac{1}{2\pi^2} P_{\theta\theta}(k) dk \tag{6.25}$$

where σ_v is the one-dimensional linear velocity dispersion. As usual, dividing equation 6.25 with $H^2(z)(1+z)^2$ will give σ_v in km/s and the damping function is now a Gaussian (Scoccimarro, 2004). Percival and White (2009) using simulations found that a Gaussian function is a better fit at large scales, although the study they performed was primarily aiming to get the best estimates of the growth rate.

In order to see more clearly how these theoretical predictions of the scale-depencence of the pvd can be tested, we define a much larger volume limited sample with the following redshift and absolute magnitude limits $z < 0.22$ and $M_r - 5\log h < -20$ respectively. This would allow for reduced error bars on $\sigma_{12}(s_\perp)$. We repeat the same procedure described in Section 6.4.3 in order to measure the pvd from the anisotropic correlation function of

Table 6.3: Redshift space distortion measurements at transverse separations $\sigma_{12}(s_\perp = 1.5\ h^{-1}\mathrm{Mpc})$ for volume limited samples defined in absolute magnitude bins. Uncertainties come from the scatter of the fitted parameters from the 9 mocks.

Redshift range	z_{median}	Magnitude limit	N_g	$f_g\sigma_8$	$b\sigma_8$	σ_{12} (km s^{-1})
$0.26 < z < 0.32$	0.29	$-22 < M_r < -21$	10582	0.24 ± 0.52	1.09 ± 0.06	450 ± 140
$0.05 < z < 0.26$	0.19	$-22 < M_r < -21$	10323	1.09 ± 0.39	1.07 ± 0.06	518 ± 51
$0.17 < z < 0.22$	0.20	$-21 < M_r < -20$	11006	0.33 ± 0.31	1.00 ± 0.11	412 ± 162
$0.02 < z < 0.17$	0.13	$-21 < M_r < -20$	10525	0.42 ± 0.23	1.00 ± 0.12	492 ± 109
$0.005 < z < 0.14$	0.11	$-20 < M_r < -19$	9594	0.65 ± 0.34	1.00 ± 0.09	409 ± 139
$0.005 < z < 0.09$	0.07	$-19 < M_r < -18$	3202	1.52 ± 0.36	0.62 ± 0.18	755 ± 224

this $L \gtrsim L^*$ galaxy sample, but again without marginalizing over the $f_g\sigma_8$ and $b\sigma_8$. There are 28 258 galaxies in this sample. The pvd as a function of transverse separation for this sample is shown in Fig. 6.12, where the theoretically expected scale-dependence of pvd is more evident, something which suggests that the current data as shown in Fig. 6.11 (i.e. splitting the main sample to very fine luminosity bins) is not good enough to capture all information about pvd for small samples.

Fig. 6.12 presents our best estimates for the pairwise velocity dispersion as a function of separation for $L \gtrsim L^*$ galaxies. For small, intra-halo scales we have $\sigma_{12}(s_\perp = 1.5\ h^{-1}\mathrm{Mpc}) = 551 \pm 71$ km/s, which as we go larger separations rises to $\sigma_{12}(s_\perp = 1.5\ h^{-1}\mathrm{Mpc}) = 608 \pm 96$ km/s. After this peak $\sigma_{12}(s_\perp)$ decreases and remains approximately constant with a large-scale value of $\sigma_{12}(s_\perp = 1.5\ h^{-1}\mathrm{Mpc}) \approx 400$ lm/s.

6.5 Comparison with previous studies

We now seek to compare our results of the pairwise velocity dispersion as a function of luminosity with previous studies. As we mentioned before, the process of fitting the anisotropic correlation function is involved and therefore it is not surprising that different groups follow different approaches to it. In Sections 6.4.3 and 6.4.4 we showed that our main results consist of the luminosity dependence of a constant as well as a function of scale pairwise velocity dispersion.

Our results of the scale-independent pairwise velocity dispersion favor a constant value of $\sigma_{12} \approx 400$ km s^{-1} across all luminosity ranges, apart from the faintest luminosity bin, for

Table 6.4: Redshift space distortion measurements at transverse separations $\sigma_{12}(s_\perp = 1.5\ h^{-1}\mathrm{Mpc})$ for volume limited samples defined in absolute magnitude thresholds. Uncertainties come from the scatter of the fitted parameters from the 9 mocks.

Redshift range	z_{median}	Magnitude limit	N_g	$f_g\sigma_8$	$b\sigma_8$	σ_{12} (km s^{-1})
$0.26 < z < 0.32$	0.29	$M_r < -21$	11642	0.46 ± 0.35	1.11 ± 0.04	506 ± 115
$0.05 < z < 0.26$	0.19	$M_r < -21$	11118	0.87 ± 0.49	1.14 ± 0.06	460 ± 85
$0.17 < z < 0.22$	0.20	$M_r < -20$	14880	0.35 ± 0.33	1.04 ± 0.11	379 ± 193
$0.02 < z < 0.17$	0.13	$M_r < -20$	13873	0.77 ± 0.21	1.05 ± 0.11	576 ± 119
$0.005 < z < 0.14$	0.11	$M_r < -19$	17447	0.55 ± 0.24	1.07 ± 0.09	387 ± 105
$0.005 < z < 0.09$	0.07	$M_r < -18$	7547	1.66 ± 0.39	0.81 ± 0.18	559 ± 183

which we have limited statistics. This value (obtained assuming an exponential form of the pairwise velocity dispersion function) is consistent with the results of Landy (2002) who analyzed both SDSS and 2dFGRS data. Note that their fitting method is very different from ours as they do not fit the $\xi(s_\perp, s_\parallel)$ but they deconvolve it in order to extract directly the pairwise velocity dispersion function. Our results are also in agreement with the more recent 2dFGRS analysis of Hawkins et al. (2003) ($\sigma_{12} = 506 \pm 52$ km s^{-1}), as well as the higher redshift result of Guzzo et al. (2008) who found $\sigma_{12} = 412$ km s^{-1}. Finally, we note that Fisher et al. (1994) also found similar values for σ_{12} using infrared-selected galaxies. On the other hand Zehavi et al. (2002) found a somewhat higher value of $\sigma_{12} \approx 600$ km s^{-1}, a measurement for which they show that is dominated by the respective one obtained only from red galaxies. In conclusion the galaxy pairwise velocity dispersion can be very well fitted by an exponential with $\sigma_{12} \approx 400$ km s^{-1}, with very little dependence on luminosity for $L \gtrsim L^*$ galaxies. This result, aside from being an important result for galaxy evolution per se, is also a very useful information for cosmology, since the pairwise velocity dispersion enters in some form in all growth rate estimations. We expect this result to be also useful for measurements of f_g using the multiple galaxy populations technique of McDonald and Seljak (2009).

Next, we compare our results with literature results for the scale-dependent pairwise velocity dispersion. Two main studies (by the same group) precede the present GAMA one, namely Jing and Börner (2004) (2dFGRS) and Li et al. (2006) (SDSS). We note that GAMA data adds redshift information, as well due to its depth. In Fig 6.13 we compare

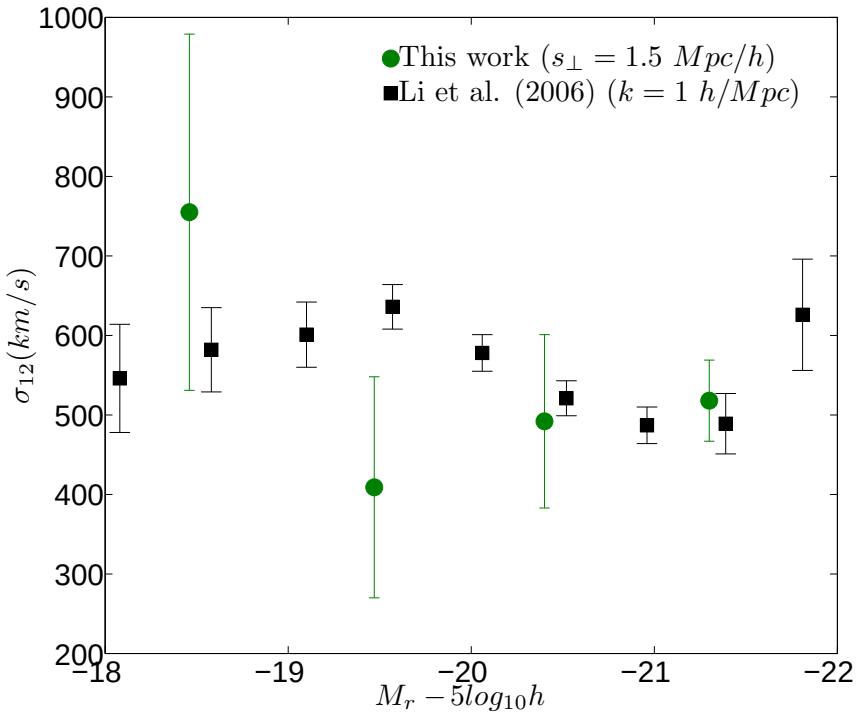

Figure 6.13: Comparison of pairwise velocity dispersion measurements between this work (circles) and Li et al. (2006) (squares). Data points show the pairwise velocity dispersion as function of r-band absolute magnitude, where we use the median magnitude of each luminosity bin. For the two bright samples we present the low-z measurement of σ_{12} to facilitate comparisons. Despite the differences between the two studies, the luminosity dependence of the pairwise velocity dispersion at separations of $\sim 1h^{-1}$Mpc is comparable. For the differences between the two methods of estimating σ_{12} see text.

the low-z results of this work at transverse scales of approximately 1.5 h^{-1}Mpc with Li et al. (2006) who worked in k-space and quote their main results at $k = 1$ hMpc^{-1}. Li et al. (2006) used SDSS data and therefore their study is more convenient for comparing luminosity ranges as GAMA also used SDSS r-band selections, albeit we utilize different k-corrections. It is argued in Li et al. (2006) that the differences between $\sigma_{12}(r)$ and $\sigma_{12}(k)$ are less than 15 per cent if one uses $r = 1/k$ (as opposed to $r = \pi/k$) and we further assume that $\sigma_{12}(s_\perp) \approx \sigma_{12}(r)$ (see Cabré and Gaztañaga, 2009b, for a systematic study of this assumption using simulations). From Fig. 6.13 we observe that our results are in good agreement for $L \gtrsim L^*$ galaxies and deviate for $L < L^*$ galaxies. Fig. 6.13 also shows that at separations of ~ 1 h^{-1}Mpc the pairwise velocity dispersion increases as we go fainter.

We expect that the differences between working either in Fourier space or configuration space to be minimal. However, the main difference between our study and Li et al. (2006) is that we explicitly take into account in our parameter fits the correlation between σ_{12} and the growth rate f_g. Thus, our results are bound to be less biased, although they will have larger uncertainties. Li et al. (2006) show that their results are generally insensitive to luminosity dependent bias (see Section 5.8.1) however they alway keep f_g fixed to its fiducial value ($f_g \approx 0.5$). The situation is somewhat similar with the case of cosmological parameter constraints using the two-point correlation function, where one first assumes some fiducial cosmology to construct $\xi(r)$ and then one varies the cosmological parameters of the model to in order to obtain joint constraints.

6.6 Discussion and conclusions

In this Chapter we have demonstrated how GAMA can reliably constrain crucial dynamical quantities like the growth rate of structure and pairwise velocity dispersion over the non-linear and quasi-linear regime ($0.25 < s < 20$ h^{-1}Mpc). For the parameter fitting we use the observed 2-dimensional correlation function, the most direct measurement from the data. The pairwise velocity dispersion is a highly non-linear effect and therefore we perform the fits on the parameters using the entire plane instead of using the multipole expansion of $\xi(s_\perp, s_\parallel)$. We also fit in the configuration space, which is preferable as observational data tends to be noisy for Fourier transformations. Furthermore, we were able to split the $\xi(s_\perp, s_\parallel)$ plane along the LOS in order to measure the pairwise velocity dispersion as a function of the transverse separation, again providing the most direct measurement from the data. We imposed no limits on the $\xi(s_\perp, s_\parallel)$ scales used in the fits, other than $r < 20$ h^{-1}Mpc, the limit where we showed that our $\xi(s_\perp, s_\parallel)$ measurements are dominated

by noise which biases the results.

The two main complications that we faced is the fact that the redshift space galaxy correlation function is strongly covariant and that there exist degeneracies between the parameters we fit $(\sigma_{12}, f_g, \sigma_8, b)$. For the case of σ_{12}, which nonetheless introduces a certain scale dependence, we break the degeneracies using combinations of the aforementioned parameters, whereas to account for the covariances we use mock catalogues and then calculate the correlations between the parameter fits from different mocks.

We note that our approach for fitting the "Kaiser plus Lorentzian" model to the data, differs from previous studies (Jing and Börner, 2004; Li et al., 2006; Cabré and Gaztañaga, 2009b) in the sense that we do not make any assumptions about the parameters of the model, i.e. the value and the form of the bias, the growth rate and σ_8. We also investigated the shape of the likelihood space, assisted by theoretical studies (Percival and White, 2009), finding evidence for a Gaussian likelihood of the parameter set $(f_g\sigma_8, b_L\sigma_8, \sigma_{12})$. Approximating the likelihood with a trivariate Gaussian and then using parameter estimations from the GAMA mocks, we efficiently marginalize over $f_g\sigma_8$ and $b_L\sigma_8$ in order to obtain the final uncertainties of the scale independent pairwise velocity dispersion.

Our results are as follows. For volume limited samples defined in luminosity bins and for a model with pvd independent of separation and taken to have the form of an exponential function, we find $\sigma_{12} \approx 400$ km s^{-1} for galaxies with $L \gtrsim L^*$. The value of the pvd doubles for $L < L^*$ galaxies,compared with $L \gtrsim L^*$ galaxies, whereof we find $\sigma_{12} = 824 \pm 146$ km s^{-1}. This is a strong indication that these galaxies are almost exclusively satellites in dense and hot haloes. The values of σ_{12} are very similar for volume limited samples defined in luminosity bins and luminosity thresholds. However, σ_{12} values differ significantly for samples with $M_r - 5\log h < -18$. In sharp contrast with the sample defined by $-19 < M_r - 5\log h < -18$ galaxies with $M_r - 5\log h < -18$ have approximately the same σ_{12} as any other sample. We believe this is due to the mixture of satellites and central galaxies in this luminosity threshold sample ($f_{\rm sat} \sim 33\%$, Zehavi et al., 2011).

We also measured the pairwise velocity dispersion as a function of transverse separation. For the same volume limited samples the statistical uncertainties were too large to clearly observe possible scale-dependence. However, when we repeated the analysis with a much larger volume limited sample we observed a scale-depedence of the pvd for $L \gtrsim L^*$ galaxies. The observed scale-dependence is in accord with theoretical expectations (Sheth et al., 2001).

We also observed a lack of substantial ($> 10\%$ uncertainties) constraining power on

$f_g\sigma_8$ when fitting each volume limited sample independently. In principle be compensated by rsd measurements for multiple galaxy populations (McDonald and Seljak, 2009; White et al., 2009), using their auto- and cross-correlations. Although our sample selection and corresponding analysis is not designed for competitive measurements of cosmological parameters, we presented $f_g\sigma_8$ measurements for 5 volume limited samples with $M_r - 5\log h < -19$ which are generally consistent with the fiducial ΛCDM value ($f_g = \Omega_m^{0.55}, \sigma_8 = 0.8$). This consistency check, demonstrates that potential biases on the estimation of σ_{12} due to galaxy coherent flows in the linear regime are minimized.

Some of the concerns raised by this study are likely to be valid for further applications of rsd in cosmology, especially when trying to measure cosmological parameters and to test general relativity via the growth rate of structure. We highlighted two major complications for redshift space distortions analyses. First, real to redshift space mapping is non-linear and detailed knowledge of the galaxy power spectrum on all scales is needed in order to use redshift space distortions in cosmology. Thus, the assumption that cosmological parameters can be measured using only large scales and consequently an effective linear bias, is no longer valid if we want to extract information from redshift space distortions. Secondly one needs to construct robust covariance matrices which will subsequently give realistic posterior distributions. Only then can one efficiently marginalize over nuisance parameters such as the bias. Ultimately it is not surprising that measurements of the growth rate of structure are more compelling than measurements of the expansion rate. As we enter the era of high precision measurements of observables from first order perturbations in cosmology (Tegmark, 2002), we should not be surprised of being challenged to work on the borderline between extragalactic astronomy and astrophysics.

Chapter 7

Conclusions

It's been shown that the clustering of galaxies can offer important insights into our understanding of the universe (Peebles, 1980). In this thesis we explored some of the aspects of that programme using data from two of the most recent galaxy surveys. The analysis presented here has been done using public data provided by the SDSS collaboration (Abazajian et al., 2009) and proprietary data from the GAMA survey (Driver et al., 2011). The detailed description of the data that we used is presented in Chapter 3. In Chapter 5 we studied the clustering of galaxies in the SDSS footprint using the angular correlation function statistic. In order to do this, we used the spectroscopic data from the GAMA survey to calibrate photometric redshifts for SDSS galaxies. This work is presented in Chapter 4. Lastly, in Chapter 6 we studied the clustering of galaxies in redshift space again using GAMA data. We now summarize our main findings.

7.1 Galaxy clustering using calibrated photometric redshifts

We obtained our photometric redshifts using an empirical technique, in particular the software package ANNz (Collister and Lahav, 2004). In the ideal case where the training set for ANNz is complete, that is we have spectroscopic redshifts for a subset of galaxies going as deep as our imaging sample, the performance of ANNz is optimized. We observed a very competitive recovery of the true spectroscopic redshift with a dispersion $\sigma_z = 0.039$ ($\sigma_z = 0.035$ for limit $z_{\rm phot} < 0.4$ which we use for this work) for a sample of 4289 223 SDSS galaxies with $r_{\rm petro} < 19.4$ mag. Subsequently using our accurate photo-z's we derived distance moduli, absolute magnitudes and colours and then constructed galaxy samples split in photo-z, absolute magnitude and colour in order to study the dependence of galaxy clustering on these properties.

The fact that we worked with a complete training set allowed us to construct robust redshift and absolute and magnitude distributions for all our samples. We achieved this assuming a Gaussian distribution of the photo-z uncertainties and then calculating the standard deviation in narrow photo-z bins for each sample. Thus, our results have a direct correspondence with other clustering studied using spectroscopic redshifts (Zehavi et al., 2011).

We then use our redshift distributions for the Limber (1953) inversion of angular clustering measurements. A major advantage of photo-z analyses is their ability to study efficiently faint galaxies. In the sample presented in this thesis there are galaxies as faint as $M_r - 5\log h = -14$ mag. Since our bright limit is $M_r - 5\log h = -24$ mag, our clustering study spans almost three orders of magnitude in L/L^*. We find that the best way to compare photo-z clustering studies with other studies that use spectroscopic redshifts is by the relative bias (Norberg et al., 2001). Moreover, we showed that the linear relation between the relative bias and L/L^* (Norberg et al., 2001; Zehavi et al., 2011) holds down to luminosities $\sim 0.03L^*$.

For samples selected in luminosity and colour, we confirmed the increase of the clustering length for sub-L^* red galaxies compared to with L^* galaxies at small scales. As we go brighter the correlation length for red galaxies with $L > L^*$ also increases. However, our large uncertainties on b/b^*, makes it difficult for constructing a robust relation with L/L^* for red galaxies. For blue galaxies the luminosity-bias relation is very similar with galaxies selected in luminosity only. Yet, compared to brighter blue samples the slope of the $-17 < M_r - 5\log h < -14$ sample is found to be steeper at the 2σ level.

For faint blue galaxies we found a significant contamination due to over-deblendend galaxies. This contamination significantly affects our small-scale clustering results and we accounted for it in our analysis. A preliminary visual inspection of a small fraction of galaxies in our sample, allowed us to quantify the effect of contamination as a function of pair separation and thus present robust measurements of $w(\theta)$.

We also studied the redshift dependence of the absolute bias of L^* galaxies in photo-z bins. We use 4 photo-z bins of width $\delta z = 0.1$ up to redshift $z_{\rm phot} = 0.4$. We found that the observed bias for L* galaxies is consistent with the passive evolution model of Tegmark and Peebles (1998).

7.2 Redshift space clustering

Using the proprietary data from the GAMA survey, we studied the dependence of the galaxy pairwise velocity dispersion as functions of luminosity and redshift. The depth of GAMA allowed us to create volume limited samples for galaxies with $L \gtrsim L^*$, defined by luminosity bins and luminosity thresholds, in two different redshift bins. For galaxies with $L < L^*$ we did not split the samples in redshift, as they cover a smaller volume. On the other hand the breadth of GAMA is such that we can work using the "flat sky" approximation. We worked with what is arguably the simplest model for redshift space distortions on all scales, the model known as "Kaiser+Lorentzian", where at small scales the correlation function is exponentially damped. We found that this model provides a satisfactory fit for all our galaxy samples in accord with other recent rsd studies (Blake et al., 2011).

Although necessary for clustering analyses the construction of the covariance matrix of the anisotropic correlation function is cumbersome. Notwithstanding, we found that using only the 9 available mocks and their respective scatter severely underestimates the uncertainties on σ_{12} and $f_g\sigma_8$. We circumvented this obstacle by constructing the covariance matrix of the derived parameters from the mocks. Furthermore, by approximating the posterior distribution of the parameter set $(\sigma_{12}, f_g\sigma_8, b\sigma_8)$ with a trivariate Gaussian, we were able to marginalize over the other two parameters, when quoting the uncertainties. This approach, although not ideal, provides more realistic uncertainties.

For galaxies with $L \gtrsim L^*$ we found that the scale-independent pvd is approximately constant with luminosity and redshift with $\sigma_{12} \approx 400$ km s^{-1}. Galaxies with $L < L^*$ exhibit a more complex behaviour. For the volume limited sample defined in the absolute magnitude range $-19 < M_r - 5\log < -18$ we found $\sigma_{12} = 824 \pm 146$ km s^{-1}, whereas for the volume limited sample defined an the absolute magnitude threshold $M_r - 5\log h < -18$ we found $\sigma_{12} = 442 \pm 129$ km s^{-1}. This difference is explained from the fraction of satellite galaxies in each sample, which generally increases as we go fainter.

In order to test the scale-dependence of the pvd, we define stripes of constant transverse separation on the $\xi(s_\perp, s_\parallel)$ plane. Although we performed fits for all our previously defined samples, we observe large uncertainties, which made the interpretation of the results difficult. Thus, we defined a much larger volume limited sample with $M_r - 5\log < -20$ and $z < 0.22$ and repeated the analysis. The pvd increased with increasing transverse separation up to scales of $s_\perp \approx 1.5\ h^{-1}$Mpc and then decreased monotonically for larger $s_\perp$.

Our constraints on $f_g\sigma_8$ derived with the approach described above are generally con-

sistent with the concordant ΛCDM+GR cosmological model. There is however, much room for improvement on this front. At the moment measurements of the growth rate of structure are severely limited by sample variance, which explains why we obtained small uncertainties for each estimation of each individual $f_g\sigma_8$ estimation from the mocks, yet large scatter on the quantity $\langle f_g\sigma_8 \rangle$. A recipe to avoid sample variance has been derived by McDonald and Seljak (2009) (see also White et al., 2009), which takes advantage of the fact that different galaxy populations, with different biases trace the same underlying dark matter field. Application of this method would result in significantly smaller error bars in Fig. 6.10.

7.3 Future work

Both photometric redshifts and redshift space distortions are likely to be important observational probes for the study of the most important open problems in cosmology today: dark energy, gravity theory, dark matter and the relation between light and mass. The study presented in this thesis left some open problems, as well as suggestions how to overcome these problems in future work. We summarize these below.

7.3.1 Photometric redshifts

- Gaussianity of photo-z uncertainties. In this thesis we model the photo-z dependent uncertainty with a Gaussian function, which is likely to be a simplification. It would be important to use more realistic modeling for the photo-z uncertainties, maybe as a sum of Gaussian functions (Ma and Bernstein, 2008), which would also automatically account for multimodal redshift distributions.

- In reality galaxy samples split in photo-z bins will be overlapping. Thus it would be necessary to account for the covariances between different photo-z redshift bins for measurements of the cosmological parameters. Recent work on this subject, which is more pressing if one is interested in measurements of the sound horizon can be found in Sánchez et al. (2011) and Crocce et al. (2011).

- Absolute magnitude distributions for HOD modeling. We showed in Chapter 5 that our results can be interpreted within the halo models of large-scale structure. However, halo models are constructed for volume limited samples and it is almost impossible to do so for photo-z derived absolute magnitude, as we showed in Section

4.3.5. A possible remedy is to convolve the HOD numbers with the observed $N(M)$ in order to directly compare data and theory.

- Strictly speaking, in photo-z clustering studies galaxies are not localized points in space, but are described by probability distributions in redshift and absolute magnitude instead. Ideally we would like to make use of all the available information [i.e. the entire $N(z)$ and $N(M)$] for every galaxy instead of just the (possibly in systematic error) values z and M.

- Photometric redshift catalogues rely on the quality of the imaging data. In this thesis we observe some contamination inssues which primarily affect our faintest galaxy sample. In the future advanced data reduction techniques might be necessary for the construction of even largest ($\sim 10^7$) galaxy catalogues, required by e.g. the DES, PanSTARRS and EUCLID projects.

7.3.2 Redshift space distortions

- It is highly non-trivial to obtain the covariances of $\xi(s_\perp, s_\parallel)$ measurements, compared to an one dimensional statistic like $w(\theta)$ (Chapter 5), due to the requirement to create numerous mock catalogues and subsequently invert very large matrices. The required number of mocks is even more amplified by the fact that mocks with different clustering properties are needed in flux-limited surveys like GAMA. The use of the covariances of the fitted parameters yields more realistic uncertainties, however we would like to have more understanding of the parameter space, than a simple multivariate Gaussian approximation.

- The HOD is a very successful model for the small-scale galaxy correlation function (see e.g. Zehavi et al., 2011). Although it has been translated to redshift space by Tinker (2007) it has yet not been applied to data. Application of HOD modeling in rsd will not only provide measurements of new dynamical parameters like the velocity bias of galaxies, but it will also reduce potential biases on growth rate measurements originating from incomplete modeling of the small-scale correlation function.

- Application of the McDonald and Seljak (2009) method to reduce the effects of cosmic variance on measurements of f_g, despite the difficulties outlines above. This requires simultaneous fit of the correlation functions (or power spectra) of two or more galaxy populations. GAMA is suited for such a clustering analysis.

Bibliography

K. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, K. Anderson et al. (2004). The Second Data Release of the Sloan Digital Sky Survey. *AJ*, 128:502–512, `arXiv:astro-ph/0403325`. Cited on 35

K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto et al. (2009). The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182:543–558, `arXiv:0812.0649`. Cited on 34, 130

Filipe B. Abdalla, Manda Banerji, Ofer Lahav, and Valery Rashkov (2008). A comparison of six photometric redshift methods applied to 1.5 million luminous red galaxies. *ArXiv Astrophysics e-prints*, `0812.3831v1`. Cited on 46

J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, K. S. J. Anderson et al. (2008). The Sixth Data Release of the Sloan Digital Sky Survey. *ApJS*, 175:297–313, `arXiv:0707.3413`. Cited on 38

H. Aihara, C. Allende Prieto, D. An, S. F. Anderson, É. Aubourg et al. (2011). The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. *ApJS*, 193:29, `arXiv:1101.1559 [astro-ph.IM]`. Cited on 86

C. Alcock and B. Paczynski (1979). An evolution free test for non-zero cosmological constant. *Nature*, 281:358. Cited on 103

L. Amendola, S. Appleby, D. Bacon, T. Baker, M. Baldi et al. (2012). Cosmology and fundamental physics with the Euclid satellite. *ArXiv e-prints*, `arXiv:1206.1225 [astro-ph.CO]`. Cited on 4

L. Amendola and S. Tsujikawa (2010). *Dark Energy: Theory and Observations*. Cambridge University Press. Cited on 4, 8, 10, 101

R. E. Angulo, C. M. Baugh, C. S. Frenk, and C. G. Lacey (2008). The detectabil-

ity of baryonic acoustic oscillations in future galaxy surveys. *MNRAS*, 383:755–776, arXiv:astro-ph/0702543. Cited on 33

I. K. Baldry, A. S. G. Robotham, D. T. Hill, S. P. Driver, J. Liske et al. (2010). Galaxy And Mass Assembly (GAMA): the input catalogue and star-galaxy separation. *MNRAS*, 404:86–100, arXiv:0910.5120 [astro-ph.CO]. Cited on 38, 40

J. D. Barrow (2005). *The Infinite Book: A Short Guide to the Boundless, Timeless and Endless*. Vintage. Cited on 4

J. Benjamin, L. van Waerbeke, B. Ménard, and M. Kilbinger (2010). Photometric redshifts: estimating their contamination and distribution using clustering information. *MNRAS*, 408:1168–1180, arXiv:1002.2266 [astro-ph.CO]. Cited on 65

A. A. Berlind and D. H. Weinberg (2002). The Halo Occupation Distribution: Toward an Empirical Determination of the Relation between Galaxies and Mass. *ApJ*, 575:587–616, arXiv:astro-ph/0109001. Cited on 29, 97

F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro (2002). Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rept.*, 367:1–248, arXiv:astro-ph/0112551. Cited on 21, 28

D. Bianchi, L. Guzzo, E. Branchini, E. Majerotto, S. de la Torre et al. (2012). Statistical and systematic errors in redshift-space distortion measurements from large surveys. *ArXiv e-prints*, arXiv:1203.1545 [astro-ph.CO]. Cited on 116

C. Blake and S. Bridle (2005). Cosmology with photometric redshift surveys. *MNRAS*, 363:1329–1348, arXiv:astro-ph/0411713. Cited on 44

C. Blake, S. Brough, M. Colless, C. Contreras, W. Couch et al. (2011). The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z=0.9. *MNRAS*, 415:2876–2891, arXiv:1104.2948 [astro-ph.CO]. Cited on 99, 101, 102, 112, 118, 132

M. R. Blanton, D. W. Hogg, N. A. Bahcall, J. Brinkmann, M. Britton et al. (2003). The Galaxy Luminosity Function and Luminosity Density at Redshift z = 0.1. *ApJ*, 592:819–838, arXiv:astro-ph/0210215. Cited on 56

M. R. Blanton, R. H. Lupton, D. J. Schlegel, M. A. Strauss, J. Brinkmann et al. (2005a). The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. *ApJ*, 631:208–230, arXiv:astro-ph/0410164. Cited on 71

M. R. Blanton and S. Roweis (2007). K-Corrections and Filter Transformations in the Ultraviolet, Optical, and Near-Infrared. *AJ*, 133:734–754, arXiv:astro-ph/0606170. Cited on 41, 56, 105

M. R. Blanton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, D. Finkbeiner et al. (2005b). New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *AJ*, 129:2562–2578, arXiv:astro-ph/0410166. Cited on 73

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk et al. (2006). Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, arXiv:astro-ph/0511338. Cited on 41

M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat (2012). The Milky Way's bright satellites as an apparent failure of ΛCDM. *MNRAS*, 422:1203–1218, arXiv:1111.2048 [astro-ph.CO]. Cited on 42

S. Brough, A. M. Hopkins, R. G. Sharp, M. Gunawardhana, D. Wijesinghe et al. (2011). Galaxy and Mass Assembly (GAMA): galaxies at the faint end of the Hα luminosity function. *MNRAS*, 413:1236–1243, arXiv:1012.3771 [astro-ph.CO]. Cited on 86

R. J. Brunner, A. S. Szalay, and A. J. Connolly (2000). Evolution in the Clustering of Galaxies for Z<1.0. *ApJ*, 541:527–534, arXiv:astro-ph/9907403. Cited on 46

T. Budavári (2009). A Unified Framework for Photometric Redshifts. *ApJ*, 695:747–754, arXiv:0811.2600. Cited on 47

T. Budavári, A. J. Connolly, A. S. Szalay, I. Szapudi, I. Csabai et al. (2003). Angular Clustering with Photometric Redshifts in the Sloan Digital Sky Survey: Bimodality in the Clustering Properties of Galaxies. *ApJ*, 595:59–70, arXiv:astro-ph/0305603. Cited on 65, 71

A. Cabré, P. Fosalba, E. Gaztañaga, and M. Manera (2007). Error analysis in cross-correlation of sky maps: application to the Integrated Sachs-Wolfe detection. *MNRAS*, 381:1347–1368, arXiv:astro-ph/0701393. Cited on 27, 28, 76

A. Cabré and E. Gaztañaga (2009a). Clustering of luminous red galaxies - I. Large-scale redshift-space distortions. *MNRAS*, 393:1183–1208, arXiv:0807.2460. Cited on 99, 112, 114

A. Cabré and E. Gaztañaga (2009b). Clustering of luminous red galaxies - II. Small-scale redshift-space distortions. *MNRAS*, 396:1119–1131, arXiv:0807.2461. Cited on 97, 99, 101, 112, 118, 127, 128

S. M. Carroll (2001). The Cosmological Constant. *Living Reviews in Relativity*, 4:1, arXiv:astro-ph/0004075. Cited on 2, 10

S. M. Carroll (2004). *Spacetime and geometry. An introduction to general relativity*. Addison Wesley. Cited on 7, 8, 9, 11, 13

S. M. Carroll, W. H. Press, and E. L. Turner (1992). The cosmological constant. *ARA&A*, 30:499–542. Cited on 18, 101

L. Christodoulou, C. Eminian, J. Loveday, P. Norberg, I. K. Baldry et al. (2012). Galaxy And Mass Assembly (GAMA): colour- and luminosity-dependent clustering from calibrated photometric redshifts. *MNRAS*, 425:1527–1548, arXiv:1206.0943 [astro-ph.CO]. Cited on 44, 47, 71, 116

Adrian Collister, Ofer Lahav, Chris Blake, Russell Cannon, Scott Croom et al. (2007). Megaz-lrg: A photometric redshift catalogue of one million sdss luminous red galaxies. *MNRAS*, 375:68–76, astro-ph/0607630v2. Cited on 53, 60

A. A. Collister and O. Lahav (2004). ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP*, 116:345–351, arXiv:astro-ph/0311058. Cited on 46, 47, 48, 55, 130

A. J. Connolly, I. Csabai, A. S. Szalay, D. C. Koo, R. G. Kron et al. (1995). Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry. *AJ*, 110:2655, arXiv:astro-ph/9508100. Cited on 46

C. J. Conselice, C. Yang, and A. F. L. Bluck (2009). The structures of distant galaxies - III. The merger history of over 20000 massive galaxies at z $<$ 1.2. *MNRAS*, 394:1956–1972, arXiv:0812.3237. Cited on 93

A. Cooray and R. Sheth (2002). Halo models of large scale structure. *Phys. Rept.*, 372:1–129, arXiv:astro-ph/0206508. Cited on 28, 29, 30, 87, 101, 104, 119

H. G. Corwin, Jr., R. J. Buta, and G. de Vaucouleurs (1994). Corrections and additions to the third reference catalogue of bright galaxies. *AJ*, 108:2128–2144. Cited on 85

M. Crocce, A. Cabré, and E. Gaztañaga (2011). Modelling the angular correlation function and its full covariance in photometric galaxy surveys. *MNRAS*, 414:329–349, arXiv:1004.4640 [astro-ph.CO]. Cited on 25, 133

R. A. C. Croft and M. Dailey (2011). On the measurement of cosmological parameters. *ArXiv e-prints*, arXiv:1112.3108 [astro-ph.CO]. Cited on 33

Carlos E. Cunha, Marcos Lima, Hiroaki Oyaizu, Joshua Frieman, and Huan Lin (2009). Estimating the Redshift Distribution of Photometric Galaxy Samples II. Applications and Tests of a New Method. *MNRAS*, 396:2379–2398, arXiv:0810.2991 [astro-ph]. Cited on 67

M. Davis and M. J. Geller (1976). Galaxy Correlations as a Function of Morphological Type. *ApJ*, 208:13–19. Cited on 71

M. Davis and P. J. E. Peebles (1983). A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, 267:465–482. Cited on 26, 97, 101

S. de la Torre and L. Guzzo (2012). Modelling non-linear redshift-space distortions in the galaxy clustering pattern: systematic errors on the growth rate parameter. *ArXiv e-prints*, arXiv:1202.5559 [astro-ph.CO]. Cited on 103, 104, 118

G. de Vaucouleurs, A. de Vaucouleurs, H. G. Corwin, Jr., R. J. Buta, G. Paturel et al. (1991). *Third Reference Catalogue of Bright Galaxies*. Springer-Verlag. Cited on 85

A. Dressler (1980). Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *ApJ*, 236:351–365. Cited on 71

S. P. Driver, D. T. Hill, L. S. Kelvin, A. S. G. Robotham, J. Liske et al. (2011). Galaxy and Mass Assembly (GAMA): survey diagnostics and core data release. *MNRAS*, 413:971–995, arXiv:1009.0614 [astro-ph.CO]. Cited on 38, 40, 42, 47, 62, 105, 130

S. Dye, S. J. Warren, N. C. Hambly, N. J. G. Cross, S. T. Hodgkin et al. (2006). The UKIRT Infrared Deep Sky Survey Early Data Release. *MNRAS*, 372:1227–1252, arXiv:astro-ph/0603608. Cited on 40

D. J. Eisenstein and W. Hu (1998). Baryonic Features in the Matter Transfer Function. *ApJ*, 496:605, arXiv:astro-ph/9709112. Cited on 21, 23

D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton et al. (2005). Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574, arXiv:astro-ph/0501171. Cited on 21, 31

C. Eminian (2008). *Photometric Redshifts and Physical Properties of Low Redshift Galaxies*. PhD thesis, University of Sussex. Cited on 86

F. Feroz and M. P. Hobson (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384:449–463, arXiv:0704.3704. Cited on 116

F. Feroz, M. P. Hobson, and M. Bridges (2009). MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398:1601–1614, arXiv:0809.3437. Cited on 116

K. B. Fisher, M. Davis, M. A. Strauss, A. Yahil, and J. P. Huchra (1994). Clustering in the 1.2-JY IRAS Galaxy Redshift Survey - Part Two - Redshift Distortions and Xi/r/p, PI. *MNRAS*, 267:927, arXiv:astro-ph/9308013. Cited on 98, 125

W. L. Freedman, B. F. Madore, B. K. Gibson, L. Ferrarese, D. D. Kelson et al. (2001). Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *ApJ*, 553:47–72, arXiv:astro-ph/0012376. Cited on 9

E. Gaztañaga, A. Cabré, F. Castander, M. Crocce, and P. Fosalba (2009). Clustering of luminous red galaxies - III. Baryon acoustic peak in the three-point correlation. *MNRAS*, 399:801–811, arXiv:0807.2448. Cited on 21

T. Goto, C. Yamauchi, Y. Fujita, S. Okamura, M. Sekiguchi et al. (2003). The morphology-density relation in the Sloan Digital Sky Survey. *MNRAS*, 346:601–614, arXiv:astro-ph/0312043. Cited on 71

J. R. Gott, III, M. Jurić, D. Schlegel, F. Hoyle, M. Vogeley et al. (2005). A Map of the Universe. *ApJ*, 624:463–484, arXiv:astro-ph/0310571. Cited on 36, 37, 75

L. Guzzo, M. Pierleoni, B. Meneux, E. Branchini, O. Le Fèvre et al. (2008). A test of the nature of cosmic acceleration using galaxy redshift distortions. *Nature*, 451:541–544, arXiv:0802.1944. Cited on 4, 99, 101, 103, 114, 116, 125

L. Guzzo, M. A. Strauss, K. B. Fisher, R. Giovanelli, and M. P. Haynes (1997). Redshift-Space Distortions and the Real-Space Clustering of Different Galaxy Types. *ApJ*, 489:37, arXiv:astro-ph/9706150. Cited on 71

A. J. S. Hamilton (1992). Measuring Omega and the real correlation function from the redshift correlation function. *ApJ*, 385:L5–L8. Cited on 97, 101

A. J. S. Hamilton (1993). Toward Better Ways to Measure the Galaxy Correlation Function. *ApJ*, 417:19. Cited on 26, 27

A. J. S. Hamilton (1998). Linear Redshift Distortions: a Review. In D. Hamilton, editor, *The Evolving Universe*, volume 231 of *Astrophysics and Space Science Library*, page 185. Cited on 96, 97, 100

A. J. S. Hamilton (2000). Uncorrelated modes of the non-linear power spectrum. *MNRAS*, 312:257–284, `arXiv:astro-ph/9905191`. Cited on 18, 23, 90, 113

A. J. S. Hamilton (2001). Formulae for growth factors in expanding universes containing matter and a cosmological constant. *MNRAS*, 322:419–425, `arXiv:astro-ph/0006089`. Cited on 18, 93

A. J. S. Hamilton and M. Tegmark (2004). A scheme to deal accurately and efficiently with complex angular masks in galaxy surveys. *MNRAS*, 349:115–128, `arXiv:astro-ph/0306324`. Cited on 73, 105

E. Hawkins, S. Maddox, S. Cole, O. Lahav, D. S. Madgwick et al. (2003). The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. *MNRAS*, 346:78–96, `arXiv:astro-ph/0212375`. Cited on 97, 101, 116, 125

H. Hildebrandt, S. Arnouts, P. Capak, L. A. Moustakas, C. Wolf et al. (2010). PHAT: PHoto-z Accuracy Testing. *A&A*, 523:A31, `arXiv:1008.0658 [astro-ph.CO]`. Cited on 47

D. W. Hogg, I. K. Baldry, M. R. Blanton, and D. J. Eisenstein (2002). The K correction. *ArXiv Astrophysics e-prints*, `arXiv:astro-ph/0210394`. Cited on 41

D. W. Hogg, M. R. Blanton, D. J. Eisenstein, J. E. Gunn, D. J. Schlegel et al. (2003). The Overdensities of Galaxy Environments as a Function of Luminosity and Color. *ApJ*, 585:L5–L9, `arXiv:astro-ph/0212085`. Cited on 72, 94

E. Hubble (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173. Cited on 1

E. Hubble (1934). The Distribution of Extra-Galactic Nebulae. *ApJ*, 79:8. Cited on 3

E. P. Hubble (1926). Extragalactic nebulae. *ApJ*, 64:321–369. Cited on 2

A. Jenkins, C. S. Frenk, S. D. M. White, J. M. Colberg, S. Cole et al. (2001). The mass function of dark matter haloes. *MNRAS*, 321:372–384, arXiv:astro-ph/0005260. Cited on 28

E. Jennings, C. M. Baugh, and S. Pascoli (2011). Modelling redshift space distortions in hierarchical cosmologies. *MNRAS*, 410:2081–2094, arXiv:1003.4282 [astro-ph.CO]. Cited on 103, 118

Y. P. Jing and G. Börner (2004). The Pairwise Velocity Dispersion of Galaxies: Luminosity Dependence and a New Test of Galaxy Formation Models. *ApJ*, 617:782–793, arXiv:astro-ph/0406077. Cited on 97, 101, 125, 128

N. Kaiser (1987). Clustering in real space and in redshift space. *MNRAS*, 227:1–21. Cited on 96, 97, 100

J. Kwan, G. F. Lewis, and E. V. Linder (2012). Mapping Growth and Gravity with Robust Redshift Space Distortions. *ApJ*, 748:78, arXiv:1105.1194 [astro-ph.CO]. Cited on 103

O. Lahav and A. R Liddle (2010). The Cosmological Parameters 2010. *ArXiv e-prints*, arXiv:1002.3488 [astro-ph.CO]. Cited on 30

S. D. Landy (2002). The Pairwise Velocity Distribution Function of Galaxies in the Las Campanas Redshift Survey, Two-Degree Field Survey, and Sloan Digital Sky Survey. *ApJ*, 567:L1–L1, arXiv:astro-ph/0202130. Cited on 101, 125

S. D. Landy and A. S. Szalay (1993). Bias and variance of angular correlation functions. *ApJ*, 412:64–71. Cited on 26, 27, 73, 104

S. D. Landy, A. S. Szalay, and T. J. Broadhurst (1998). The Pairwise Velocity Distribution of Galaxies in the Las Campanas Redshift Survey. *ApJ*, 494:L133, arXiv:astro-ph/9711045. Cited on 101

G. Lemaître (1927). Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Societe Scietifique de Bruxelles*, 47:49–59. Cited on 1

G. Lemaître (1931a). Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *MNRAS*, 91:483–490. Cited on 1

G. Lemaître (1931b). The Beginning of the World from the Point of View of Quantum Theory. *Nature*, 127:706. Cited on 2

A. Lewis and S. Bridle (2002). Cosmological parameters from CMB and other data: A Monte Carlo approach. *Phys. Rev. D*, 66(10):103511, `arXiv:astro-ph/0205436`. Cited on 30, 116

Antony Lewis, Anthony Challinor, and Anthony Lasenby (2000). Efficient computation of CMB anisotropies in closed FRW models. *ApJ*, 538:473–476, `astro-ph/9911177`. Cited on 21, 22, 90, 112

C. Li, Y. P. Jing, G. Kauffmann, G. Börner, X. Kang et al. (2007). Luminosity dependence of the spatial and velocity distributions of galaxies: semi-analytic models versus the Sloan Digital Sky Survey. *MNRAS*, 376:984–996, `arXiv:astro-ph/0701218`. Cited on 97

C. Li, Y. P. Jing, G. Kauffmann, G. Börner, S. D. M. White et al. (2006). The dependence of the pairwise velocity dispersion on galaxy properties. *MNRAS*, 368:37–47, `arXiv:astro-ph/0509874`. Cited on 97, 101, 125, 126, 127, 128

A. Liddle (2003). *An Introduction to Modern Cosmology, Second Edition*. Wiley. Cited on 11

P. B. Lilje and G. Efstathiou (1988). The cross-correlation of Abell clusters with the Lick galaxy counts. *MNRAS*, 231:635–655. Cited on 103

S. J. Lilly, O. Le Fèvre, A. Renzini, G. Zamorani, M. Scodeggio et al. (2007). zCOSMOS: A Large VLT/VIMOS Redshift Survey Covering $0 < z < 3$ in the COSMOS Field. *ApJS*, 172:70–85, `arXiv:astro-ph/0612291`. Cited on 55

D. N. Limber (1953). The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field. *ApJ*, 117:134. Cited on 24, 25, 131

H. Lin, H. K. C. Yee, R. G. Carlberg, S. L. Morris, M. Sawicki et al. (1999). The CNOC2 Field Galaxy Luminosity Function. I. A Description of Luminosity Function Evolution. *ApJ*, 518:533–561, `arXiv:astro-ph/9902249`. Cited on 41

E. V. Linder (2005). Cosmic growth history and expansion history. *Phys. Rev. D*, 72(4):043529, `arXiv:astro-ph/0507263`. Cited on 4, 100, 114, 119

J. M. Lotz, P. Jonsson, T. J. Cox, D. Croton, J. R. Primack et al. (2011). The Major and Minor Galaxy Merger Rates at z $<$ 1.5. *ApJ*, 742:103, arXiv:1108.2508 [astro-ph.CO]. Cited on 93

J. Loveday, S. J. Maddox, G. Efstathiou, and B. A. Peterson (1995). The Stromlo-APM redshift survey. 2: Variation of galaxy clustering with morphology and luminosity. *ApJ*, 442:457–468, arXiv:astro-ph/9410018. Cited on 71

J. Loveday, P. Norberg, I. K. Baldry, S. P. Driver, A. M. Hopkins et al. (2012). Galaxy and Mass Assembly (GAMA): ugriz galaxy luminosity functions. *MNRAS*, 420:1239–1262, arXiv:1111.0166 [astro-ph.CO]. Cited on 41, 42, 56, 57, 58, 71, 105

R. H. Lupton, J. E. Gunn, and A. S. Szalay (1999). A Modified Magnitude System that Produces Well-Behaved Magnitudes, Colors, and Errors Even for Low Signal-to-Noise Ratio Measurements. *AJ*, 118:1406–1410, arXiv:astro-ph/9903081. Cited on 34

Z. Ma and G. Bernstein (2008). Size of Spectroscopic Calibration Samples for Cosmic Shear Photometric Redshifts. *ApJ*, 682:39–48, arXiv:0712.1562. Cited on 133

M. Masjedi, D. W. Hogg, R. J. Cool, D. J. Eisenstein, M. R. Blanton et al. (2006). Very Small Scale Clustering and Merger Rate of Luminous Red Galaxies. *ApJ*, 644:54–60, arXiv:astro-ph/0512166. Cited on 72

T. Matsubara (2000). The Correlation Function in Redshift Space: General Formula with Wide-Angle Effects and Cosmological Distortions. *ApJ*, 535:1–23, arXiv:astro-ph/9908056. Cited on 104

H. J. McCracken, O. Ilbert, Y. Mellier, E. Bertin, L. Guzzo et al. (2008). Clustering properties of a type-selected volume-limited sample of galaxies in the CFHTLS. *A&A*, 479:321–334, arXiv:0711.4204. Cited on 71

P. McDonald and U. Seljak (2009). How to evade the sample variance limit on measurements of redshift-space distortions. *JCAP*, 10:7, arXiv:0810.0323. Cited on 4, 125, 129, 133, 134

C. J. Miller (1974). The Jack-knife - a review. *Biometrica*, 61(1):1–15. Cited on 75

J. F. Navarro, C. S. Frenk, and S. D. M. White (1996). The Structure of Cold Dark Matter Halos. *ApJ*, 462:563, arXiv:astro-ph/9508025. Cited on 29

P. Norberg, C. M. Baugh, E. Gaztañaga, and D. J. Croton (2009). Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics. *MNRAS*, 396:19–38, arXiv:0810.1885. Cited on 27, 28, 37, 76, 80, 103, 112, 116

P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, D. Madgwick et al. (2002). The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type. *MNRAS*, 332:827–838, arXiv:astro-ph/0112043. Cited on 71, 72, 88, 94

P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, J. A. Peacock et al. (2001). The 2dF Galaxy Redshift Survey: luminosity dependence of galaxy clustering. *MNRAS*, 328:64–70, arXiv:astro-ph/0105500. Cited on 72, 88, 89, 93, 131

P. Norberg, E. Gaztañaga, C. M. Baugh, and D. J. Croton (2011). Statistical analysis of galaxy surveys - IV. An objective way to quantify the impact of superstructures on galaxy clustering statistics. *MNRAS*, 418:2435–2450, arXiv:1106.5701 [astro-ph.CO]. Cited on 37, 75

N. Padmanabhan, D. J. Schlegel, D. P. Finkbeiner, J. C. Barentine, M. R. Blanton et al. (2008). An Improved Photometric Calibration of the Sloan Digital Sky Survey Imaging Data. *ApJ*, 674:1217–1233, arXiv:astro-ph/0703454. Cited on 35

T. Padmanabhan (1993). *Structure Formation in the Universe*. Cambridge University Press, pp. 499. Cited on 2, 14

T. Padmanabhan (2008). Dark energy and gravity. *General Relativity and Gravitation*, 40:529–564, arXiv:0705.2533 [gr-qc]. Cited on 2

H. Parkinson (2012). *PhD Thesis*. PhD thesis, University of Edinburgh. Cited on 55, 56, 62

J. A. Peacock (1999). *Cosmological Physics*. pp. 704. Cambridge, UK: Cambridge University Press. Cited on 20

P. J. Peebles and B. Ratra (2003). The cosmological constant and dark energy. *Reviews of Modern Physics*, 75:559–606, arXiv:astro-ph/0207347. Cited on 2, 3

P. J. E. Peebles (1980). *The large-scale structure of the universe*. Princeton University Press. Cited on 3, 6, 15, 16, 20, 21, 24, 25, 30, 65, 93, 97, 100, 118, 130

P. J. E. Peebles (1993). *Principles of Physical Cosmology*. Princeton University Press. Cited on 2, 17, 71

P. J. E. Peebles (2002). From Precision Cosmology to Accurate Cosmology. *ArXiv Astrophysics e-prints*, arXiv:astro-ph/0208037. Cited on 33

P. J. E. Peebles (2003). Open Problems in Cosmology. *ArXiv Astrophysics e-prints*, arXiv:astro-ph/0311435. Cited on 4

P. J. E. Peebles (2012). The natural science of cosmology. *ArXiv e-prints*, arXiv:1203.6334 [astro-ph.CO]. Cited on 3

A. A. Penzias and R. W. Wilson (1965). A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419–421. Cited on 2

W. J. Percival, B. A. Reid, D. J. Eisenstein, N. A. Bahcall, T. Budavari et al. (2010). Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *MNRAS*, 401:2148–2168, arXiv:0907.1660. Cited on 31

W. J. Percival and M. White (2009). Testing cosmological structure formation using redshift-space distortions. *MNRAS*, 393:297–308, arXiv:0808.0003. Cited on 103, 114, 123, 128

S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent et al. (1999). Measurements of Omega and Lambda from 42 High-Redshift Supernovae. *ApJ*, 517:565–586, arXiv:astro-ph/9812133. Cited on 14

V. Petrosian (1976). Surface brightness and evolution of galaxies. *ApJ*, 209:L1–L5. Cited on 35

M. Postman and M. J. Geller (1984). The morphology-density relation - The group connection. *ApJ*, 281:95–99. Cited on 71

W. H. Press and P. Schechter (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438. Cited on 28

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical recipes in FORTRAN. The art of scientific computing*. Cambridge University Press. Cited on 23, 90

A. Raccanelli, L. Samushia, and W. J. Percival (2010). Simulating redshift-space distortions for galaxy pairs with wide angular separation. *MNRAS*, 409:1525–1533, arXiv:1006.1652 [astro-ph.CO]. Cited on 104

B. A. Reid, W. J. Percival, D. J. Eisenstein, L. Verde, D. N. Spergel et al. (2010). Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. *MNRAS*, 404:60–85, arXiv:0907.1659 [astro-ph.CO]. Cited on 31, 32, 33

B. A. Reid, L. Samushia, M. White, W. J. Percival, M. Manera et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measurements of the growth of structure and expansion rate at z=0.57 from anisotropic clustering. *ArXiv e-prints*, arXiv:1203.6641 [astro-ph.CO]. Cited on 121

B. A. Reid and M. White (2011). Towards an accurate model of the redshift-space clustering of haloes in the quasi-linear regime. *MNRAS*, 417:1913–1927, arXiv:1105.4165 [astro-ph.CO]. Cited on 103

A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks et al. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038, arXiv:astro-ph/9805201. Cited on 14

A. Robotham, S. P. Driver, P. Norberg, I. K. Baldry, S. P. Bamford et al. (2010). Galaxy and Mass Assembly (GAMA): Optimal Tiling of Dense Surveys with a Multi-Object Spectrograph. *PASA*, 27:76–90, arXiv:0910.5121 [astro-ph.CO]. Cited on 38

A. S. G. Robotham, I. K. Baldry, J. Bland-Hawthorn, S. P. Driver, J. Loveday et al. (2012). Galaxy And Mass Assembly (GAMA): in search of Milky Way Magellanic Cloud analogues. *MNRAS*, 424:1448–1453, arXiv:1208.4293 [astro-ph.CO]. Cited on 43

A. S. G. Robotham, P. Norberg, S. P. Driver, I. K. Baldry, S. P. Bamford et al. (2011). Galaxy and Mass Assembly (GAMA): the GAMA galaxy group catalogue (G^3Cv1). *MNRAS*, 416:2640–2668, arXiv:1106.1994 [astro-ph.CO]. Cited on 41, 42

I. G. Roseboom, R. J. Ivison, T. R. Greve, A. Amblard, V. Arumugam et al. (2012). The Herschel Multi-tiered Extragalactic Survey: SPIRE-mm photometric redshifts. *MNRAS*, 419:2758–2773, arXiv:1109.2887 [astro-ph.CO]. Cited on 47

A. J. Ross, S. Ho, A. J. Cuesta, R. Tojeiro, W. J. Percival et al. (2011a). Ameliorating systematic uncertainties in the angular clustering of galaxies: a study using the SDSS-III. *MNRAS*, 417:1350–1373, arXiv:1105.2320 [astro-ph.CO]. Cited on 63

A. J. Ross, W. J. Percival, and R. J. Brunner (2010). Evolution of the clustering of photometrically selected SDSS galaxies. *MNRAS*, 407:420–434, arXiv:1002.1476 [astro-ph.CO]. Cited on 72, 94

A. J. Ross, R. Tojeiro, and W. J. Percival (2011b). Understanding the faint red galaxy population using large-scale clustering measurements from SDSS DR7. *MNRAS*, 413:2078–2086, arXiv:1010.1403 [astro-ph.CO]. Cited on 72, 87

L. Samushia, B. A. Reid, M. White, W. J. Percival, A. J. Cuesta et al. (2012). The Clustering of Galaxies in the SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Testing Deviations from Λ and General Relativity using anisotropic clustering of galaxies. *ArXiv e-prints*, arXiv:1206.5309 [astro-ph.CO]. Cited on 121

A. G. Sánchez, M. Crocce, A. Cabré, C. M. Baugh, and E. Gaztañaga (2009). Cosmological parameter constraints from SDSS luminous red galaxies: a new treatment of large-scale clustering. *MNRAS*, 400:1643–1664, arXiv:0901.2570. Cited on 28

E. Sánchez, A. Carnero, J. García-Bellido, E. Gaztañaga, F. de Simoni et al. (2011). Tracing the sound horizon scale with photometric redshift surveys. *MNRAS*, 411:277–288, arXiv:1006.3226 [astro-ph.CO]. Cited on 133

A. Sandage (1961). The Ability of the 200-INCH Telescope to Discriminate Between Selected World Models. *ApJ*, 133:355. Cited on 30

W. Saunders, M. Rowan-Robinson, and A. Lawrence (1992). The spatial correlation function of IRAS galaxies on small and intermediate scales. *MNRAS*, 258:134–146. Cited on 103, 105, 112

D. J. Schlegel, D. P. Finkbeiner, and M. Davis (1998). Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *ApJ*, 500:525, arXiv:astro-ph/9710327. Cited on 35

R. Scoccimarro (2004). Redshift-space distortions, pairwise velocities, and nonlinearities. *Phys. Rev. D*, 70(8):083007, arXiv:astro-ph/0407214. Cited on 98, 100, 102, 103, 118, 123

Hee-Jong Seo and Daniel J. Eisenstein (2003). Probing Dark Energy with Baryonic Acoustic Oscillations from Future Large Galaxy Redshift Surveys. *Astrophys. J.*, 598:720–740, arXiv:astro-ph/0307460. Cited on 44

R. K. Sheth (1996). The distribution of pairwise peculiar velocities in the non-linear regime. *MNRAS*, 279:1310, arXiv:astro-ph/9511068. Cited on 101

R. K. Sheth and A. Diaferio (2001). Peculiar velocities of galaxies and clusters. *MNRAS*, 322:901–917, arXiv:astro-ph/0009166. Cited on 101

R. K. Sheth, L. Hui, A. Diaferio, and R. Scoccimarro (2001). Linear and non-linear contributions to pairwise peculiar velocities. *MNRAS*, 325:1288–1302, arXiv:astro-ph/0009167. Cited on 123, 128

R. K. Sheth and G. Tormen (1999). Large-scale bias and the peak background split. *MNRAS*, 308:119–126, arXiv:astro-ph/9901122. Cited on 29

P. Simon (2007). How accurate is Limber's equation? *A&A*, 473:711–714, arXiv:astro-ph/0609165. Cited on 24, 25

F. Simpson and J. A. Peacock (2010). Difficulties distinguishing dark energy from modified gravity via redshift distortions. *Phys. Rev. D*, 81(4):043512, arXiv:0910.3834 [astro-ph.CO]. Cited on 4

D. S. Sivia and J. Skilling (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press, pp. 246. Cited on 117

R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk et al. (2003). Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, 341:1311–1332, arXiv:astro-ph/0207664. Cited on 21, 22, 90, 102, 112, 114

D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett et al. (2007). Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170:377–408, arXiv:astro-ph/0603449. Cited on 31

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, arXiv:astro-ph/0504097. Cited on 2, 41

C. Stoughton, R. H. Lupton, M. Bernardi, M. R. Blanton, S. Burles et al. (2002). Sloan Digital Sky Survey: Early Data Release. *AJ*, 123:485–548. Cited on 35, 75

M. A. Strauss, D. H. Weinberg, R. H. Lupton, V. K. Narayanan, J. Annis et al. (2002). Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *AJ*, 124:1810–1824, arXiv:astro-ph/0206225. Cited on 35

M. E. C. Swanson, M. Tegmark, M. Blanton, and I. Zehavi (2008a). SDSS galaxy clustering: luminosity and colour dependence and stochasticity. *MNRAS*, 385:1635–1655, arXiv:astro-ph/0702584. Cited on 72, 94

M. E. C. Swanson, M. Tegmark, A. J. S. Hamilton, and J. C. Hill (2008b). Methods for rapidly processing angular masks of next-generation galaxy surveys. *MNRAS*, 387:1391–1402, arXiv:0711.4352. Cited on 73, 105

A. Taruya, T. Nishimichi, and S. Saito (2010). Baryon acoustic oscillations in 2D: Modeling redshift-space power spectrum from perturbation theory. *Phys. Rev. D*, 82(6):063522, arXiv:1006.0699 [astro-ph.CO]. Cited on 103, 118

M. Tegmark (2002). Measuring Spacetime: From the Big Bang to Black Holes. *Science*, 296:1427–1433. Cited on 129

M. Tegmark, M. R. Blanton, M. A. Strauss, F. Hoyle, D. Schlegel et al. (2004). The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, 606:702–740, arXiv:astro-ph/0310725. Cited on 4

M. Tegmark, S. Dodelson, D. J. Eisenstein, V. Narayanan, R. Scoccimarro et al. (2002). The Angular Power Spectrum of Galaxies from Early Sloan Digital Sky Survey Data. *ApJ*, 571:191–205, arXiv:astro-ph/0107418. Cited on 27

M. Tegmark, D. J. Eisenstein, M. A. Strauss, D. H. Weinberg, M. R. Blanton et al. (2006). Cosmological constraints from the SDSS luminous red galaxies. *Phys. Rev. D*, 74(12):123507, arXiv:astro-ph/0608632. Cited on 2, 30, 31, 44, 72

M. Tegmark and P. J. E. Peebles (1998). The Time Evolution of Bias. *ApJ*, 500:L79, arXiv:astro-ph/9804067. Cited on 72, 93, 94, 131

J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren et al. (2008). Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality. *ApJ*, 688:709–728, arXiv:0803.2706. Cited on 29

J. L. Tinker (2007). Redshift-space distortions with the halo occupation distribution - II. Analytic model. *MNRAS*, 374:477–492, arXiv:astro-ph/0604217. Cited on 4, 102, 103, 134

J. L. Tinker, P. Norberg, D. H. Weinberg, and M. S. Warren (2007). On the Luminosity Dependence of the Galaxy Pairwise Velocity Dispersion. *ApJ*, 659:877–889, arXiv:astro-ph/0603543. Cited on 97

J. L. Tinker, D. H. Weinberg, and Z. Zheng (2006). Redshift-space distortions with the halo occupation distribution - I. Numerical simulations. *MNRAS*, 368:85–108, arXiv:astro-ph/0501029. Cited on 103

R. Tojeiro, W. J. Percival, J. Brinkmann, J. R. Brownstein, D. J. Eisenstein et al. (2012). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measuring structure growth using passive galaxies. *MNRAS*, 424:2339–2344, arXiv:1203.6565 [astro-ph.CO]. Cited on 116

Y. Wang (2008). Differentiating dark energy and modified gravity with galaxy redshift surveys. *JCAP*, 5:21, arXiv:0710.3885. Cited on 18, 19

Y. Wang, X. Yang, H. J. Mo, and F. C. van den Bosch (2007). The Cross-Correlation between Galaxies of Different Luminosities and Colors. *ApJ*, 664:608–632, arXiv:astro-ph/0703253. Cited on 71

D. F. Watson, A. A. Berlind, C. K. McBride, D. W. Hogg, and T. Jiang (2012). The Extreme Small Scales: Do Satellite Galaxies Trace Dark Matter? *ApJ*, 749:83, arXiv:1108.1195 [astro-ph.CO]. Cited on 72

S. Weinberg (1989). The cosmological constant problem. *Reviews of Modern Physics*, 61:1–23. Cited on 2

M. White, Y.-S. Song, and W. J. Percival (2009). Forecasting cosmological constraints from redshift surveys. *MNRAS*, 397:1348–1354, arXiv:0810.1518. Cited on 4, 129, 133

S. D. M. White (2007). Fundamentalist physics: why Dark Energy is bad for astronomy. *Reports on Progress in Physics*, 70:883–897, arXiv:0704.2291. Cited on 33

D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis et al. (2000). The Sloan Digital Sky Survey: Technical Summary. *AJ*, 120:1579–1587, arXiv:astro-ph/0006396. Cited on 34

I. Zehavi, M. R. Blanton, J. A. Frieman, D. H. Weinberg, H. J. Mo et al. (2002). Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data. *ApJ*, 571:172–190, arXiv:astro-ph/0106476. Cited on 72, 88, 123, 125

I. Zehavi, D. H. Weinberg, Z. Zheng, A. A. Berlind, J. A. Frieman et al. (2004). On Departures from a Power Law in the Galaxy Correlation Function. *ApJ*, 608:16–24, arXiv:astro-ph/0301280. Cited on 87

I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall et al. (2011). Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *ApJ*, 736:59, arXiv:1005.2413 [astro-ph.CO]. Cited on 28, 57, 58, 60, 71, 72, 75, 79, 87, 88, 89, 90, 94, 110, 119, 128, 131, 134

I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind et al. (2005). The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, 630:1–27, arXiv:astro-ph/0408569. Cited on 71, 72, 75, 87, 88, 94, 95

Z. Zheng, A. A. Berlind, D. H. Weinberg, A. J. Benson, C. M. Baugh et al. (2005). Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies. *ApJ*, 633:791–809, arXiv:astro-ph/0408564. Cited on 87

Z. Zheng and D. H. Weinberg (2007). Breaking the Degeneracies between Cosmology and Galaxy Bias. *ApJ*, 659:1–28, arXiv:astro-ph/0512071. Cited on 97