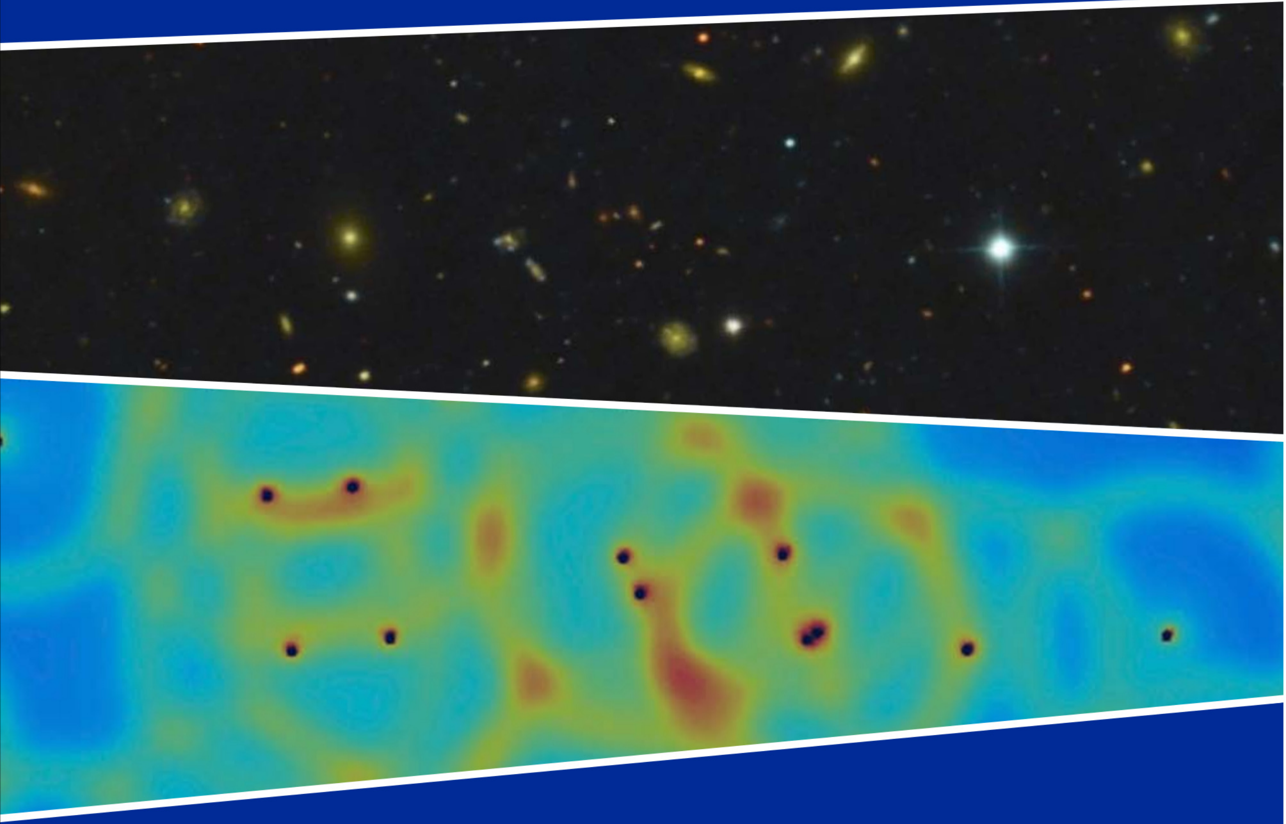
VNIVERSITAT DE VALÈNCIA
Departament d´Astronomia i Astrofísica



Clustering of galaxies: evolution, segregation and baryon acoustic oscillations

Pablo Arnalte Mur
Tesi Doctoral
València, 2011

VNIVERSITAT DE VALÈNCIA

Clustering of galaxies: evolution, segregation and baryon acoustic oscillations

Pablo Arnalte Mur

Tesi Doctoral

Observatori Astronòmic

Departament d'Astronomia i Astrofísica

València, 2011

Disseny de la portada: Raquel Bachiller Martínez.
(Fotografia processada per Vicent Peris a partir de dades d'ALHAMBRA).

Dr. Alberto Fernández Soto,
científic titular CSIC a l'Instituto de Física de Cantabria,

I

Dr. Vicent Josep Martínez García,
catedràtic d'Astronomia i Astrofísica a la Universitat de València,

Certifiquen

Que la present memòria, "Clustering of galaxies: evolution, segregation and baryon acoustic oscillations", ha estat realitzada sota llur direcció per Pablo Arnalte Mur, i que constitueix la seua tesi doctoral per optar al grau de Doctor en Física.

I per què hi quede constància i tinga els efectes oportuns, signem el present document en Paterna a 24 d'octubre de 2011.

Signat: Alberto Fernández Soto Vicent J. Martínez García

a mis padres y mi hermana,
y a Raquel...

Contents

1

Introduction

Cosmology is the science that aims at understanding the origin, evolution, and structure of the universe as a whole. It has been an integral part of the development of science through history, since its very beginning. During the last century, several theoretical and mainly observational developments made it a mature science, getting us closer to understanding the basic properties of our universe.

One of the main observational tools in this endeavour has been the construction of large galaxy surveys covering significant fractions of the sky. By analysing the distribution of galaxies in space, they allow for the study of the large-scale structure (LSS) of the universe. They are also focused on the study of the processes of formation and evolution of galaxies. In the last three decades, together with other cosmological tools, galaxy surveys underwent a great improvement in terms of coverage, depth, and accuracy, thanks to important technical developments. At the same time, the theoretical and statistical tools needed to analyse all these data were developed greatly.

In this thesis, we tackle several open problems in the study of large-scale structure through the clustering of galaxies. To this end, we analyse data from some of the latest surveys, and we also develop new statistical techniques needed for this analysis in specific cases. In the first part, we focus on small and intermediate scales, where the relation between galaxy properties and their clustering (known as *segregation*) is important. The work in this part is driven by the exploitation of data from the ongoing Advanced Large Homogeneous Area Medium-Band Redshift Astronomical (ALHAMBRA) survey, which is perfectly suited to study the evolution of segregation trough cosmic time. In the second part, we focus on the study of a large scale feature of the galaxy distribution, the baryon acoustic oscillations (BAO).

We start this Introduction chapter with an overview of the cosmological model considered standard today (Section 1.1). In particular, we review the description of the cosmological density field in this model, and the process of structure formation. In Section 1.2 we describe some of the standard statistical tools used for the description of the galaxy distribution. Section 1.3 presents a brief review of the properties of galaxy redshift surveys, focusing on those whose data we use in this thesis. Finally, in Section 1.4 we summarise the aims of this thesis, and present an outline of the work.

1.1 The standard cosmological model

In this section, we briefly summarise the standard cosmological model, known as Λ – Cold Dark Matter (ΛCDM) model, which is the basis for any work on cosmology today, including the work presented in this thesis. For a more complete exposition of the cosmological model, and the observational evidence for it, see e.g. Peebles (1993); Peacock (1999); Dodelson (2003).

The development of modern cosmology and of the model of the universe considered as standard today is based on important advancements, both theoretical and observational, that happened during the first third of the XXth century. On the theoretical side, Albert Einstein published its general theory of relativity in 1915 (Einstein, 1915), which changed significantly our understanding of gravity. Several models for the universe were built in the following years based on the new theory by Willem de Sitter, Georges Lemaître, Alexander Friedmann or Einstein himself.

On the observational side, Edwin Hubble made two important contributions. In the first place, he used the period–luminosity relation for Cepheids, previously established by Henrietta Swan Leavitt (Leavitt, 1908; Leavitt and Pickering, 1912), to measure accurately the distances to nebulae, establishing its extragalactic nature (Hubble, 1925a,b). In this way, the issues raised in the ‘Great Debate’ between Heber Curtis and Harlow Shapley (Hoskin, 1976; Trimble, 1995) were settled, and our Milky Way was then regarded as just another galaxy in a vastly larger universe. Hubble’s other important contribution was to establish a relation between the recession velocities v of galaxies and their distances d (Hubble, 1929, although it could be argued that this relation was first established by Lemaître, 1927). He proposed a linear relation,

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

now known as *Hubble’s law*, where H_0 is a constant[1]. This relation implies that the universe is expanding.

From these starting points, the *Hot Big Bang* model was developed during the last century, and is now considered as the standard cosmological model. It is based on three main hypothesis:

---

[1] Hubble gave a value for this constant of $H_0 \simeq 500\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, which is nearly an order of magnitude larger than the current accepted value.

- On large scales, the distribution of matter in the universe is homogeneous and isotropic. This is known as the *cosmological principle.*
- The basic dynamics of the universe is driven by gravity, as described by Einstein's general relativity.
- The universe is expanding from a primordial hot and dense state.

According to this model, the universe started as a very dense and hot plasma formed by photons, fermions and quarks. As time passed, expansion made the density and temperature decrease. This allowed the formation of the first nuclei of light elements (hydrogen, helium and lithium), in a process known as *Big Bang nucleosynthesis.* $380,000$ years after the Big Bang, the temperature dropped to ~ 3000K, and it was possible for nuclei and electrons to form atoms, in a process known as *recombination* (although they had never been 'combined' before). In this way, baryons[2] and radiation decoupled. Once decoupled from the photons, baryons evolved jointly with the Dark Matter, mainly through gravitational collapse, to form stars, galaxies, and the rest of structures we see in the universe today. Radiation from the moment of decoupling simply cooled down, and gave way to the Cosmic Microwave Background (CMB) we observe today.

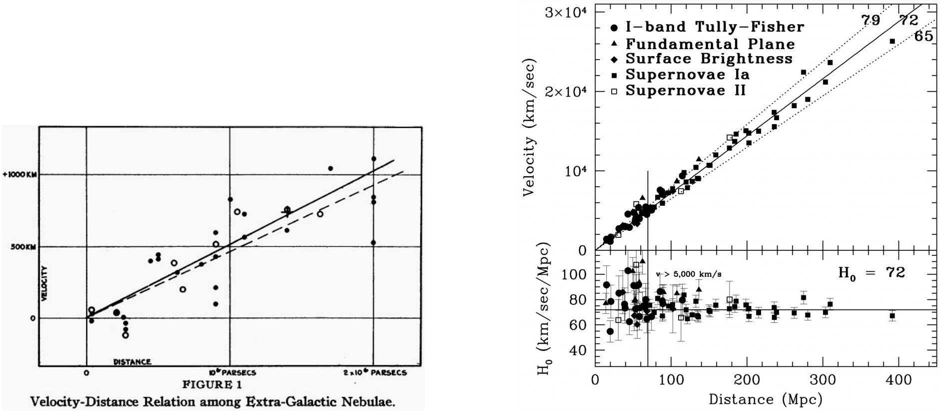
It is standard today to assume that, at the very early stages, around $10^{-34}\,$s after the Big Bang, the universe underwent an exponential expansion, known as *inflation.* This scenario, first proposed by Guth (1981), would solve two fine-tuning problems that appear otherwise in the Big Bang model. On the one side, the fact that the early universe was so nearly homogeneous, and on the other side, the fact that the geometry of the universe is so close to being flat. Moreover, the inflation scenario gives a prediction for some characteristics of the small perturbations present in the early universe, which were the seeds for present day structures.

The validity of this *Hot Big Bang* model is mainly based on three observational facts:

- *Hubble's law* for the recession velocities of galaxies (equation 1.1). This relation implies that the universe is expanding, as all galaxies (on large scales) are receding from each other. It was first established by Slipher's redshift measurements and Hubble's distance determinations using Cepheids or other distance indicators. It has since been confirmed and extended to much larger distances. At the same time, the value of H_0 has been measured with high precision. The most recent results using data from the Hubble Space Telescope (HST), shown in Fig 1.1, gave a value of $H_0 = 72 \pm 8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Freedman et al., 2001). For a review, see Freedman and Madore (2010).

---

[2] In cosmology, we usually call *baryons* all matter particles in the Standard Model of Particle Physics, so we include also fermions, and, in particular, electrons.



(a) (b)

Figure 1.1: Distance — recession velocity diagrams for different types of galaxies, which illustrate Hubble's law (equation 1.1). We show the results obtained by Hubble in 1929 (a), and by the recent HST Key Project (b). Note the huge change in the range of distances studied. The solid line in (a) corresponds to the linear fit to the 24 individual galaxies for which Hubble had distance measurements (black points), while the dashed line is the fit obtained when the galaxies are combined into nine groups (open circles). Different symbols in (b) correspond to different methods to measure extragalactic distances. The solid line is the best-fit relation obtained in this case (with $H_0 = 72\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$), and the dashed lines correspond to different values of H_0, as indicated. Figures from Hubble (1929) and Freedman et al. (2001).

- The detection of the *Cosmic Microwave Background* (CMB) radiation by Penzias and Wilson (1965). This radiation is a relic of the photons emitted at the time of recombination. It has a nearly perfect black body spectrum with a temperature of $T = 2.73$K, shown in Fig. 1.2, and is nearly isotropic, with only small anisotropies of the order of one in 10^5. The presence of this radiation implies that the universe went through a hotter epoch, and has been cooling ever since. Many observations, including the Cosmic Background Explorer (COBE) and Wilkinson Microwave Anisotropy Probe (WMAP) satellites, have measured the spectrum of this radiation, and have mapped its tiny anisotropies, which provide important information about the parameters in the cosmological model. Fig. 1.3 shows the most recent full-sky map of these anisotropies, measured by WMAP.

- The measurement of the *abundance of light elements* in the universe, which

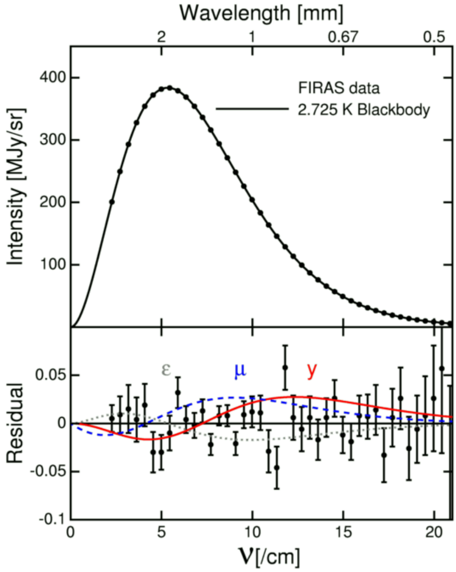


Figure 1.2: Spectrum of the CMB radiation observed by the FIRAS instrument on board of COBE (Mather et al., 1994), showing a remarkable agreement with a black body spectrum. The points with errors are the measurements of FIRAS, and the black solid line is the best-fit black body spectrum. The other lines in the bottom plot show the spectra for possible deviations from the black body case: that of a body with a reflectivity different from zero (dotted grey), and the effect of hot electrons adding an excess energy to the CMB either at $z \gtrsim 10^5$ (dashed blue) or $z \lesssim 10^5$ (solid red). In all cases, the curves shown are the maximum allowed deviations (at 95% confidence level) by FIRAS data. Figure by Ned Wright, http://www.astro.ucla.edu/∼wright/cosmolog.htm.

is in good overall agreement with that predicted by *Big Bang nuclesynthesis* (BBN). During this process, 75% (by mass) went to single protons (hydrogen nuclei), 25% formed helium (^{4}He) nuclei, and only traces appeared of other nuclei: deuterium (^{2}D), ^{3}He and lithium (^{7}Li). Fig. 1.4 shows the agreement between the predicted and observed abundances of ^{4}He and ^{2}D. As seen in the figure, however, there is a discrepancy between the observed and predicted (for the value of Ω_b determined from the CMB) abundances of ^{7}Li (Cyburt et al., 2008). Several possible solutions for this problem have been proposed. Given that the abundance of ^{7}Li is measured from metal-poor population II stars in the Galactic halo, a possibility is the presence of additional systematic effects in the stellar physics involved, such as incorrect

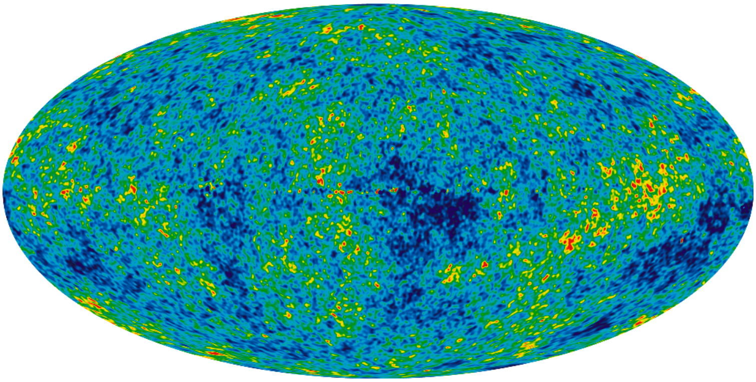


Figure 1.3: Full-sky map, in galactic coordinates, of the CMB temperature anisotropies (of order $\sim 10^{-5}$), as measured by the WMAP satellite after seven years of observations (Jarosik et al., 2011). The map shown here is the 'Internal Linear Combination' map, which is constructed in a way in which all the foregrounds from the Galaxy (due to synchroton, free-free, and dust emission) are removed (see Gold et al., 2011 for details). Figure by the WMAP Science Team, http://lambda.gsfc.nasa.gov/product/map/current/.

determinations of the effective temperatures of the stars (Meléndez and Ramírez, 2004), or additional depletion of ^{7}Li during the life of the star (Korn et al., 2006). An alternative possibility is that this discrepancy is a sign of new physics beyond the standard model of particle physics, such as variations of the fine structure constant α (Dmitriev et al., 2004), or the decay of some additional particle occurred during BBN (Jedamzik, 2004). For a review on BBN, see Olive et al. (2000).

1.1.1 The Friedmann-Lemaître-Robertson-Walker model

Based on the cosmological principle, we can derive the metric that describes the geometry of the universe at large scales. The metric that follows from this assumption is the Robertson-Walker (RW) metric, which has the form[3]

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\right], \tag{1.2}$$

where r, θ and ϕ are comoving spatial spherical coordinates, t is the time, and $a(t)$ is a function that describes the expansion, and is known as *cosmic scale factor*.

---

[3] We use units where the speed of light $c = 1$.

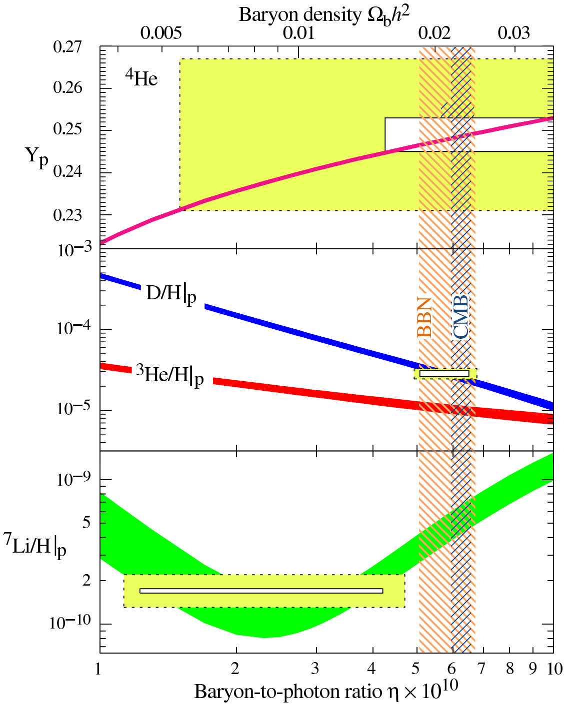


Figure 1.4: Abundance of light elements as a function of the baryon-to-photon ratio η or, equivalently, baryon density Ω_b. The bands show the theoretical predictions, and the boxes the measured constraints on the abundances. Data agrees very well for helium and deuterium, for a value of Ω_b coincident with the value obtained independently from the CMB anisotropies (shaded vertical band), but there is some discrepancy for the lithium abundance (see the text for details). Figure from Nakamura et al. (2010).

By convention, its value at the present-day time t_0 is $a_0 \equiv a(t_0) \equiv 1$[4]. k is the curvature of the three-space. $k = 0$ corresponds to a spacially flat universe, $k > 0$ to a universe with positive curvature (spherical), and $k < 0$ to a universe with negative curvature (hyperbolic).

The Hubble parameter H is defined as

$$H = \frac{\dot{a}}{a},$$

where the dot means a time derivative, and it is, in general, a function of time.

---

[4] By convention, we indicate the values that the different quantities take today by a subscript 0.

As the physical distances between galaxies are given by

$$d = ar\,,$$

we recover Hubble's law (equation 1.1) as a first order approximation for $t \simeq t_0$, with 'Hubble's constant' $H_0 = H(t_0)$. Hubble's constant is usually expressed in terms of the a-dimensional parameter h as $H_0 \equiv 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, with the value $h \simeq 0.7$.

In the case of a photon travelling from a distant galaxy to us, its wavelength would expand with the metric, so that

$$\frac{\lambda_o}{\lambda_e} = \frac{1}{a}$$

where λ_e is the emission wavelength of the photon, λ_o is the wavelength we observe, and a is the scale factor at the time of emission. This means that, if a galaxy emitted some light when the scale factor of the universe was a, we now observe its spectrum redshifted by a quantity

$$z \equiv \frac{\lambda_o - \lambda_e}{\lambda_e} = \frac{1}{a} - 1\,.$$

The quantity z is known as the *cosmological redshift*, and it is used as a measure of the cosmic time t corresponding to the scale factor a. It has the advantage of being directly measurable from the galaxy spectra.

If we introduce the form of the RW metric (equation 1.2) into Einstein's field equations of General Relativity, and assuming that we can describe the contents of the universe as a perfect fluid with density ρ and pressure p, we obtain the Friedmann-Lemaître equations, which are the basic equations to describe the dynamics of the Friedmann-Lemaître-Robertson-Walker (FLRW) universe model:

$$H^2 = \frac{8\pi G\rho}{3} - \frac{k}{a^2} + \frac{\Lambda}{3} \tag{1.3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p\right) + \frac{\Lambda}{3}\,, \tag{1.4}$$

where G is Newton's gravitational constant, and Λ is the *cosmological constant*. This can be interpreted simply as an additional constant of GR Theory appearing in Einstein's equations, or as an additional component of the universe in the form of a vacuum energy (also known as *dark energy*) if we make the identification

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}$$
$$p_\Lambda = -\rho_\Lambda$$

The total energy content of the universe can be expressed in terms of its constituent species, each of them defined by a equation of state $p_i = w_i \rho_i$. For non-relativistic (or 'cold') matter, either baryons of dark matter, we have $w_m = 0$, while for radiation, $w_r = \frac{1}{3}$, and for dark energy, $w_\Lambda = -1$, as shown above. For relativistic matter, such as neutrinos, w_i is no longer a constant, but it varies with time. The equation of state of each component determines the evolution of its density with redshift. From the conservation of energy equation $\mathrm{d}(a^3\rho_i) = -p_i \mathrm{d}a^3$ we obtain, assuming a constant w_i,

$$\rho_i = \rho_{i,0}\,(1+z)^{3(1+w_i)}\ ,$$

where $\rho_{i,0}$ is the density of the species today.

From equation (1.3), we see that there is a clear relation between the energy content of the universe (via ρ), and its geometry (via k). We can define the *critical density*, ρ_c, as the value of the total density (including vacuum energy) today which corresponds to a flat universe (i.e. $k = 0$),

$$\rho_c = \frac{3H_0^2}{8\pi G}\ .$$

This results in a value of $\rho_c = 2.775\times10^{11}\,h^2\,M_\odot\,\mathrm{Mpc}^{-3} = 1.878\times10^{-29}\,h^2\,\mathrm{g\,cm}^{-3}$, which corresponds approximately to 6 hydrogen atoms per cubic metre. We conventionally describe the densities of the different species today using the density parameters $\Omega_i \equiv \frac{\rho_{i,0}}{\rho_c}$,

$$\Omega_m \equiv \frac{\rho_{m,0}}{\rho_c}\ ,\ \Omega_r \equiv \frac{\rho_{r,0}}{\rho_c}\ ,\ \Omega_\Lambda \equiv \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda}{3H_0^2}\ .$$

Analogously, we can define the parameter

$$\Omega_k = -\frac{k}{H_0^2}\ ,$$

to characterise the curvature of the universe, so that

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_k = 1\ .$$

Now, from equation (1.3), the evolution of the expansion rate is given by

$$H(z) = H_0 E(z)\ ,\ E(z) = \sqrt{\Omega_r\,(1+z)^4 + \Omega_m\,(1+z)^3 + \Omega_k\,(1+z)^2 + \Omega_\Lambda}\ .$$

The acceleration of the expansion is described by the *deceleration parameter* $q \equiv -\frac{\ddot{a}a}{\dot{a}^2}$ which, from equation (1.4), has the value at the present time

$$q_0 = \Omega_r + \frac{\Omega_m}{2} - \Omega_\Lambda\ . \tag{1.5}$$

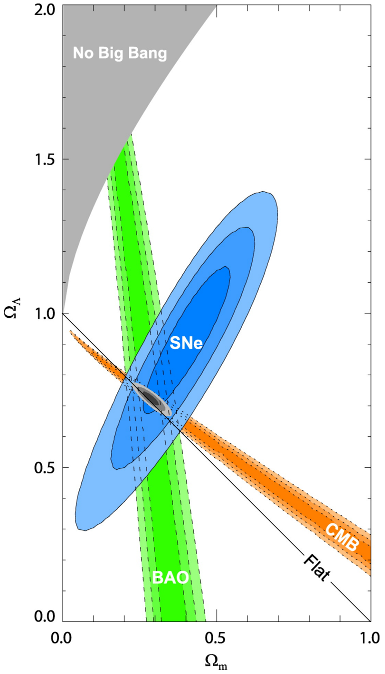


Figure 1.5: Constraints on Ω_m and Ω_Λ from a combination of different observational probes: CMB anisotropies, high-z supernovae (SNe), and baryon acoustic oscillations (BAO) in the galaxy distribution. Contours show the 68%, 95%, and 99.7% likelihood constraints for the different probes, and the combined ones (in grey). Figure from Kowalski et al. (2008).

The accurate determination of these parameters defining the cosmological model has been one of the main research goals in cosmology during recent years. They can be measured through a combination of several observations, mainly the CMB radiation, the distance–redshift relationship for different *standard candles* (such as type Ia supernovae), and the clustering of galaxies as measured in large surveys. Fig. 1.5 shows an example of how these observations give consistent results, implying strong constraints on the parameters. For the most recent determination of these values, see Tegmark et al. (2006); Kowalski et al. (2008); Komatsu et al. (2011). The picture that we obtain from these measurements is that of a (nearly) flat universe, $\Omega_k \simeq 0$. Most of the energy content of the universe is in the form of a cosmological constant, or dark energy, $\Omega_\Lambda \simeq 0.73$, and the rest is matter, $\Omega_m \simeq 0.27$. The radiation content of the universe today is much smaller. Moreover, the matter content can be divided into baryons, with a density $\Omega_b \simeq 0.05$, and dark matter, with $\Omega_{DM} \simeq 0.22$.

Dark matter and dark energy

According to the standard cosmological model, therefore, about 95% of the energy content of the universe is of an unknown type, either dark matter or dark energy.

Dark matter comprises any form of matter which does not couple, either directly or indirectly, to radiation. Hence it does not emit light, and it only interacts via gravity. The first indication for the need of some form of extra matter additional to the luminous matter came from Zwicky (1933). He studied the dynamics of the Coma cluster, and postulated the need for a large amount of extra matter in order to maintain the virial equilibrium, while compensating for the high velocity dispersion. Evidence in favour of the existence of some form of dark matter has grown since then, based on different observations of galaxy clusters, such as velocity dispersion of galaxies, X-ray observations of the hot gas (Kellogg et al., 1971; Gursky et al., 1971), or strong gravitational lensing (Lynds and Petrosian, 1986; Soucail et al., 1987; Paczynski, 1987), but also on the dynamics of galaxies themselves, through the study of the rotation curves of spiral galaxies (Babcock, 1939; Rubin and Ford, 1970; Rubin et al., 1978). From these observations emerged the idea that dark matter is the dominant component in the universe (Einasto et al., 1974; Ostriker et al., 1974). Cosmological measurements give further evidence along this line, from the fact that the total matter density Ω_m needed to accomodate measurements of the expansion and geometry of the universe is much larger than the baryon density Ω_b obtained from Big Bang nucleosynthesis (see Fig. 1.4) or other cosmological observables. Finally, weak lensing measurements are directly sensitive to all forms of matter, including dark matter. Recent lensing studies of interacting clusters, such as the *Bullet cluster* (Clowe et al., 2006), provide concluding evindence for the need of a dark matter component, separated from the baryonic matter.

There are several candidate explanations for dark matter, mostly in the form of new types of particles that appear in Particle Physics theories beyond the Standard Model. Several projects are aiming at the detection of this kind of particles, either in astrophysical phenomena, or at large particle accelerators, such as the Large Hadron Collider (LHC) (see e.g. Feng, 2010, for a review).

The cosmological constant term was first introduced by Einstein himself, in order to reconcile his GR theory with a model for a static universe.[5] Once the expansion of the universe was established, this term was dropped for reasons of simplicity. However, it was never fully forgotten, as it was an alternative that could solve some inconsistencies between different observational results. The main example for this reasoning is that Λ could solve the discrepancy between the age of the universe implied by measurements of H_0, and the individual ages derived for some old objects such as globular clusters. Therefore, the cosmological constant term was present in the cosmological literature, and there was strong (although

---

[5] This is seen by setting $\Lambda = \frac{3k}{a^2} - 8\pi G\rho$ in equation (1.3). However, such a model for a static universe is unstable, as seen from equation (1.4).

circumstantial) evidence for $\Lambda \neq 0$ by the mid-1990s (see e.g. Carroll et al., 1992, for a review). The general acceptance of this term came after two groups used the distance — redshift relation for a set of high redshift Type Ia supernovae to show that the expansion of the universe was accelerating (Riess et al., 1998; Perlmutter et al., 1999). They obtained $q_0 < 0$, which implies the need for $\Lambda > 0$ (see equation 1.5). In recent years, complementary probes have confirmed this need for a positive cosmological constant, and have measured its value with reasonable precision (see Fig. 1.5). However, we should note that, unlike dark matter, all evidence for Λ comes from cosmological observations, as it would not have any measurable effect at smaller scales, such as those typical, e.g., of clusters or galaxies.

We can interpret this Λ term simply as an extra parameter of the theory, or an integration constant, which modifies the way in which the metric is affected by the energy-moment tensor (McVittie, 1956). Alternatively, as explained above, it can be also interpreted as a contribution to the energy content coming from a perfect fluid with negative pressure, $p_{DE} = w_{DE}\rho_{DE}$, with $w_{DE} = -1$, which at least mathematically is equivalent to a cosmological constant. This fluid is usually known as dark energy. In the framework of quantum field theory, a term like this can be interpreted as a vacuum energy. However, when estimating the expected contribution from the well established theories describing particle physics, one obtains a discrepancy of at least ~ 60 orders of magnitude between the expected vacuum energy density and the observed Ω_Λ (Carroll, 2001)[6]. A wealth of alternatives have been proposed to explain dark energy, mainly in the form of new scalar fields giving a equation of state $w_{DE} \neq -1$ (collectively known as 'quintessence'), or modifications of General Relativity. A radical alternative explanation, which somehow contradicts the *cosmological principle*, assumes that we happen to be located at the centre of a very large-scale under-dense region. This could, in principle, explain the observations even for $\Lambda = 0$ (Tomita, 2001; Alnes et al., 2006; Garcia-Bellido and Haugbølle, 2008). For reviews on the observational evidence in favour of dark energy, and the possible explanations, see e.g. Peebles and Ratra (2003); Frieman et al. (2008).

Trying to find a satisfactory explanation for both dark matter and dark energy is one of the hottest topics on astronomy and physics research today, which combines large observational efforts with theoretical developments. However, some argue that the need for dark matter and dark energy may be a sign that a fundamental change is needed in our cosmological paradigm (Martínez and Trimble, 2009; Horvath, 2009).

---

[6] This discrepancy is often quoted as ~ 120 orders of magnitude, from the extrapolation of these known theories to the Planck scale.

1.1.2 Inhomogeneities in the universe

From the first maps that revealed the galaxy distribution in space, it was clear that matter is not uniformly distributed in the universe. On the contrary, galaxies are highly clustered, forming a *Cosmic Web* composed of a variety of structures, such as groups, clusters, filaments, walls and voids. Therefore, we need a way to describe the matter density field, and an explanation for the formation of these structures in the context of the ΛCDM model.

Cosmological random fields

Although, according to the Cosmological Principle, the universe is statistically homogeneous at large scales, this does not mean that their properties are exactly uniform in space. We therefore need a way to describe those physical quantities, such as density or temperature, that are continuous functions of the spatial coordinates and time. However, we are not interested in the specific details of the observed field, but in an statistical description of them. That is, we consider the observed field as a single random realisation from an ideal ensemble of fields. Any cosmological model we may consider will give a description of this "ensemble of universes", but not of the specific universe we observe.[7]

A difference between the cosmological case and the typical situation in which we use statistical ensembles is that we only have one universe to observe. As all statistics of the random fields (e.g. variance) are defined as averages over the ensemble, this would mean that we can not measure them from our single realisation. The way to circumvent this issue is to assume that averaging over widely separated parts of space in a single realisation is equivalent to averaging over the ensemble. Random fields which fulfil this condition are called "ergodic". We assume always that our cosmological fields are ergodic, and therefore we refer without distinction to both kinds of averages.

Here, we are interested in describing the density field $\rho(\mathbf{x}, t)$. It is convenient to use the dimensionless density perturbation field $\delta(\mathbf{x}, t)$ defined as

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \,, \tag{1.6}$$

where $\bar{\rho(t)}$ is the mean density. The first moment of the field $\delta(\mathbf{x})$ is given by the mean, which in this case is zero by the definition above,

$$\langle \delta(\mathbf{x}) \rangle = 0 \,.$$

The second moment is given by the two-point correlation function

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle \ .$$

---

[7] Note that this is not in contradiction with having a causal cosmological model. In principle, we could calculate the details of any specific model if we knew completely the initial conditions. However, this does not happen in this case, as we only have a statistical description of them.

As we assume that the density field is homogeneous and isotropic, the correlation function $\xi(r) = \xi(\mathbf{x}_1, \mathbf{x}_2)$ in this case depends only on the distance between the points, $r \equiv |\mathbf{x}_2 - \mathbf{x}_1|$. Higher-order moments can be defined in an analogous way, but we will not study them here.

For many applications it is useful to work in Fourier space. We can decompose the field $\delta(\mathbf{x})$ in its Fourier components $\delta_{\mathbf{k}}$ as

$$\delta(\mathbf{x}) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \delta_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} \,.$$

The analogue to the correlation function in Fourier space is the power spectrum $P(\mathbf{k})$ defined by

$$\langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle = (2\pi)^3 \delta_D(\mathbf{k}_1 + \mathbf{k}_2) P(\mathbf{k}_1) \,, \tag{1.7}$$

where $\delta_D(\cdot)$ is Dirac's delta function. As the field is isotropic, the power spectrum $P(k) = P(\mathbf{k})$ only depends on the norm $k \equiv |\mathbf{k}|$.

The power spectrum and correlation function also form a Fourier pair in 3D which, assuming homogeneity and isotropy, is defined as

$$\xi(r) = 4\pi \int_0^\infty P(k) \frac{\sin(kr)}{kr} \frac{k^2 \mathrm{d}k}{(2\pi)^3} \,.$$

Therefore, $\xi(r)$ and $P(k)$ contain the same information about the density field $\delta(\mathbf{x})$. However, the fact that Fourier modes are statistically independent (equation 1.7) makes the study of the field in Fourier space much easier *a priori*.

A special case of random fields are Gaussian fields. In this case, the probability distribution function for the field at N points in space is given by a multivariate Gaussian distribution. This means that the statistical properties of the field are fully described by the second-order moments $\xi(r)$ or $P(k)$. This is an important case in cosmology, as it is assumed to be a very good approximation of the initial density perturbations, and also of the present-day perturbations at very large scales, where $|\delta| \ll 1$. In the general case, however, second-order moments are not enough to describe completely the field, and higher-order moments are needed.

Dynamics of structure formation

As explained above, despite the fact that the universe is statistically homogeneous, we observe a large variety of structures in it at very different scales. In the standard cosmological paradigm, the large-scale structure we see today has grown by the effect of gravity from initial density fluctuations, due to the instability of the Friedmann models. In this picture, we start with a Gaussian field of perturbations with very small amplitude ($|\delta| \ll 1$), and with a power spectrum which is close to scale-invariant, $P_i(k) \propto k^n$, with $n \simeq 1$. These initial perturbations are adiabatic, meaning that they affect in the same way all the species. These tiny fluctuations grow gravitationally, and are responsible both for the inhomogeneities (of order $\sim 10^{-5}$) in the CMB at $z \sim 1100$, and for the different structures we observe in the present-day universe.

We briefly show here how this gravitational growth works in the linear approximation, for a full review of the perturbation theory in this context, see Bernardeau et al. (2002). It is valid when the perturbations are small ($\delta \ll 1$). The approximation is valid, therefore, at early times or, nowadays, at very large scales. If we study scales which are much smaller than the radius of curvature of the universe, we do not need to use a general relativistic approach, and we can use a Newtonian approach.

The evolution of the matter distribution is then described by the usual equations for fluid motion,

$$\begin{aligned}
&\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla \cdot (1+\delta)\,\mathbf{u} = 0 \\
&\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}\left[(\mathbf{u}\cdot\nabla)\,\mathbf{u} + \nabla\phi\right] = 0 \\
&\nabla^2\phi = 4\pi G\bar{\rho}a^2\delta\,.
\end{aligned}$$

Here, ϕ is the gravitational potential, and $\mathbf{u}$ is the peculiar velocity field. The peculiar velocity of a particle is its total velocity minus that due to the Hubble flow. The first of these equations is the continuity equation, related to conservation of energy. The second is Euler's equation of motion of the fluid, and the last one is Poisson's equation. Here, we have neglected the effect of pressure, therefore we describe the evolution of the perturbations for the case of non-relativistic matter alone.

In the linear approximation, these equations can be combined, keeping only the first term in perturbation theory, to give

$$\ddot{\delta} + 2H\dot{\delta} - 4\pi G\bar{\rho}\delta = 0\,.$$

This equation does not depend explicitly on the spatial coordinates, hence the evolution of the perturbations is local. This is a second order differential equation in δ, so it has two partial solutions, $D_+(t)$ (the "growing mode") and $D_-(t)$ (the "decaying mode"), which can be combined to give the general solution

$$\delta(\mathbf{x},t) = A(\mathbf{x})D_+(t) + B(\mathbf{x})D_-(t)\,,$$

where $A(\mathbf{x})$ and $B(\mathbf{x})$ are initial conditions. $D_-(t)$ is negligible when we study evolution over large periods of time, so we are left with $\delta(t) \propto D_+(t)$. The actual calculation of $D_+(t)$ involves a numerical integration over the expansion factor a, but a good approximation is given by (Carroll et al., 1992)

$$D_+(z) \simeq \frac{5}{2(1+z)}\frac{\Omega_m(z)}{\Omega_m(z)^{4/7} - \Omega_\Lambda(z) + \left(1 - \frac{\Omega_m(z)}{2}\right)\left(1 + \frac{\Omega_\Lambda}{70}\right)}\,.$$

We have therefore a description of the initial perturbations, and of the basic process which drives their evolution. This means that we can make a prediction for the present-day linear theory matter power spectrum

$$P_{\rm lin}(k,z) = D_+^2(z)T^2(k,z)P_i(k) = AD_+^2(z)T^2(k,z)k^n\,,$$

where A is a constant that sets the global normalisation of the spectrum, and $T(k,z)$ is the "transfer function". Instead of A, the usual parameter used to characterise the normalisation of the spectrum is the rms fluctuations of mass in a sphere of radius $R = 8\,h^{-1}\,\text{Mpc}$, σ_8, defined as

$$\sigma_8^2 = 4\pi \int_0^\infty \widetilde{W}_{TH}^2(k, R = 8\,h^{-1}\,\text{Mpc}) P_{\rm lin}(k) \frac{k^2\text{d}k}{(2\pi)^3}\,,$$

where $\widetilde{W}_{TH}(k,R)$ is the Fourier transform of a spherical top-hat filter of radius R. The accepted value for this parameter from observations is $\sigma_8 \sim 0.8$.

The transfer function describes all the physical effects that we have not included in this simple model, mainly due to the relativistic species present. By definition, $T(k) = 1$ for modes k that evolve exactly as described above. The main effect comes from the fact that, during the radiation dominated era, the pressure of photons stops the gravitational growth described above. However, this only happens for perturbation on scales smaller than the horizon at that time, while larger scale modes do grow as described above. The overall result is that $P_{\rm lin}(k)$ shows a wide bump, with the location of the maximum roughly indicating the scale of the horizon at the time of radiation–matter equality.

Other, less prominent, features in $T(k)$ come from the effect of baryons. One of such features are baryon acoustic oscillations (BAO). These are a series of damping oscillations that appear in $P_{\rm lin}(k)$ due to acoustic waves that travelled in the baryon–photon plasma prior to recombination. They also imprint a characteristic scale in the matter power spectrum, which corresponds to the sound horizon at recombination, and which can be used to study the geometry of the universe. We describe this feature in more detail in Section 5.1.

A precise calculation of the transfer function $T(k)$ requires solving numerically the Boltzmann equations for the different species. Several codes are available for this purpose, being the most used CMBFAST (Seljak and Zaldarriaga, 1996) and CAMB (Lewis et al., 2000). There also exist several approximate fitting functions to $T(k)$ which are accurate enough for many applications (Bond and Efstathiou, 1984; Holtzman, 1989; Eisenstein and Hu, 1998, 1999). In Fig. 1.6, we plot the $T(k)$ and $P(k)$ obtained using CAMB for standard values of the cosmological parameters, showing the different features described above.

This linear model breaks down at small scales (large values of k) when the density fluctuations grow larger and can not be treated as small perturbations anymore. Several methods exist to study the non-linear evolution in this case

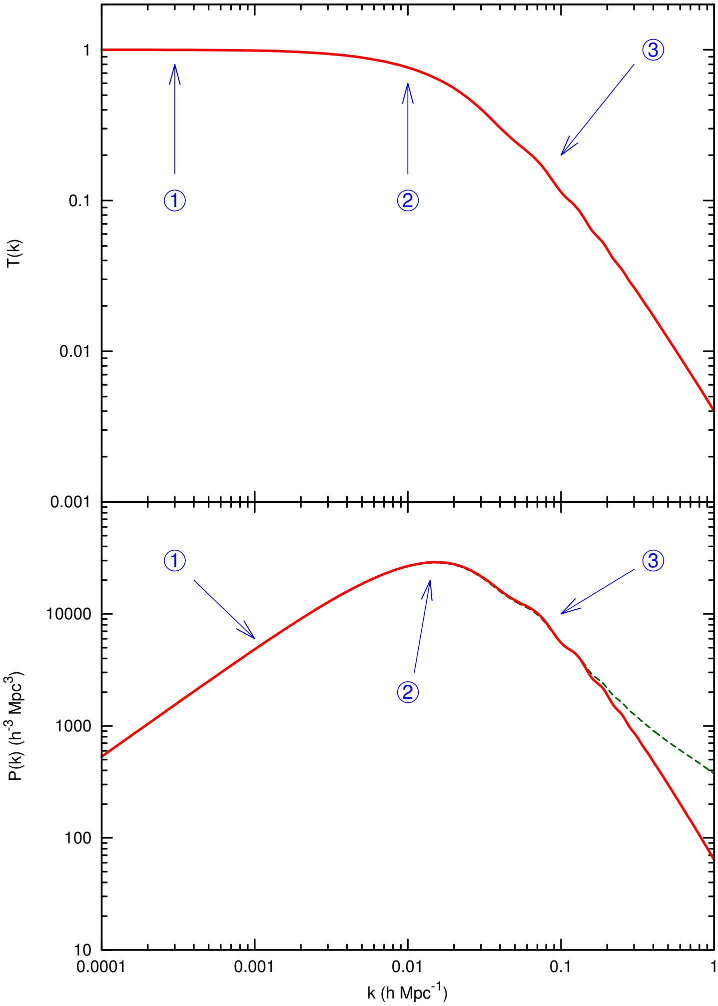


Figure 1.6: Transfer function $T(k)$ (top) and linear-theory matter power spectrum $P_{\rm lin}(k)$ (bottom) at $z = 0$ obtained using the CAMB software. They are calculated for the values of the parameters in Komatsu et al. (2011). The dashed green line corresponds to the non-linear power spectrum, using the HALOFIT approximation. We can see the behaviour at different scales described in the text. (1): The largest scales (small k) did not have time to enter the horizon during the radiation domination era. They are therefore not affected by pressure effects, so $T(k) \simeq 1$, and the power spectrum corresponds to the initial one (apart from the normalisation), $P(k) \propto k^n \simeq k$. (2): At the scales corresponding to the horizon scale at matter–radiation equality we see a change in this behaviour, as $T(k) < 1$, and there is a wide bump in $P(k)$ at roughly that scale. (3): At smaller scales, we see one of the effects of baryons: a series of oscillations (BAO) in both $T(k)$ and $P(k)$.

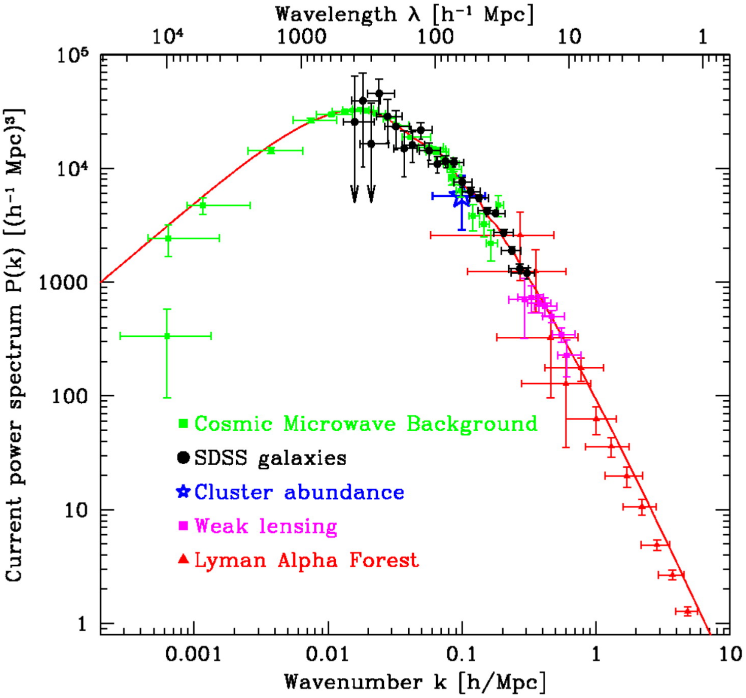


Figure 1.7: Matter power spectrum estimated through different cosmological observables. Figure from Tegmark et al. (2004). See that reference for details about the different data used, and the way in which they were converted into a measure of $P(k)$.

(see e.g. Hamilton et al., 1991; Crocce and Scoccimarro, 2006). One of the most successful of such methods is based on the halo model (Seljak, 2000; Peacock and Smith, 2000; Cooray and Sheth, 2002). In this model, the matter density distribution is decomposed into a population of massive virialised dark matter haloes. The study of the properties of these haloes can be related to the non-linear matter power spectrum. This approach was used by Smith et al. (2003) to produce the Halofit approximation. Halofit is a fitting formula, calibrated with N-body simulations, to obtain accurate "non-linear corrections" to the power spectrum $P(k)$. The resulting non-linear power spectrum is also shown in Fig. 1.6 by the dashed line.

The density field have been studied using several, complementary, cosmological observables. An illustration is given in Fig. 1.7, reproduced from Tegmark et al. (2004). The figure shows the measurements from different observations transformed into constraints on the matter power spectrum. At the largest scales (low k) we get the best constraints from the CMB anisotropies. At the smallest scales, we have good constraints from the study of absorption lines in the spectra of quasars (the Ly-α forest), and from the study of the shear in galaxies' shapes produced by weak gravitational lensing.

In this thesis, we focus on the study of the matter density field through the spatial distribution of galaxies. This is, at least conceptually, the most direct method, as we are studying the distribution of the part of the matter we can directly observe (luminous galaxies). However, there are many caveats in this simple approach, and we briefly review them below in Section 1.2. The range of scales that can be studied through galaxy clustering varies greatly depending on the characteristics of the survey used (specially its volume and sampling density).

1.2 Statistical analysis of galaxy clustering

When performing observations in practice through extra-galactic surveys, we do not study directly the matter density field, but we can obtain information only from the distribution of the observed positions of galaxies. In this case, when trying to obtain cosmological information we must address two important issues. On one side, we need a model to describe the relation between the galaxy distribution and the matter density field or, in other words, understand how well "galaxies trace mass". On the other side, the galaxy distribution can be considered as a realisation of a random point process (or field), and we should use the statistical machinery associated to this kind of processes, instead of the continuous random field we used to describe the matter distribution.

For a complete review of spatial statistics, see Stoyan et al. (1995); Illian et al. (2008), and for their applications in cosmology, Martínez and Saar (2002). The random point process consists on a random sequence of N discrete points, $\Phi = \{\mathbf{x}_i\}_{i=1}^{N}$, corresponding to the observed positions of galaxies, inside a finite region of space, or "window", $W \subset \Re^3$. The window W is typically defined by the characteristics –depth, sky area covered– of the survey. As usual, we consider that the point process in this case is homogeneous and isotropic.

The two-point correlation function $\xi(\mathbf{x}_1, \mathbf{x}_2)$ for a point process can be defined as follows. Let us consider two infinitesimally small spheres of volumes $\mathrm{d}V_1$ and $\mathrm{d}V_2$ centred at $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively. Then, the probability that in each of those spheres lies a point belonging to the point process is

$$\mathrm{d}P_{12} = \lambda_2(\mathbf{x}_1, \mathbf{x}_2)\mathrm{d}V_1\mathrm{d}V_2 = \lambda^2 \left[1 + \xi(\mathbf{x}_1, \mathbf{x}_2)\right] \mathrm{d}V_1\mathrm{d}V_2$$

where $\lambda_2(\mathbf{x}_1, \mathbf{x}_2)$ is the 'second order intensity function' of the process, and λ is its number density. For a homogeneous and isotropic process, the correlation function depends only on the distance between the two points, $r \equiv |\mathbf{x}_2 - \mathbf{x}_1|$.

In this way, the correlation function $\xi(\mathbf{x}_1, \mathbf{x}_2)$ can be understood as a measure of clustering in excess ($\xi > 0$) or in defect ($\xi < 0$) with respect to a homogeneous

Poisson process. This is a process for which the number of points in any region A with volume $V(A)$ follows a Poisson distribution with mean $\lambda \cdot V(A)$, and the number of points in any two non-overlapping regions of space are independent. This means that the points of the process are uniformly distributed inside the window W, independently from each other. Poisson samples are used for the estimation of the correlation function, as explained below.[8]

Several point process models have been used to attempt to explain the observed distribution of galaxies, starting with the model of Neyman and Scott (1952), based on the random superposition of galaxy groups. After the first measurements of the galaxy correlation function (Totsuji and Kihara, 1969; Peebles, 1974) showed its power-law behaviour, several models based on the use of fractals appeared (e.g. Mandelbrot, 1975; Soneira and Peebles, 1978). Although the use of fractal models at certain scales can provide useful methods for the analysis, it has been shown that these models can not be extrapolated to large scales (see e.g. Martínez, 1999). However, the power-law model is still a good and widely used approximation for the behaviour of the galaxy correlation function for limited ranges of scales (Tucker et al., 1997; Jing et al., 1998; Zehavi et al., 2002; Hawkins et al., 2003; Zehavi et al., 2005a; Coil et al., 2006). Other statistical models have been used, for examples those based on the use of Voronoi tessellations (Icke and van de Weygaert, 1987; van de Weygaert and Icke, 1989), or on a thermodynamical description of clustering (Saslaw and Sheth, 1993; Sheth and Saslaw, 1996). For a review of these models, see Jones et al. (2005); Martínez and Saar (2002).

In the ΛCDM paradigm, the galaxy distribution is related to the total matter distribution using a Cox process model. In this model, the expected galaxy density $\rho_g(\mathbf{x})$ at each point is described as a realisation of a continuous random field. Then, the galaxy point process is a inhomogeneous Poisson process which follows this density field. This means that the number of galaxies inside an infinitesimal region at $\mathbf{x}$ with volume $\mathrm{d}V$ is random variable distributed according to a Poisson distribution function with mean $\rho_g(\mathbf{x}) \cdot \mathrm{d}V$. Therefore, this model is doubly stochastic. On the one side because of the random nature of the galaxy density field, and on the other side because of the Poisson sampling of this field.

The galaxy density field ρ_g is obviously related to the total matter density field ρ described in Section 1.1.2. In fact, the relation between these two fields encodes important information about galaxy formation and evolution. A simple, widely used, approximation is to assume a global linear relation between these

---

[8] The homogeneous Poisson process is often confused with the binomial random field. The difference lies on the fact that, for a binomial field the total number of points N inside the window W is fixed, while for the Poisson process it is itself a random variable (following a Poisson distribution). In most practical applications, where N is large, this small difference has no influence. In cosmology, usually the name "Poisson sample" is used to name both concepts interchangeably.

fields through the *bias* parameter b, defined as

$$\delta_g(\mathbf{x}) = b\delta(\mathbf{x}) \,, \tag{1.8}$$

where $\delta(\mathbf{x})$ is the perturbation field defined in equation (1.6), and $\delta_g(\mathbf{x})$ is defined analogously with respect to the ρ_g field. In the "un-biased" case $b = 1$, we say that *galaxies trace mass*, and in this case the expected value of the galaxy correlation function $\xi_g(r)$ coincides with that of the matter density field $\xi_m(r)$. In the general case $b \neq 1$, we get $\xi_g(r) = b^2\xi_m(r)$, and similar relations for higher-order correlations. More general bias models exist in which the bias parameter depends on position or scale (see e.g. Fry and Gaztanaga, 1993; Mo et al., 1997; Matsubara, 1999).

The presence of this bias parameter can be understood in a natural way if we assume that galaxies form at the peaks, or local maxima, of the matter density field. In this case, the bias for the peaks can be calculated analytically, at least if we assume a Gaussian random field (Kaiser, 1984; Bardeen et al., 1986). In the context of the halo model introduced above (see Section 1.1.2), the bias parameter for dark matter haloes can be modelled, and depends on the properties of the halo such as its mass (see, e.g., Mo and White, 1996, 2002). Studying the clustering of a certain type of galaxies gives therefore information on the characteristics of the haloes that host them. A refinement of this model is the study of the Halo Occupation Distribution (HOD). In this framework, one not only studies the clustering of haloes, but also the spatial distribution of galaxies inside them (Scoccimarro et al., 2001; Berlind and Weinberg, 2002; Kravtsov et al., 2004).

As the bias parameter encodes information about the galaxy formation and evolution process, it is logical to expect that it will be different for different galaxy populations. In other words, the clustering properties of galaxies should depend on some of their intrinsic properties, such as their mass, and should evolve with time. This coincides with the observations. This phenomenon, known as *galaxy segregation*, can be observed when studying the dependence of clustering on different observables. In general, luminous, red, elliptical galaxies are more strongly clustered than faint, blue, spiral ones. We study the issue of galaxy segregation, its evolution, and appropriate methods to analyse it in Chapters 3 and 4 of this thesis.

In this thesis, specifically in Chapter 3, we use the simple linear bias model defined in equation (1.8). This is a valid approximation for our purposes, given that we do not attempt a detailed modelling of our data, and we restrict our analysis to a limited range of scales. However, when a detailed model for the correlation function is needed, possible deviations from this simple bias model should be taken into account. A possible problem is, for example, that a scale-dependent bias could be masking variations in the model due to changes in the cosmological parameters.

We focus in this Section in the estimation methods and properties of the two-point correlation function $\xi(r)$, which is the tool on which we base the work done in Chapters 2, 3, and 5. In Chapter 4 we will explain in more detail some methods based on mark statistics which, although closely related to $\xi(r)$, are tailored for the study of galaxy segregation. Finally, in Chapter 6 we will introduce a different method for the analysis of the galaxy distribution, based on the use of wavelets.

1.2.1 Estimation of the correlation function

One of the ways to analyse the clustering properties of a given galaxy sample is to estimate the value of the correlation function. We describe here the methods used to make this estimation based on a sample obtained from a galaxy redshift survey. In this case, once the galaxy sample to study is properly defined, we start from a catalogue containing the three-dimensional position of each galaxy inside the survey window. There are several issues associated to the fact that the galaxy positions are calculated from their redshifts, and to the selection effects inherent to any survey, which we discuss later.

First, we assume that the positions of the N_D galaxies in the sample are known exactly. All methods for the estimation of $\xi(r)$ are based on some type of average on the counts of neighbours at different scales. The main problem appears from the fact that, close to the boundaries of the window considered, these counts are under-estimated.

There are different approaches to correct for this border effect. Most of the methods used in cosmology are based on comparing the neighbour counts in the data sample to the neighbour counts for an auxiliary sample generated as a homogeneous Poisson process in the same survey window. These methods then estimate $\xi(r)$ by a combination of the quantities $DD(r)$, $RR(r)$ and $DR(r)$. These are the number of pairs at a given scale r between points in the data sample (DD), between points in the auxiliary Poisson sample (RR), and for crossed pairs between points in both samples (DR). Different such estimators were proposed by Peebles and Hauser (1974); Davis and Peebles (1983); Hamilton (1993); Landy and Szalay (1993), and are given by

$$\begin{aligned}
\hat{\xi}_{PH}(r) &= \left(\frac{N_R}{N_D}\right)^2 \frac{DD(r)}{RR(r)} - 1 \\
\hat{\xi}_{DP}(r) &= \frac{N_R}{N_D}\frac{DD(r)}{DR(r)} - 1 \\
\hat{\xi}_{HAM}(r) &= \frac{DD(r)\cdot RR(r)}{[DR(r)]^2} - 1 \\
\hat{\xi}_{LS}(r) &= 1 + \left(\frac{N_R}{N_D}\right)^2 \frac{DD(r)}{RR(r)} - 2\frac{N_R}{N_D}\frac{DR(r)}{RR(r)} ,
\end{aligned}$$

where N_R is the number of points in the auxiliary Poisson sample. The calculation of the pair counts (and hence of $\hat{\xi}$) can be done either in discrete bins in r, or continuously in r by the use of a finite kernel. The differences between both approaches are minimal when the bin or kernel widths are similar.

Several works compared the performance, in terms of bias and variance, of these estimators in a cosmological context (Pons-Bordería et al., 1999; Kerscher, 1999; Labatie et al., 2011). The general conclusion is that, as first noted by Hamilton (1993), the bias is lower for $\hat{\xi}_{HAM}(r)$ and $\hat{\xi}_{LS}(r)$, specially at large scales, and they are therefore the usual estimators of choice. Through this thesis (Chapters 2, 3, and 5) we use the estimator of Landy and Szalay (1993) for the estimation of $\xi(r)$.

An alternative is to explicitly calculate the correction factors needed to avoid the border effects. Examples of this approach are the estimators proposed by Rivolo (1986) and Stoyan and Stoyan (1994), although they are rarely used in cosmology. In Chapter 4 we derive the estimators for other two-point statistics from the estimator of Stoyan and Stoyan (1994), and we explain there its properties in more detail.

When dealing with data from a real survey, however, the situation is not so straightforward, due to the different selection effects. On the one side, the angular selection function can be complicated due to the design of the survey, and to different observational issues. This not only defines the window covered by the survey, but could also imply a variable completeness for different sky regions. On the other side, the selection function can change greatly with distance, depending on the way in which the galaxy sample is selected. The latter issue can be addressed by making an appropriate selection in absolute magnitude and redshift to produce a "volume limited" sample, although this usually implies losing a large amount of valid data. The way to address all these selection effects is to introduce them into the calculation through the auxiliary Poisson catalogue. This means that the auxiliary catalogue used is a *inhomogeneous* Poisson catalogue, generated using a density field in the window which follows the selection function of the sample. Therefore, the precise modelling of this selection function is crucial, as any inaccuracy in it could introduce a bias in the estimation of $\xi(r)$.

1.2.2 Real and redshift space

In order to estimate $\xi(r)$, we need to know the three-dimensional position of each object in the sample. However, in cosmology a problem appears because this position is determined in a different way for different directions in space. In the direction perpendicular to the line of sight, the position is obtained from the angular position of the object, which is not an issue.

However, in the line-of-sight direction, the position is obtained from the redshift z of the object. In the case of spectroscopic surveys, the redshift is measured with great accuracy (typically, $\sigma(z) \sim 10^{-4}$). The problem arises because the measured redshift is a combination of two effects: the cosmological redshift due to the Hubble flow, which in principle is directly linked to the position, and any additional peculiar velocity of the object. This has a direct effect on the measured

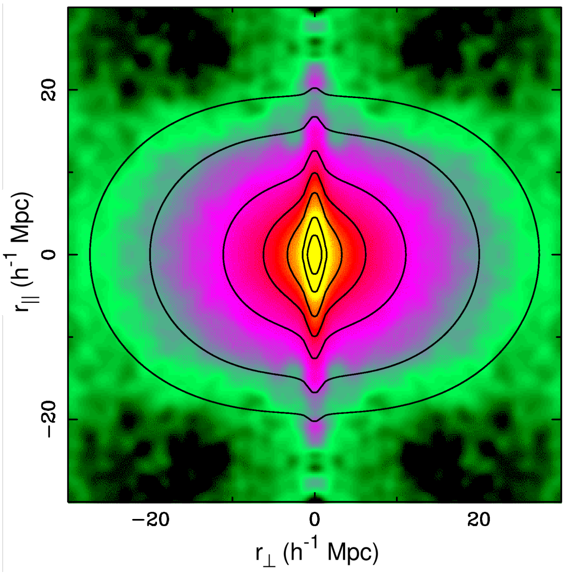


Figure 1.8: Two-dimensional correlation function for the 2dFGRS survey, expressed as a function of the line of sight ($r_{\parallel}$) and transverse ($r_{\perp}$) separations, showing the effects of redshift space distortions. The cigar-like shape at $r_{\perp} \simeq 0$ is due to random peculiar velocities in virialised groups, while the oval-like shape at larger transverse separations is the signature of coherent infall. Figure from Peacock et al. (2001).

correlation function in “redshift space”.

At small scales, the apparent correlations along the line of sight are partially erased, due to the random peculiar velocities of objects in virialised groups or clusters. However, at large scales, the effect is the opposite, as the peculiar velocities of objects that are falling towards collapsing structures are coherent.

Overall, these distortions imply that the redshift-space galaxy distribution is no longer isotropic. It is useful, therefore, to consider the two-dimensional correlation function $\xi(r_{\parallel}, r_{\perp})$, where we separated the dependence on scale into the line of sight separation $r_{\parallel}$, and the transverse separation $r_{\perp}$ [9]. Fig. 1.8 shows the two-dimensional correlation function for the Two-degree Field Galaxy Redshift Survey (2dFGRS) obtained by Peacock et al. (2001). These redshift space distortions on the correlation function can be modelled (Kaiser, 1987; Hamilton, 1998), and used to constrain the cosmological parameters, and in particular Ω_m (Peacock et al., 2001; Hawkins et al., 2003; Cabré and Gaztañaga, 2009). The

---

[9] In the literature, the symbol π is typically used for $r_{\parallel}$, and σ or r_p for $r_{\perp}$. We chose this notation in this thesis to avoid possible confusion with other symbols used through the text.

real space correlations can still be studied, when one is not interested on these distortions, by projecting the correlation function into the transverse direction (Davis and Peebles, 1983).

In the case of traditional photometric surveys, the error in the determination of redshifts is much larger, so most information on radial correlations is lost. However, when the number of filters used is large, this uncertainty can be reduced to $\sigma(z) \sim 0.01(1+z)$. Chapter 2 of this thesis is devoted to the study of real space correlations in this case using the projected correlation function method, which we describe in detail in Section 2.1. In Chapter 3 we exploit this method for the analysis of data from the ALHAMBRA Survey.

1.3 Galaxy redshift surveys

Galaxy surveys are one of the main tools of observational cosmology. The basic idea is to collect observational information about all galaxies (following certain selection criteria) in a given region of the sky. The main selection criterion is usually the apparent magnitude in a given band, but there can be more complicated criteria in some cases. The aim of this approach is to study statistically the properties of the different galaxy populations, and their distribution in space. In this way, we obtain information on the galaxy formation and evolution processes, and also on the properties of the cosmological density field, as described above.

Early galaxy surveys, such as those of the Lick (Shane and Wirtanen, 1967) and Palomar (Abell, 1958; Zwicky et al., 1961) observatories, gave basically the angular position and flux for each galaxy, so that only two-dimensional information on the positions was obtained. In the early 1980's, however, the first surveys in which spectroscopy was used to systematically measure galaxy redshifts appeared. The use of spectroscopy not only provides accurate three-dimensional positions through the measurement of the redshift, but also accurate information about several properties of each galaxy. The first of these surveys were those produced at the Center for Astrophysics (CfA, Huchra et al., 1983; de Lapparent et al., 1986), and several followed in the next years, such as the Southern Sky Redshift Survey (SSRS, da Costa et al., 1991), the Las Campanas Redshift Survey (LCRS, Shectman et al., 1996), or the Point Source Catalogue survey (PSCz Saunders et al., 2000), see Jones et al. (2005) for a review.

Currently, the largest completed redshift surveys are the 2dFGRS (Colless et al., 2001), and the Sloan Digital Sky Survey (SDSS, York et al., 2000), which both cover a large fraction of the sky to a moderate depth. Fig. 1.9 shows two slices drawn from both the 2dFGRS and SDSS, compared to a slice from the CfA-II survey. Other kind of surveys, sometimes known as “pencil-beam” surveys, cover a much smaller area, but are much deeper. In this way, they are able to cover a large range in redshift, allowing for detailed studies of cosmic evolution. Recent examples of the latter kind of surveys are the Deep Extragalactic Evolutionary

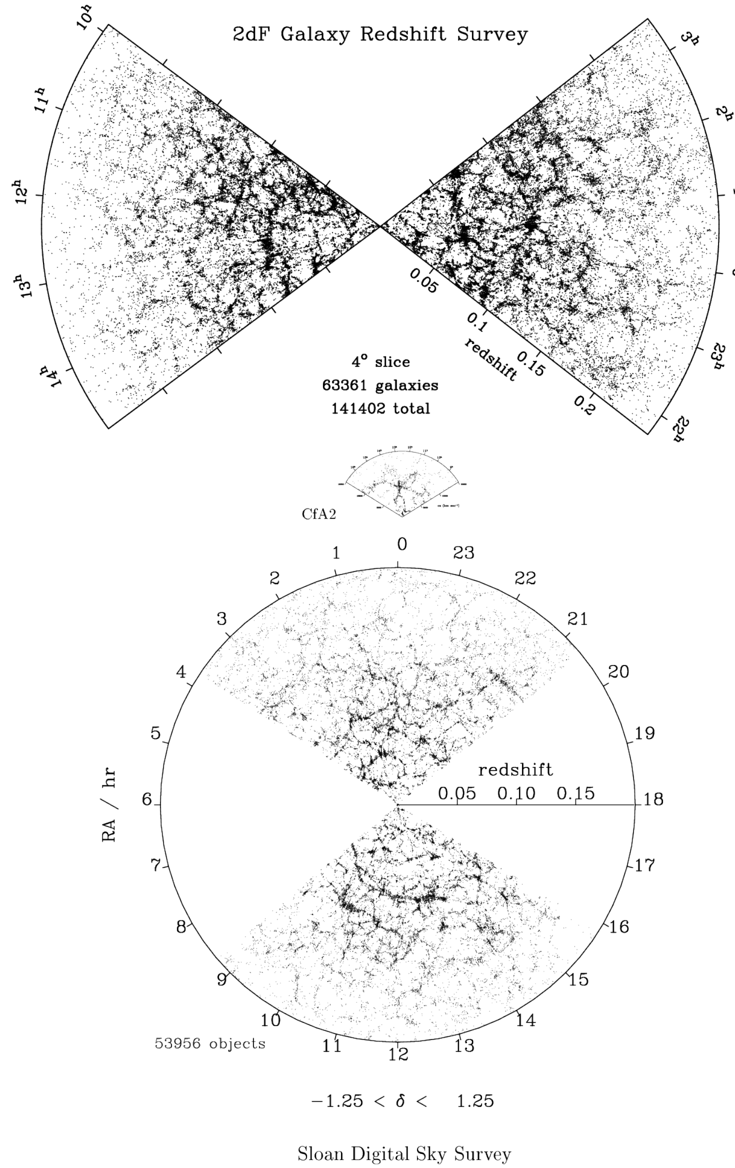


Figure 1.9: Slices drawn from three different galaxy redshift surveys: the CfA2, 2dFGRS, and SDSS. The figure gives an idea of the different depths of the surveys. The overall characteristics of the galaxy distribution, forming a *Cosmic web*, can be appreciated, as well as the effects of redshift space distortions. Figure from Jones et al. (2005).

Probe (DEEP2, Davis et al., 2003), and the VIMOS VLT Deep Survey (VVDS, Le Fèvre et al., 2005).

In this thesis, we use publicly available data from both the 2dFGRS and the SDSS (in Chapters 4, 5, and 6). The 2dFGRS survey[10] was carried out in the period 1996 – 2003 using the 2dF multifiber spectrograph on the 3.9-meter Anglo-Australian Telescope. It covered a total of $\sim 1500\,\mathrm{deg}^2$ distributed in two strips, one in the Southern Galactic Hemisphere (covering $\sim 75^\circ \times 15^\circ$), and one in the Northern one (covering $\sim 75^\circ \times 7.5^\circ$). There are 100 additional fields (with a 2° diameter) scattered randomly around the southern strip. The region of sky observed by 2dFGRS is shown in Fig. 1.10a. Reliable redshifts were obtained for $221,414$ galaxies in total. Galaxies were selected for spectroscopy from an extended version of the Automatic Plate Measuring (APM) machine survey (Maddox et al., 1990), by an apparent magnitude cut in the b_J band, $b_J \leq 19.45$. The median redshift of the survey is $z_m = 0.11$, and contains galaxies with redshifts up to $z \sim 0.3$. The final data for the survey was released publicly (Colless et al., 2003), together with the masks describing the angular selection function of the survey, needed for clustering studies. This selection function takes into account several factors, such as bright stars or defects in the APM imaging, varying magnitude limits, or redshift completeness in the spectroscopic observations.

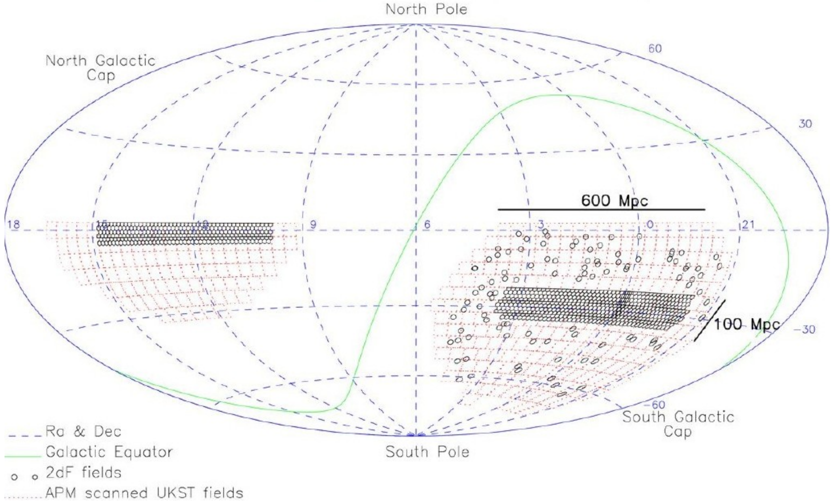
The SDSS[11] was started in 2000, and uses a dedicated 2.5-meter telescope at Apache Point Observatory (APO), equipped with a purpose-built imaging camera and two spectrographs. The SDSS project consists on several complementary surveys. We focus here on the survey following the original design, later known as 'Legacy'. This survey was completed in 2008, and the final data were made public in Data Release 7 (DR7, Abazajian et al., 2009). The full survey consists on a 5-band optical photometric survey, and a spectroscopic survey which uses the former for selecting the target galaxies. The area covered by the spectroscopy survey totals $\sim 8000\,\mathrm{deg}^2$, most of it corresponding to a large continuous region in the Northern Galactic Hemisphere ($\sim 7300\,\mathrm{deg}^2$), and the rest distributed in three strips in the South, as shown in Fig. 1.10b.

The survey consists of two different galaxy samples. The 'Main' galaxy sample is basically a magnitude limited sample, selected through a cut in the r magnitude, $r \leq 17.77$ (Strauss et al., 2002). The redshift coverage of this sample is similar to that of 2dFGRS, with a median redshift of $z_m \simeq 0.1$, and it contains reliable spectra for $\sim 560,000$ galaxies. A second sample targets Luminous Red Galaxies (LRG), selected through a set of colour and magnitude cuts (Eisenstein et al., 2001). This results in a sample which is approximately volume limited up to $z \simeq 0.37$, and extends further to $z \sim 0.5$. This sample contains spectra for $\sim 200,000$ galaxies (including those at low redshift also included in the 'Main' sample), and it is well suited for correlation studies at very large scales, due to the

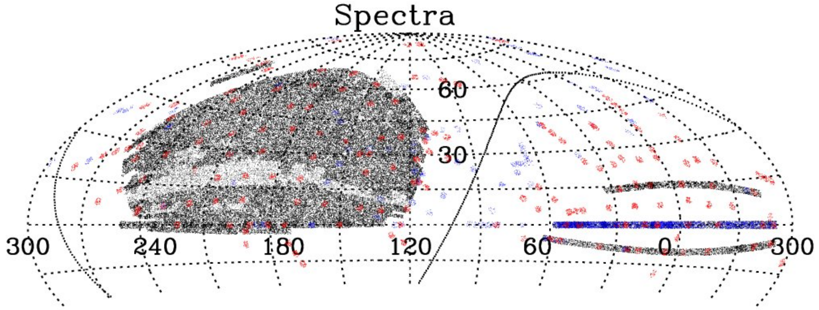
---

[10] http://www.mso.anu.edu.au/2dFGRS/

[11] http://www.sdss.org/



(a) 2dFGRS



(b) SDSS

Figure 1.10: Sky coverage for the two wide area surveys described in the text, in equatorial coordinates. (a) The fields for spectroscopy covered by the 2dFGRS are shown as black circles. This includes two wide strips, and additional fields distributed randomly in the South. Figure from Colless et al. (2003). (b) The area of the sky covered by spectroscopy in the DR7 of SDSS. The area covered by the Legacy survey corresponds to the contiguous area in the North, plus the three long strips in the South (shown by black and grey dots). Figure from Abazajian et al. (2009).

large volume covered (more than $1\,h^{-3}\,\mathrm{Gpc}^3$), and to the large bias characterising the LRG. The New York University Value Added Galaxy Catalogue (NYU-VAGC Blanton et al., 2005) provides as one of its products the detailed description of the angular selection function of SDSS. The SDSS-III project is implementing several additional surveys (Eisenstein et al., 2011), extending the SDSS original project, after some technical upgrades of the camera and spectrograph.

All these redshift surveys are based on the selection of objects via previous imaging observations, and targeting them for spectroscopy. An alternative method is the direct use of photometric redshifts, or "photo-z" (Baum, 1962; Koo, 1985), when photometry is available in several bands. There are several methods for photo-z determination (Fernández-Soto et al., 1999; Benítez, 2000; Bolzonella et al., 2000; Collister and Lahav, 2004), but the most widely used method is based on fitting the observed magnitudes and colours of each object to a set of template spectra (with varying redshift). In this approach, it is obvious that the spectral characterisation of each object, and hence its redshift, is much less accurate than when we use spectroscopy. However, the observational cost (in terms, e.g., of observation time needed per object) is much lower in this case. This allows this kind of surveys to cover larger areas, to improve the sampling completeness, or to increase the depth and/or number density of the survey. This results in an increase of the total number of objects available for the analysis, or in exploring larger parameter spaces (in terms of luminosity or redshift). Recent examples of wide area photometric redshift surveys are the five band survey which is part of SDSS, or the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS, Ilbert et al., 2006).

By increasing the number of photometric filters used, one can design a survey with characteristics in between those of spectroscopic and traditional photometric surveys (Hickson et al., 1994). The surveys Calar Alto Deep Imaging Survey (CADIS, Meisenheimer et al., 1998) and Classifying Objects by Medium-Band Observations in 17 filters (COMBO-17, Wolf et al., 2003) followed this approach, by combining observations in a set of filters of different widths.

The ALHAMBRA survey (Moles et al., 2008) is currently ongoing, and it intends to improve on this idea. It uses a set of 20 medium-band optical filters (plus three near infrared ones), which were specifically designed to maximise the output of the survey in terms of number of objects and photometric redshift accuracy (Benítez et al., 2009b). The final survey will cover $4\,\mathrm{deg}^2$ in 8 separate fields, and aims at obtaining a typical accuracy of $\Delta z \leq 0.015(1+z)$ for $\sim 3 \times 10^5$ objects at a median redshift of $z_m = 0.74$. Fig. 1.11 compares its characteristics (area, depth, number of objects and volume covered) to different completed surveys, in particular to those mentioned above.

In Chapter 2 of this thesis, we explore the possibilities of a survey with the redshift accuracy of ALHAMBRA for studies of the large-scale structure and, in particular, for the measurement of $\xi(r)$. Then, in Chapter 3, we use preliminary

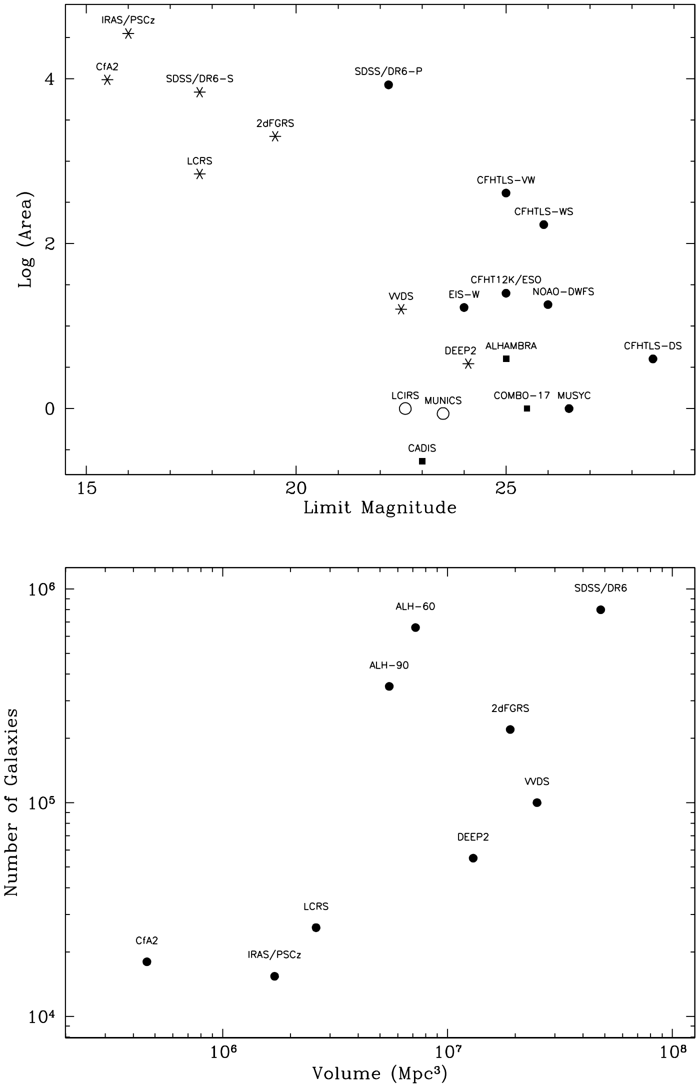


Figure 1.11: Comparison of the expected properties of the ALHAMBRA Survey to other completed surveys. The top panel shows the limit magnitude – area diagram, comparing ALHAMBRA to different spectroscopic (stars) and photometric (circles) surveys. The bottom panel shows the volume – number of galaxies diagram, making the comparison only to a set of spectroscopic surveys. Figure from Moles et al. (2008).

data from ALHAMBRA to study the clustering of galaxies as function of their luminosity, and its evolution. In Section 3.1 we describe in detail the characteristics of the survey, and its current status.

1.4 Aims of this thesis

In this thesis, we study different subjects related to the analysis of the large-scale structure of the universe based on the clustering of galaxies. We focus on the analysis of data from recent galaxy redshift surveys. We also develop a series of new statistical methods aimed at tackling some of the open problems in the study of galaxy clustering.

The thesis is divided in two parts. In Part I, we focus on galaxy clustering at small and intermediate scales. At these scales, the galaxy distribution is as much affected by the overall large-scale matter distribution, as by the processes governing galaxy formation and evolution, through bias. It is therefore essential to understand how the clustering of galaxies depends on the intrinsic properties of the galaxies in the samples considered (segregation). It is also important to study the evolution with time of the clustering properties, trying to disentangle the LSS evolution from the evolution of the properties of the galaxies themselves. We focus here on the possible studies that can be done using the ALHAMBRA survey.

First, in Chapter 2 we develop a method to recover the real-space correlation function from photometric redshifts, focusing on the order of accuracy expected for ALHAMBRA. We test this method using mock catalogues from the pencil-beam halo simulation of Heinämäki et al. (2005). We then apply this method to preliminary data from the ALHAMBRA survey in Chapter 3. We include a description of the characteristics of the survey there. We calculate the correlation function for several galaxy samples at different redshifts (up to $z = 1.5$), and luminosities. This allows us to study the luminosity segregation of galaxies, and its evolution, in a period that covers $\sim 70\%$ of the life of the universe. Finally, in Chapter 4 we review a series of statistical methods that allow a more thorough analysis of galaxy segregation. These methods are based on the mark statistics formalism, widely used in other fields. Here, we illustrate them using data from the 2dFGRS, and introduce for the first time the mark connection function for its use in cosmology.

In Part II, we study larger scales, and in particular we focus on baryon acoustic oscillations (BAO) in the galaxy distribution. BAO are a faint feature imprinted in the galaxy distribution by acoustic waves travelling in the baryon-photon plasma before recombination. As such, they provide a connection between the present-day LSS and the physics of the early universe. More importantly, BAO provide a standard ruler that can be used to trace with detail the expansion history of the universe.

BAO were detected in the galaxy distribution only recently (Eisenstein et al., 2005; Cole et al., 2005), using the largest available galaxy surveys. In Chapter 5 we review the physics of BAO and then focus on studying the reliability of their detection in current data. To this end, we calculate the large-scale correlation functions for several galaxy samples drawn from both the SDSS and 2dFGRS. Finally, in Chapter 6 we develop a new complementary method for the detection of BAO. This method is based on the use of wavelets, and it is aimed at detecting the structures responsible for the BAO. Moreover, thanks to the wavelet approach, this method provides more information than the usual two-point statistics, by keeping information about the localisation of the BAO signal in configuration space. In that chapter we present the method, and apply it to a galaxy catalogue from the SDSS.

Finally, Appendix A contains a list of the articles published, or submitted for publication, which contain part of the work presented in this thesis. Appendix B contains a summary of the contents of this thesis in Catalan.

Part I

Clustering evolution and galaxy segregation at small scales

2

Recovering the real-space correlation function from photometric redshift surveys

The Advanced Large Homogeneous Area Medium-Band Redshift Astronomical (ALHAMBRA) Survey (Moles et al., 2008) is a deep photometric redshift survey covering 4 deg^2. It has as one of its main objectives extending large-scale structure studies towards higher redshifts than it is possible using spectroscopic surveys. Several such deep photometric surveys have been proposed and carried out in recent years. They observe a region of the sky through a number of filters, and use the photometry obtained to determine the redshifts, z, and spectral energy distributions (SEDs) of galaxies. Using photometry instead of spectra allows them to get much deeper, but the uncertainty in the determination of redshifts is larger (Baum, 1962; Koo, 1986; Connolly et al., 1995; Fernández-Soto et al., 2001; Blake and Bridle, 2005). This uncertainty in z translates into an uncertainty in the determination of distances to the galaxies, and hence in their three-dimensional positions (Coe et al., 2006). This uncertainty is the main problem one has to tackle in order to being able to study the large-scale structure with these photometric surveys.

In this chapter, we study the way to recover the real-space correlation function from data of this kind of surveys, taking into account the presence of large redshift errors. We study the case of deep pencil-beam photometric surveys, as ALHAMBRA. However, most of this work should also be valid for the analysis of data from wide photometric redshift surveys, such as the Sloan Digital Sky Survey (SDSS, York et al., 2000), the Panoramic Survey Telescope and Rapid Response System (PanSTARRS) survey (Kaiser et al., 2002) or the Physics of the

Accelerating Universe (PAU) survey (Benítez et al., 2009a), as long as we take into account the possible differences due e.g. to redshift accuracies and range of scales studied.

Restricting ourselves to pencil-beam photometric surveys, there exist two other projects similar to ALHAMBRA, Classifying Objects by Medium-Band Observations (COMBO-17, Wolf et al., 2003), and the Ilbert et al. (2009) catalogue of photometric redshifts in the COSMOS field. COMBO-17 surveyed a total area of $\sim 1\,\mathrm{deg}^2$ using a combination of 17 broad-band and medium-band filters. It provided photometric redshifts for ~ 25000 galaxies with magnitude $R < 24$ in the redshift range $0.2 < z < 1.2$, with a typical error of $\Delta z \simeq 0.03$. Ilbert et al. compiled their photometric redshift catalogue using 30 broad-, medium- and narrow-band filters ranging from the ultraviolet to the mid-infrared. Their observations covered an area of $2\,\mathrm{deg}^2$ in the COSMOS field. They obtained a typical error of $\Delta z = 0.007(1 + z)$ for their 'bright' ($i^+_{AB} < 22.5$) sample, and $\Delta z = 0.033(1 + z)$ for their 'faint' ($i^+_{AB} > 23$) sample.

For comparison, the ALHAMBRA survey uses 20 medium-band equal-width filters in the optical range, plus the standard J, H, K_s near-infrared filters. The expectations are to obtain photometric redshifts for $\sim 3 \times 10^5$ galaxies with $I_{AB} \leq 24.7$, and $\Delta z \lesssim 0.015(1 + z)$, at $z_{med} = 0.74$ (see Section 3.1 for details). These numbers give us an idea of the typical redshift errors achievable with this kind of surveys.

In this chapter, we focus on the two-point correlation function, $\xi(r)$, which is the basic statistic for the analysis of large-scale structure. As discussed in Section 1.2, the use of this statistic in galaxy surveys is affected by the fact that we measure distances to galaxies through their redshifts. This adds two sources of uncertainty. In the first place, the measured redshift includes a contribution from the peculiar motion of the galaxy. In the second place, any uncertainties in the determination of redshifts affects the measured positions. The former is important for spectroscopic surveys, where redshift errors are typically small, while the latter dominates the uncertainties in photometric surveys. Moreover, the effect of peculiar velocities can be modelled and used to extract additional cosmological information from the survey.

We study how $\xi(r)$ is affected by redshift errors, and describe a method to recover its real-space value from photometric redshift survey data. The method we use is based on measuring the two-dimensional correlation function, $\xi(r_\parallel, r_\perp)$ (where $r_\parallel$ is the line-of-sight separation, and $r_\perp$ is the transverse separation), obtaining from it the projected correlation function, $w(r_\perp)$, and deprojecting it. This method (outlined in Section 2.1) was first proposed by Davis and Peebles (1983) as a way to avoid the uncertainties due to peculiar velocities in spectroscopic surveys, and has been used successfully in subsequent analyses (e.g. Saunders et al., 1992; Peacock et al., 2001; Hawkins et al., 2003; Madgwick et al., 2003b; Zehavi et al., 2004, 2005a).

Phleps et al. (2006) studied the correlation function of galaxies in COMBO-17. They used $w(r_\perp)$ as a measure of real-space clustering, and compared it to the predictions of halo occupation models. However, they did not attempt to recover $\xi(r)$ from their data.

We tested this method using data from the light-cone simulation of Heinämäki et al. (2005). From the simulation, we produced three mock photometric redshift catalogues, corresponding to different accuracies in the determination of redshifts. We then compared the correlation function $\xi(r)$ obtained by our method in each case to the real-space one, computed from the original catalogue. We performed additional tests using simulated catalogues based on point process models with a known analytical $\xi(r)$. The work presented in this chapter was published as Arnalte-Mur et al. (2009).

2.1 Method for recovering $\xi(r)$

The methods to estimate $\xi(r)$ described in Section 1.2 can not be used directly when the studied catalogue comes from a photometric survey. The large errors in redshift and hence in the line-of-sight positions produce two effects that have to be considered. On one side, these random shifts in position erase correlations between points (Snethlage et al., 2002), and hence the $\xi(r)$ estimated directly would be much lower than the real $\xi(r)$. On the other side, as the shifts are only in the line-of-sight direction, isotropy of the distribution is lost. Correlation is only lost along the longitudinal direction, but it is conserved in the transverse plane.

The method we used to recover real-space $\xi(r)$ from the photometric redshift catalogues is the same described in Davis and Peebles (1983) and Saunders et al. (1992) for spectroscopic surveys. It is based in the decomposition of pair separations in parallel and perpendicular distances ($r_\parallel$ and $r_\perp$, respectively).

We define $r_\parallel$ and $r_\perp$ as shown in an schematic way in Fig. 2.1. Let $\mathbf{s_1}$ and $\mathbf{s_2}$ be the measured positions (in 'observed redshift space') of two points in the catalogue. We then define the separation vector, $\mathbf{s} \equiv \mathbf{s_2} - \mathbf{s_1}$, and the line-of-sight vector, $\mathbf{l} \equiv \mathbf{s_1} + \mathbf{s_2}$, of the pair. From these, we now define the parallel and perpendicular distances of the pair as

$$r_\parallel \equiv \frac{|\mathbf{s} \cdot \mathbf{l}|}{|\mathbf{l}|} \quad , \quad r_\perp \equiv \sqrt{\mathbf{s} \cdot \mathbf{s} - r_\parallel^2} \, . \tag{2.1}$$

Once we have defined $r_\parallel$ and $r_\perp$ for each pair of points, we can define the two-dimensional correlation function, $\xi(r_\perp, r_\parallel)$, in an analogous way to $\xi(r)$ (see Section 1.2), dropping the assumption of isotropy[1], simply substituting the dependence on r by a dependence on $(r_\perp, r_\parallel)$. From $\xi(r_\perp, r_\parallel)$, we define the projected

---

[1] We note that we only drop this assumption partially, as we still assume that the point process is isotropic in the plane perpendicular to the line of sight.

correlation function as

$$w(r_\perp) \equiv 2\int_0^\infty \xi(r_\perp, r_\parallel)\mathrm{d}r_\parallel \,. \tag{2.2}$$

As w depends only on $r_\perp$, and the angle between any pair of points is small, it will not be affected significantly by redshift errors, as these will mainly produce shifts in $r_\parallel$.

Assuming that the real-space distribution is isotropic, we can relate w to the real-space correlation function, ξ_r, as

$$w(r_\perp) = 2\int_{r_\perp}^\infty \xi_r(r)\frac{r\mathrm{d}r}{(r^2 - r_\perp^2)^{1/2}} \,. \tag{2.3}$$

This relation can be inverted, obtaining ξ_r in terms of w as the Abel integral:

$$\xi_r(r) = -\frac{1}{\pi}\int_r^\infty \frac{\mathrm{d}w(r_\perp)}{\mathrm{d}r_\perp}\frac{\mathrm{d}r_\perp}{(r_\perp^2 - r^2)^{1/2}} \,. \tag{2.4}$$

Therefore, the method proposed to compute $\xi(r)$ from photometric survey data consists of the following steps. We first obtain $\xi(r_\perp, r_\parallel)$ from one of the estimation methods described in Section 1.2. The projected correlation function, $w(r_\perp)$, is then obtained by integration of equation (2.2). Finally, the real-space correlation function, $\xi(r)$, is calculated from equation (2.4).

Some problems arise in the numerical integration of equations (2.2) and (2.4). Both integrals extend formally to $+\infty$. However, when computing them numerically, we have to set finite upper limits, $r_{\parallel,\mathrm{max}}$ and $r_{\perp,\mathrm{max}}$. When integrating equation (2.4), the upper limit $r_{\perp,\mathrm{max}}$ is fixed, for pencil-beam surveys, by the maximum transverse separation allowed by the geometry of the survey. The value $r_{\parallel,\mathrm{max}}$ used in the integration of equation (2.2) is, however, more problematic. The value used should be large enough to include almost all the correlated pairs. However, if it is too large, this will introduce extra noise in the calculation. Typical values used in spectroscopic surveys are in the range $40 - 80\,h^{-1}$ Mpc (Hawkins et al., 2003; Madgwick

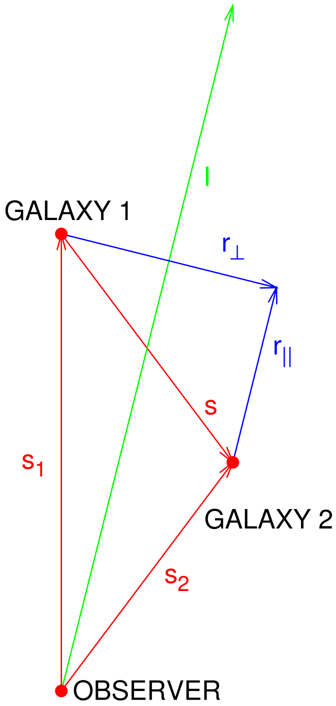


Figure 2.1: Diagram showing the separation of distances into the radial ($r_\parallel$) and transverse ($r_\perp$) directions, with the definition of the quantities entering equation (2.1).

et al., 2003b; Zehavi et al., 2004, 2005a), which is enough to take into account the effect of peculiar velocities. In the case of photometric surveys, however, we will need to integrate to larger values of $r_{\|}$, and moreover the value of $r_{\|,\mathrm{max}}$ will depend on the typical redshift error, Δz, of the catalogue. We will show below the tests we performed to obtain the optimal value in this case.

Redshift errors will influence the result in two ways. First, these errors change the apparent line-of-sight direction $\mathbf{l}/|\mathbf{l}|$ (see equation (2.1)), and through that, the apparent line-of-sight distance $r_{\|}$, and, most important, the perpendicular distance $r_{\perp}$. These errors grow with the redshift error and with the angular separation of the galaxy pair. Given the typical angles in 'pencil-beam' surveys, this source of error is negligible in our case.

Another, and much stronger source of errors is the assumption that the apparent distance in redshift space is the real distance between two galaxies – this assumption is necessary to obtain our basic integral relation (2.3). In the case of photometric errors, this assumption is hardly justified, but we will see that the inverted correlation functions are close to the real one, anyway. The errors caused by this assumption grow with the redshift errors.

2.2 Mock photometric catalogues

2.2.1 The dark matter halo simulation

We constructed a set of mock photometric catalogues based on the cosmological simulation of Heinämäki et al. (2005). This is a light-cone simulation, reproducing the geometry of a pencil-beam type survey, instead of a snapshot of the Universe at a given redshift. The output of the simulation is a catalogue containing dark matter haloes in the simulation, as they did not simulate the formation of galaxies in these haloes.

The method used to simulate the light-cone is as follows. They start from the simulation of a cube of comoving size 200 h^{-1} Mpc. Then, they choose the position and orientation of the light-cone with respect to this cube and its periodic replicas. Then, for each redshift snapshot, they compute which simulation particles are inside the light-cone. The orientation of the cone is chosen in such a way as to avoid possible spurious periodicities in the final catalogue due to the use of several copies of the same cube.

The code used to generate the simulation is the Multi Level Adaptive Particle Mesh (Knebe et al., 2001). The simulated cube contains 256^3 dark matter particles in a 256^3-point grid, and evolves from the initial conditions set at $z = 30$. They used the standard ΛCDM parameters ($\Omega_{\mathrm{DM}} = 0.226$, $\Omega_{\mathrm{b}} = 0.044$, $\Omega_{\Lambda} = 0.73$, $\sigma_8 = 0.84$). The mass of each particle in the simulation is $3.57 \times 10^{10}\, h^{-1}\, \mathrm{M}_{\odot}$. This means that a typical galaxy corresponds to a halo composed by a few simulation particles.

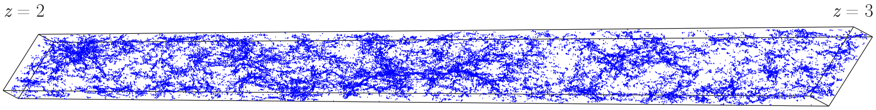


Figure 2.2: The halo catalogue from Heinämäki et al. (2005) used in this chapter. This covers the redshift range from $z = 2$ (left) to $z = 3$ (right). For clarity, the catalogue is diluted randomly so that only 20 per cent of the points are shown.

The simulated light cone covers an area of $2^\circ \times 0\overset{\circ}{.}5$ in the sky, and reaches a maximum redshift of $z = 6$. The haloes in the output catalogue are selected using a friends-of-friends algorithm. This algorithm collects together groups of particles with spacing smaller than a given linking length, ℓ. This length depends mildly on redshift, but is always $\ell \simeq 0.2\,h^{-1}$ Mpc. Finally, in order not to include in the catalogue random associations of particles which do not belong to gravitationally bound haloes, a virial equilibrium condition is applied. This means that they only include in the halo catalogue the selected groups for which $E_k/|E_p| < 0.5$ (where E_k is the kinetic energy, and E_p the potential energy of the group). The output catalogue contains all the haloes after this selection containing a number of particles $n \geq 2$, corresponding to a halo mass $M \geq 7.14 \times 10^{10}\,h^{-1}\,\mathrm{M}_\odot$.

As we are not interested in this chapter in the evolution of the correlation function with redshift, we restricted our analysis to the redshift bin $z \in [2, 3]$. The volume considered, in co-moving coordinates, is $864\,h^{-1}$ Mpc long in the line-of-sight direction, while its transverse section varies between $130 \times 32\,h^{-1}$ Mpc in its close end to $160 \times 40\,h^{-1}$ Mpc in its far end. The total volume is $4.56 \times 10^6\,h^{-3}\,\mathrm{Mpc}^3$, and the number of haloes in the catalogue for this region is $\sim$ 180000. We show the resulting catalogue in Fig. 2.2

We chose the interval at such a high redshift in order to be able to study a relatively large range of scales. At lower redshift, the light cone is narrower, and we would have to restrict the study to the shortest scales. Obviously, the range of scales to study in a real survey such as ALHAMBRA will depend on the actual geometry and depth of the survey.

Our intention in this chapter is to use this halo catalogue, and its correlation function, to test the method presented in Section 2.1. Therefore, we do not make any selection of the haloes (e.g. by mass) to better reproduce the correlation function of any particular galaxy population. In this way, we use the full halo catalogue, so that we minimise the influence of Poisson noise in our analysis. Overall, the behaviour of the halo correlation function should be similar enough to the galaxy correlation function as to correctly assess the validity of our method.

2.2.2 Simulating the photometric redshift errors

From this simulated halo catalogue, we generated three mock 'photometric redshift catalogues', corresponding to surveys with redshift uncertainties $\Delta z/(1+z) = 0.05, 0.015, 0.005$. The first case, $\Delta z/(1+z) = 0.05$, corresponds typically to a survey using ~ 5 broad-band filters (see e.g. Fernández-Soto et al., 2001). $\Delta z/(1+z) = 0.015$ corresponds to the value expected from the ALHAMBRA survey (Moles et al., 2008). The last case, $\Delta z/(1+z) = 0.005$, would correspond to a survey using even more filters. As explained above, Ilbert et al. (2009) obtained a value of $\Delta z/(1+z) = 0.007$, although only for the bright end of their sample. Also, the PAU survey project aims at obtaining photometric redshifts for Luminous Red Galaxies (LRGs) with uncertainties $\Delta z/(1+z) \sim 0.0035$ for $z \lesssim 0.9$. As the uncertainty in photometric redshifts decreases for high redshift galaxies ($z \gtrsim 2.5$), when the Lyman-α wavelength enters into the visible domain, it should also be possible, in principle, to get such a small Δz in this case.

In creating our mock catalogues, we assumed Gaussian errors for the photometric redshifts. This is not generally the case for real surveys, due to the existence of a fraction of 'catastrophic' redshift determinations, and to the mix of different classes of objects with a variety of photometric redshift errors. However, our assumption of single-peaked Gaussian-distributed errors would be valid for a catalogue selected to contain only "good" redshifts. This catalogue could be built combining the selection of a given class of objects (e.g. LRGs), with the use of some estimate of the redshift determination quality. The latter could be the knowledge of the full redshift probability distribution function (Fernández-Soto et al., 2002), or the 'odds' parameter in the case of Bayesian methods (Benítez, 2000). Existing experience indicates that, depending on the survey design, it is possible to obtain "good" redshifts for objects down to magnitudes $m_{\rm lim} - 1$ or $m_{\rm lim} - 2$, where $m_{\rm lim}$ is the limit magnitude of the survey (Fernández-Soto et al., 2001).

In Section 2.4, we assess the robustness of our results to the presence of 'catastrophic' redshifts. We consider catalogues with a fraction of such outliers of 5 per cent. This is a conservative value, typical of broad-band, non-optimised photometric redshift surveys, and it should be significantly smaller in the case of "good" redshifts. As an example, Ilbert et al. (2009) obtained just 0.7 per cent of outliers when comparing their bright sample ($i^+_{AB} < 22.5$) to spectroscopic redshifts.

To generate each mock catalogue, we modified the position of each point in the simulation following these steps:

1. We calculated the 'cosmological redshift' of the object from its real-space position.

2. We added to this ‘cosmological redshift’ the redshift due to the line-of-sight peculiar velocity of the object. These peculiar velocities of the haloes are provided by the simulation.

3. To simulate the expected redshift errors, we added a random shift to the resulting redshift, following a Gaussian distribution. This distribution has a standard deviation equal to Δz, which depends on the mock catalogue we simulate and on the redshift of each galaxy. The redshift obtained is the ‘observed redshift’ of the object.

4. We finally obtained the three-dimensional position of the object corresponding to this ‘observed redshift’ and included it in the mock catalogue.

This distortion process was carried out for all the points in the whole cone of the simulation. The selection of the points in the redshift bin $z \in [2,3]$ was performed using the new ‘observed redshifts’, thus simulating the selection process in a real survey.

Fig. 2.3 shows the distribution of haloes in the original catalogue and in the mock photometric catalogues. The upper panel of the figure shows the real space positions of haloes, not affected by peculiar velocities, and thus does not show the finger-of-God or coherent infall effects observed in spectroscopic surveys. Due to redshift errors, structures which are clearly seen in real space are smoothed and hardly recognisable in photometric redshift data.

2.3 Application of the method to simulated catalogues

2.3.1 Calculation of $\xi_{\mathrm{r}}(r)$ in practice

We used the method outlined in Section 2.1 to recover the real-space correlation function $\xi_{\mathrm{r}}(r)$ from the mock photometric catalogues obtained from the halo simulation. In this section, we explain some details of the calculation we made in this case.

We obtained the two-dimensional correlation function $\xi(r_\perp, r_\parallel)$ using the estimator of Landy and Szalay (1993) (see Section 1.2). We generated a random Poisson distribution in the same volume as the simulated catalogue, and then estimated $\xi(r_\perp, r_\parallel)$ as

$$\hat{\xi}(r_\perp, r_\parallel) = 1 + \left(\frac{N_R}{N_D}\right)^2 \frac{DD(r_\perp, r_\parallel)}{RR(r_\perp, r_\parallel)} - 2\frac{N_R}{N_D}\frac{DR(r_\perp, r_\parallel)}{RR(r_\perp, r_\parallel)}\,.$$

Here, N_D is the number of points in the data catalogue, N_R is the number of points in the random Poisson catalogue, and DD, DR and RR correspond to

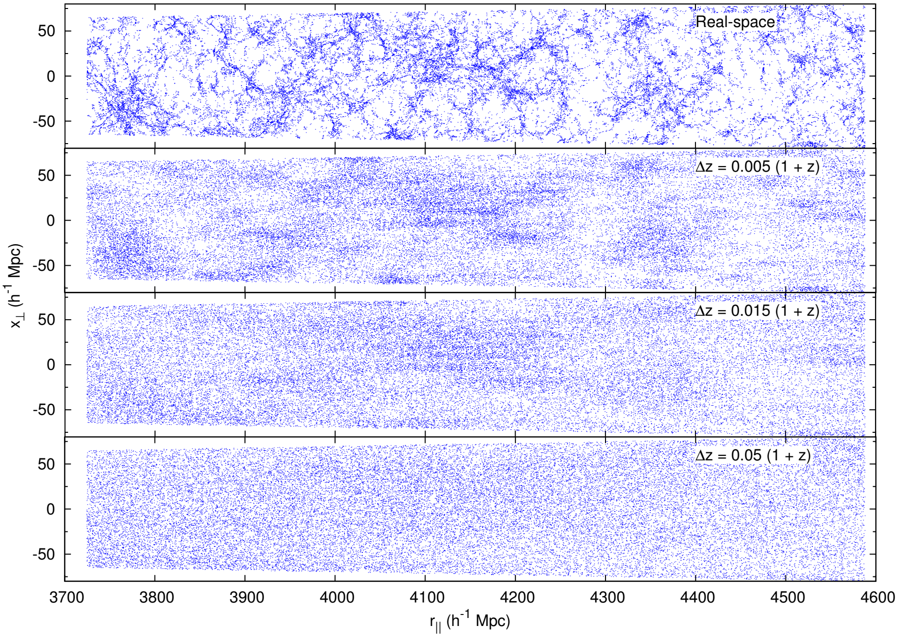


Figure 2.3: The distribution of haloes in the four catalogues used: the original real-space catalogue, and the three mock photometric catalogues. The distribution is shown projected on a longitudinal plane ($x_\perp$ is one of the coordinates contained in the transverse plane), and only 20 per cent of the points are shown, for clarity.

pairs of points with transverse separations in the interval $[r_\perp, r_\perp + \mathrm{d}r_\perp]$ and line-of-sight separations in the interval $[r_\parallel, r_\parallel + \mathrm{d}r_\parallel]$. DD counts pairs of points in the data catalogue, RR counts pairs in the random Poisson catalogue, and DR counts pairs crossed between a point in the data catalogue and a point in the Poisson catalogue. In this case, we used a Poisson catalogue with $N_R = 5N_D$ points. As the simulation does not contain any observational effects, we can assume that the catalogues are volume limited. Therefore we generated the Poisson catalogue with a uniform density over the whole volume.

From $\xi(r_\perp, r_\parallel)$ we obtained $w(r_\perp)$ using equation (2.2) modified as

$$w(r_\perp) \equiv 2 \int_0^{r_{\parallel,\max}} \xi(r_\perp, r_\parallel) \mathrm{d}r_\parallel \,. \tag{2.5}$$

We performed several tests to choose the appropriate value for $r_{\parallel,\max}$, as explained in Section 2.4.1. We finally adopted a value of $r_{\parallel,\max} \simeq 4r(\Delta z)$ in each case for our calculations. Here, $r(\Delta z)$ is the comoving scale corresponding to the redshift difference Δz calculated at the mid-point of the survey (in this case, $z = 2.5$). The exact values used were $r_{\parallel,\mathrm{max}} = 66$, 200, $600\,h^{-1}$ Mpc, respectively for $\Delta z/(1+z) = 0.005$, 0.015, 0.05.

We finally obtained the de-projected real-space correlation function from equation (2.4). The value used for the integration limit was $r_{\perp,\mathrm{max}} = 130\,h^{-1}$ Mpc. This is about 80 per cent of the maximum transverse separation allowed by the geometry of the light-cone. We used 32 bins in σ, logarithmically spaced between $0.1\,h^{-1}$ Mpc and $r_{\perp,\mathrm{max}}$. The way we used to evaluate (2.4) was that of Saunders et al. (1992). We interpolated linearly w between its values in each $r_\perp$ bin, and then integrated (2.4) analytically. Taking w_i as the value of w for the bin centred at $r_{\perp,i}$, we have

$$\xi(r_{\perp,i}) = -\frac{1}{\pi} \sum_{j \geq i} \frac{w_{j+1} - w_j}{r_{\perp,j+1} - r_{\perp,j}} \ln \left(\frac{r_{\perp,j+1} + \sqrt{r^2_{\perp,j+1} - r^2_{\perp,i}}}{r_{\perp,j} + \sqrt{r^2_{\perp,j} - r^2_{\perp,i}}} \right) . \qquad (2.6)$$

To estimate the correlation function error for each bin in r, we used the jackknife method (see e.g. Norberg et al., 2009). We divided our volume in $N_{\mathrm{jack}} = 12$ equal sub-volumes, and constructed our jackknife samples omitting one sub-volume at a time. We repeated the full calculation of $\xi(r)$ for each of these samples. Denoting by ξ_i^k the value of the correlation function obtained for bin i in jackknife sample k, the error in this bin is then

$$\sigma_i^2 = \frac{N_{\mathrm{jack}} - 1}{N_{\mathrm{jack}}} \sum_{k=1}^{N_{\mathrm{jack}}} \left(\xi_i^k - \bar{\xi}_i \right)^2 , \qquad (2.7)$$

where $\bar{\xi}_i$ is the average of the values obtained for bin i, and $N_{\mathrm{jack}} = 12$. Given that the number of jackknife regions used is small, we can not reliably estimate the full covariance matrix using this method. However, we expect this method to provide a good approximation for the individual bins' errors σ_i. We test this error estimation below using a set of realisations of a segment Cox process (Section 2.4.3).

2.3.2 Results

Effect of redshift errors on the redshift-space correlation function

To illustrate the loss of correlation due to photometric errors, we calculated directly the redshift-space correlation function $\xi_z(s)$ in each mock catalogue. For this calculation, we assumed isotropy, so we used the standard Landy and Szalay (1993) estimator, as in Section 1.2. We show the correlation functions obtained in Fig. 2.4, together with $\xi_{\mathrm{r}}(r)$ measured directly from the real-space catalogue.

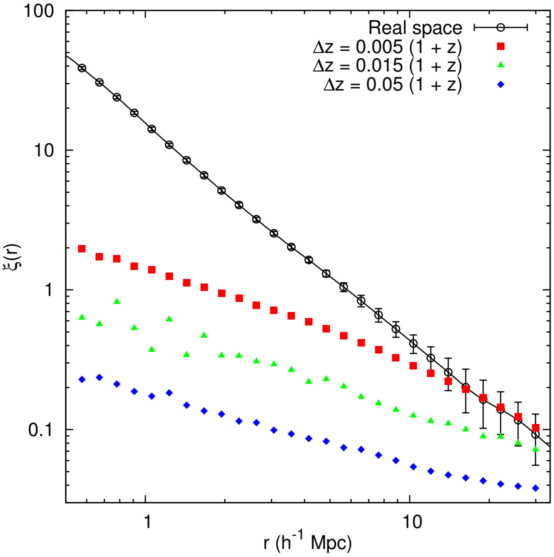
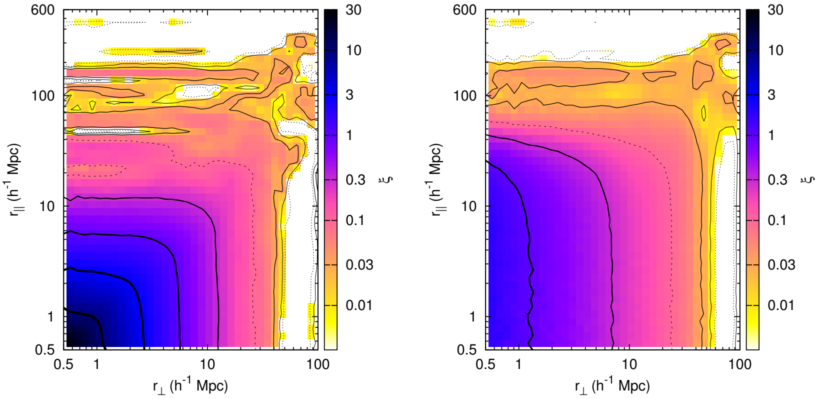


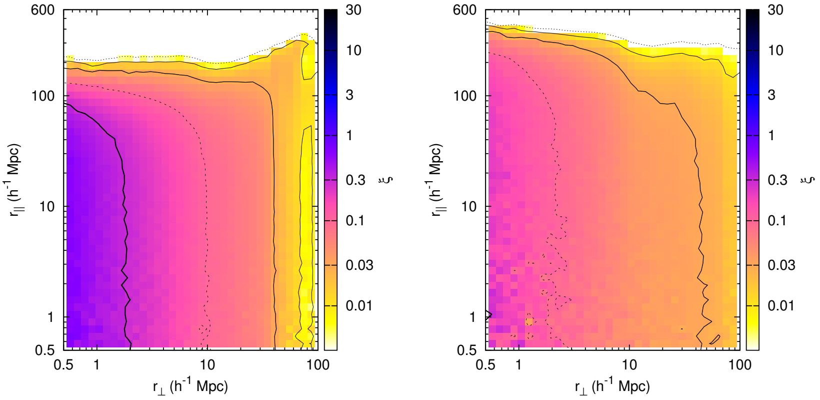
Figure 2.4: The redshift-space correlation function $\xi_z(s)$ measured directly from the catalogues, assuming isotropy. For comparison, we also show the real-space correlation function $\xi_{\mathrm{r}}(r)$ measured directly from the real-space catalogue, together with the spline approximation used for further comparisons (line). The error bars plotted correspond to the diagonal terms in the covariance matrix, $C_{ii}^{1/2}$.

We see clearly how the estimated correlation decreases with Δz. This effects is larger for smaller scales (where the real-space correlation is stronger). In order to later compare the real-space result to the one obtained using our method in the mock photometric catalogues, we fitted $\xi_r(r)$ by a third-order spline, also shown in the figure.

As a first step in the calculation of the deprojected $\xi(r)$, we calculated $\xi(r_\perp, r_\parallel)$ for each mock catalogue. The results are shown in Fig. 2.5. We also plot the $\xi(r_\perp, r_\parallel)$ obtained in the real-space catalogue, for comparison. Two effects of the redshift errors can be observed. First, correlation decreases with the value of Δz for each catalogue, as seen in Fig. 2.4. Also, there is a loss of symmetry of $\xi(r_\perp, r_\parallel)$ in these plots. In real space, due to the isotropy of the distribution, $\xi(r_\perp, r_\parallel)$ has circular symmetry (seen as a ‘boxy’ symmetry in the logarithmic scale used). However, we can see that when we calculate it for the mock photometric catalogues, the distribution gets stretched along the $r_\parallel$ axis.



(a) Real space **(b)** $\Delta z = 0.005(1 + z)$



(c) $\Delta z = 0.015(1 + z)$ **(d)** $\Delta z = 0.05(1 + z)$

Figure 2.5: The two-dimensional correlation function $\xi(r_{\perp}, r_{\parallel})$ obtained for the real space catalogue, and for the mock photometric catalogues with different values of Δz. Contours are drawn at $\xi = 10, 3, 1, 0.3, 0.1, 0.03, 0.01, 0$, with decreasing thickness. Contours at 0.1 and 0 are dashed.

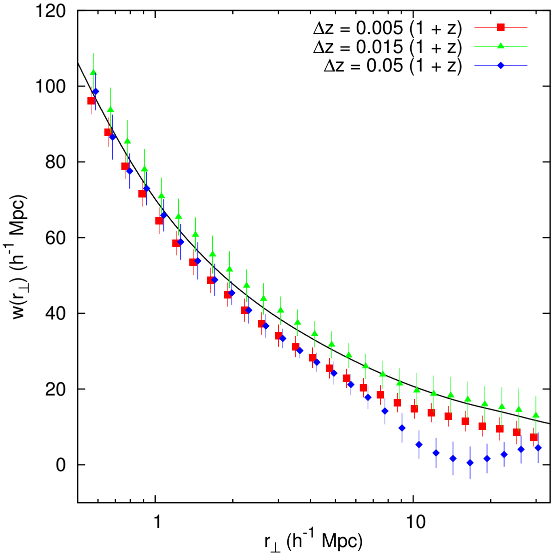


Figure 2.6: The projected correlation function obtained from the mock photometric catalogues, compared to the real-space result. The solid line corresponds to a spline fit to $\xi_r(r)$, as explained in the text. Small shifts have been applied along the $r_\perp$ axis, for clarity. The feature observed at large scales for the $\Delta z = 0.05(1+z)$ catalogue is due to the large $r_{\parallel,\max}$ value used in that case.

Tests of the deprojection method

From $\xi(r_\perp, r_\parallel)$, we obtained the projected correlation function $w(r_\perp)$ for each of the mock photometric catalogues. In Fig. 2.6 we compare the function $w(r_\perp)$ calculated for the mock catalogues to the function obtained from the spline fit to the real-space $\xi_{\rm r}(r)$, according to equation (2.3). The results obtained for the $\Delta z/(1+z) = 0.005$ and $\Delta z/(1+z) = 0.015$ catalogues follow closely the real-space result. In the case of the $\Delta z/(1+z) = 0.05$ catalogue, however, $w(r_\perp)$ falls below the real-space result for $r_\perp \gtrsim 10\,h^{-1}\,$Mpc. This feature is due to the fact that the value of $r_{\parallel,\max} = 600\,h^{-1}\,$Mpc used for that catalogue gets close to the line-of-sight length of the simulation box used. We obtained a similar feature when artificially using this large value of $r_{\parallel,\max}$ for the calculation in the two other mock catalogues. However, as $r_{\parallel,\max}$ scales with Δz in the standard calculation, this issue does not affect the other catalogues.

Our final result for the deprojected correlation function, $\xi_{\rm dep}(r)$ obtained from the mock photometric catalogues is shown in Fig. 2.7, where we compare it to the real-space correlation function, $\xi_{\rm r}(r)$. We restricted the analysis to the scale range $r \in [0.5, 30]\, h^{-1}$ Mpc. The low limit was imposed by the friends-of-friends algorithm used to select haloes in the simulation. If we had two haloes at a too small separation, they would merge into a single one. This prevents us from measuring $\xi(r)$ at such small distances. We therefore set a conservative limit corresponding to $\simeq 2\ell$ (ℓ is the linking length used). The upper limit was fixed because of the geometry of the light cone. To avoid any problems related to border effects, we set this limit approximately as the maximum separation along the short transverse axis.

From the figure, we see that, for $\Delta z \leq 0.015(1+z)$, we recover the real-space correlation function, within the errors, for the full range of scales studied. However, for the largest redshift error studied, $\Delta z = 0.05(1+z)$, the deprojected $\xi(r)$ deviates from the real-space one for large scales ($r \gtrsim 4\, h^{-1}$ Mpc). This issue is related to the feature observed in $w(r_\perp)$.

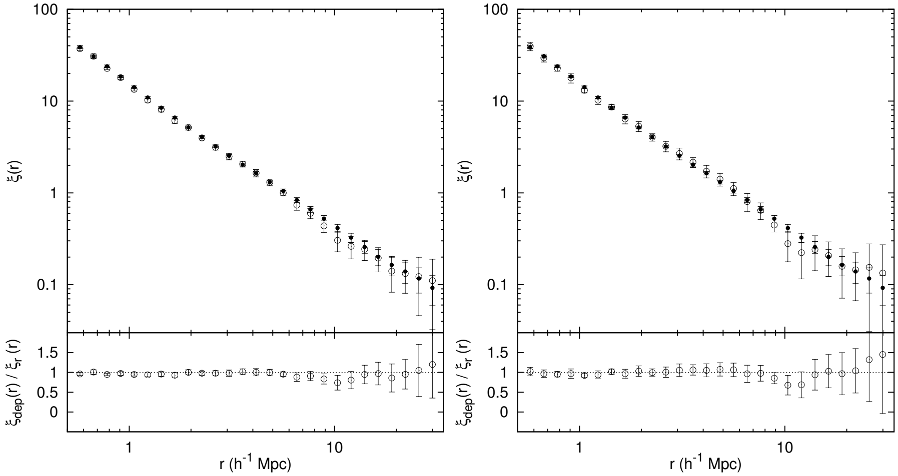
To quantify the quality of the recovery, we used an 'average normalised residual', $\Delta\xi$, as figure of merit, defined as

$$\Delta\xi = \frac{1}{N} \sum_i \left| \frac{\xi_{\rm dep}(r_i) - \xi_r(r_i)}{\xi_r(r_i)} \right| ,$$

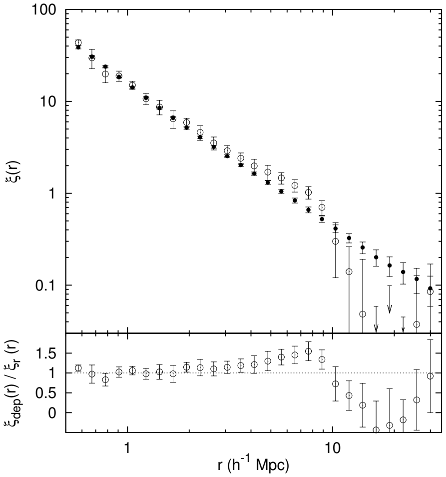
where r_i are the values of the bins in r where we measure ξ, and N is the number of such bins considered. Without prior knowledge of $\xi_r(r)$ we could anyhow estimate the quality of the recovery calculating the quantity:

$$\widehat{\Delta\xi} = \frac{1}{N} \sum_i \frac{C_{ii}^{1/2}}{|\xi_{\rm dep}(r_i)|} .$$

We computed $\Delta\xi$ and $\widehat{\Delta\xi}$ for different ranges in r, in order to assess the validity of the method at different scales. The values of $\Delta\xi$ and $\widehat{\Delta\xi}$ obtained for the different mock catalogues are shown in Table 2.1. From the values of $\Delta\xi$ we see that we recover $\xi(r)$ within a 5 per cent in the average for scales $r < 10\, h^{-1}$ Mpc for mock catalogues with $\Delta z \leq 0.015(1+z)$. At larger scales, the deviations from $\xi_{\rm r}$ are larger ($12-20$ per cent). In the case with the largest redshift errors, our method is only valid for very small scales, $r < 2\, h^{-1}$ Mpc, where the average deviations are of a 7 per cent. We note that, in all cases where the method is valid, $\widehat{\Delta\xi} > \Delta\xi$. Hence, the jackknife method allows us to estimate the errors to an acceptable precision. We remark, however, that for large values of Δz the jackknife error underestimates the real one as measured from the residuals or compared to other Δz values, specially over medium scales ($2-20\, h^{-1}$ Mpc).



(a) $\Delta z = 0.005(1+z)$ **(b)** $\Delta z = 0.015(1+z)$



(c) $\Delta z = 0.05(1+z)$

Figure 2.7: Comparison between the deprojected correlation function, $\xi_{\rm dep}(r)$ (open circles), and the real-space correlation function, $\xi_{\rm r}(r)$ (solid circles), for each mock photometric catalogue. The error bars plotted correspond to the diagonal terms in the covariance matrix, $C_{ii}^{1/2}$.

	$\frac{\Delta z}{(1+z)} = 0.005$		$\frac{\Delta z}{(1+z)} = 0.015$		$\frac{\Delta z}{(1+z)} = 0.05$	
Range (h^{-1} Mpc)	$\Delta\xi$	$\widehat{\Delta\xi}$	$\Delta\xi$	$\widehat{\Delta\xi}$	$\Delta\xi$	$\widehat{\Delta\xi}$
$0.5 < r < 30$	0.07	0.17	0.09	0.26	0.36	0.67
$0.5 < r < 2$	0.04	0.05	0.04	0.10	0.07	0.15
$2 < r < 10$	0.05	0.09	0.05	0.16	0.28	0.16
$10 < r < 30$	0.12	0.40	0.20	0.57	0.79	1.89

Table 2.1: Values of $\Delta\xi$ and $\widehat{\Delta\xi}$ obtained for the three mock photometric catalogues and for different scale ranges.

2.4 Additional tests of the method

In this section, we describe additional calculations made to assess the robustness of the method, and to find the optimal parameters for the calculation. In the first place, we describe the way in which we determined the optimal value for $r_{\parallel,\mathrm{max}}$ in the integration along the line-of-sight direction. In the second place, we test the performance of the method in the presence of catastrophic redshift errors. Finally, we test the method in a set of mock catalogues obtained from a point process model with known analytical correlation function. We use these catalogues also to test the reliability of the jackknife error estimation.

2.4.1 Determination of the optimal value for $r_{\parallel,\mathrm{max}}$

As explained in Section 2.3.1, we chose a value of $r_{\parallel,\mathrm{max}} \simeq 4r(\Delta z)$ for the upper limit in the integration of equation (2.5). Here we show the tests we performed in order to make this selection.

The selection of a value for $r_{\parallel,\mathrm{max}}$ is a tricky issue. When increasing its value, one includes more correlated pairs, and thus the recovered $\xi_{\mathrm{dep}}(r)$ gets closer to its real value. However, at the same time one includes more uncorrelated pairs, increasing the noise in the final value of $\xi_{\mathrm{dep}}(r)$. That is, a too low value for $r_{\parallel,\mathrm{max}}$ would induce a systematic underestimation of $\xi_{\mathrm{r}}(r)$, while a too high value would add extra noise. The optimal value for $r_{\parallel,\mathrm{max}}$, therefore, depends on the 'dispersion' of the correlated pairs along the line of sight, i.e. how the observed value of $r_{\parallel}$ differs from the real one for each of the pairs.

This 'dispersion' of the pairs depends linearly on the typical redshift error for the points in the sample, Δz. Hence, we make the assumption that the optimal value for $r_{\parallel,\mathrm{max}}$ depends linearly on $r(\Delta z)$ as

$$r_{\parallel,\mathrm{max}} = a\, r(\Delta z)\,.$$

The purpose of our tests is therefore to find the optimal value for the parameter a. We repeated the calculation of $\xi_{\mathrm{dep}}(r)$ for several values of a ranging from $a = 1$

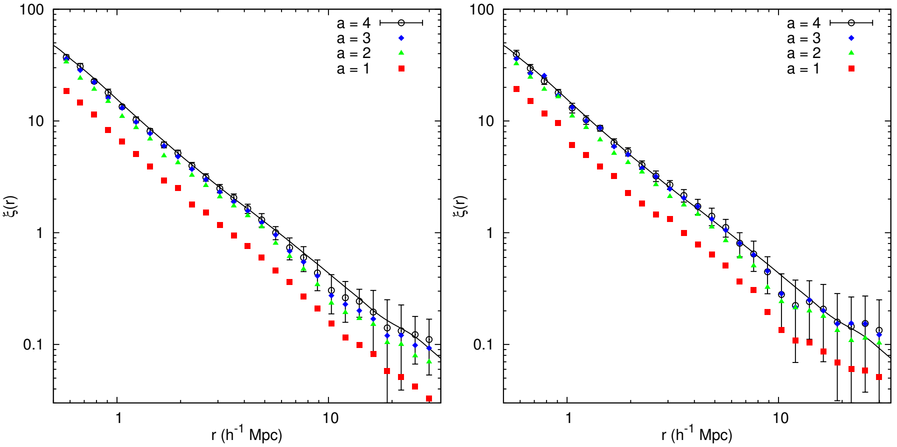
to $a = 6$. The results obtained for the different mock catalogues are shown in Fig. 2.8, where they are compared to the known real-space $\xi_{\rm r}(r)$. The behaviour is similar for the different values of Δz. For $a \leq 2$, the $\xi_{\rm dep}(r)$ obtained is clearly lower than $\xi_{\rm r}(r)$. That is, we are still loosing a significant fraction of correlated pairs. For $a \geq 3$, we approximately recover the values of $\xi_{\rm r}(r)$ (with the caveats already discussed above). However, the result is slightly noisier for larger values of a. We therefore chose the final value at $a = 4$ as a conservative choice, in order to avoid any possible systematic effect that may remain when using $a = 3$.

A possible variation of the method would be to use a smaller value of $r_{\parallel,\rm max}$, and multiply the resulting $\xi_{\rm dep}(r)$ by a constant correction factor. This would reduce the extra noise introduced by the integration along a large range in the $r_{\parallel}$ direction. From Fig. 2.8 we see that, for $r_{\parallel,\rm max} = r(\Delta z)$ (i.e., $a = 1$), a correction factor of $\simeq 2$ works well in our simulation, generally reducing the error. However, the optimal value is slightly different for each mock photometric catalogue. The main problem for the use of this method would be the accurate determination of the correction factor in each case, as any deviation from the optimal value would introduce a bias in the result. In the case of real data, we would need to obtain this correction factor externally from simulations, such as those used here. However, as shown by the small differences between mocks, the Gaussian approximation used here is probably not so close to reality as to infer that constant from our simulations.

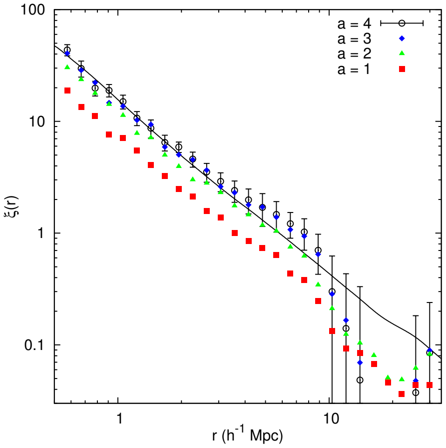
2.4.2 Test of the deprojection method in the presence of outliers

In previous sections, we assumed that the distribution of the photometric redshift errors was purely Gaussian. The main problem with this assumption comes from the outliers produced by ‘catastrophic’ redshift errors. These are typically cases in which the algorithm for computing the photometric redshift confuses two features of the spectrum. For example, assigning a break in the spectrum the (rest) wavelength of the Lyman break, when it really corresponds to the Balmer break, would give a galaxy with a real redshift of $z \simeq 0.2$ a measured redshift of $z_p \simeq 2.5$. This same problem is known to affect also spectroscopic redshift, due to confusion between spectral lines (Fernández-Soto et al., 2001). The final effect is that a fraction of the measured redshifts will be scattered randomly through the redshift interval.

We performed a simple test to assess the robustness of our method to the presence of such ‘catastrophic’ errors. We repeated our full calculation in mock catalogues containing 5 per cent of outliers, which we consider to be a conservative estimate. We generated outliers by selecting points at random in the original catalogue, and assigning them a random distance within the range considered. The results for the recovered real-space $\xi(r)$ are shown in Fig. 2.9.

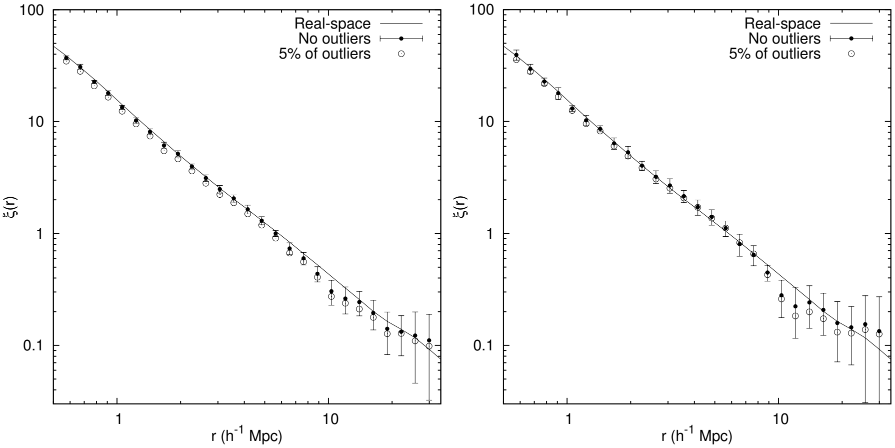


(a) $\Delta z = 0.005(1+z)$ **(b)** $\Delta z = 0.015(1+z)$

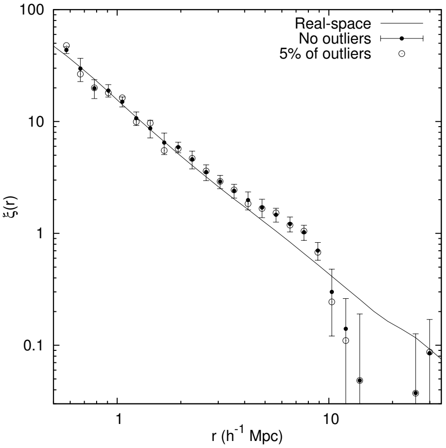


(c) $\Delta z = 0.05(1+z)$

Figure 2.8: Comparison between the de-projected correlation function $\xi_{\rm dep}(r)$ obtained using different values of $r_{\parallel,\rm max}$ for each mock photometric catalogue. The solid line in each case corresponds to the spline fit to the $\xi_{\rm r}(r)$ measured directly in the real-space catalogue.



(a) $\Delta z = 0.005(1+z)$ **(b)** $\Delta z = 0.015(1+z)$



(c) $\Delta z = 0.05(1+z)$

Figure 2.9: The deprojected correlation function obtained for the three mock catalogues containing a 5 per cent of outliers, compared to our original results (Fig. 2.7). In each case, the continuous line is the real-space correlation function.

Range (h^{-1} Mpc)	$\frac{\Delta z}{(1+z)} = 0.005$	$\frac{\Delta z}{(1+z)} = 0.015$	$\frac{\Delta z}{(1+z)} = 0.05$
$0.5 < r < 30$	0.14	0.12	0.38
$0.5 < r < 2$	0.12	0.08	0.13
$2 < r < 10$	0.13	0.05	0.28
$10 < r < 30$	0.17	0.25	0.79

Table 2.2: Values of $\Delta\xi$ obtained for the three mock photometric catalogues and for different scale ranges, in the case in which we simulate a fraction of 5 per cent of outliers.

Even with the conservative assumption of a large fraction of outliers, our method is able to recover $\xi(r)$. However, the quality of this recovery is slightly worse. This is shown by the values of the residuals $\Delta\xi$, shown in Table 2.2. When comparing these results to those obtained without outliers (Table 2.1), we see how the strongest effect of the outliers appears for the catalogue with $\Delta z = 0.005(1+z)$. For larger values of Δz, the fact that we integrate over a large range in $r_\parallel$ reduces the impact of outliers, as we recover a high fraction of the correlated pairs even if the redshift error is very large. For the smallest value of Δz, however, we use a small $r_{\parallel,\max}$, which means that we loose most of the correlation due to outliers. Fortunately, we expect the catalogues with the best redshift errors Δz to have also a smaller fraction of outliers (see e.g. Ilbert et al., 2009).

2.4.3 Application of the method to a segment Cox process

We additionally tested the recovery method, using a set of mock catalogues based on a segment Cox process. This point process model has the advantage of having a known analytical expression for the correlation function, and of being easy and fast to generate. The purpose of of this approach was two-fold. On one side, we wanted to perform an extra test of the method using a simple model with known correlation properties. On the other side, this also serves as a test of the jackknife estimation of errors, as we can compare this estimation to the variance of the results for several independent realisations of the model.

A segment Cox process is a particular case of a Cox field. Cox fields (Martínez and Saar, 2002) are a class of point processes which are generated following a two-step algorithm. In a first step, an intensity field is generated in the volume as a realisation of a random field following a given distribution. In a second step, an inhomogeneous Poisson process is generated following that intensity field. In this way, Cox processes are doubly stochastic. The particular case of the segment Cox process, which we used for our analysis, is produced in the following way: segments of a given length, l, are randomly scattered within a volume. Then, points are randomly distributed along these segments. The length density of the

	$\frac{\Delta z}{(1+z)} = 0.005$		$\frac{\Delta z}{(1+z)} = 0.015$		$\frac{\Delta z}{(1+z)} = 0.05$	
Range (h^{-1} Mpc)	$\Delta\xi$	$\widehat{\Delta\xi}$	$\Delta\xi$	$\widehat{\Delta\xi}$	$\Delta\xi$	$\widehat{\Delta\xi}$
$0.5 < r < 10$	0.08	0.16	0.18	0.23	0.35	0.40
$0.5 < r < 2$	0.06	0.08	0.07	0.12	0.23	0.21
$2 < r < 10$	0.10	0.22	0.27	0.33	0.46	0.57

Table 2.3: Values of $\Delta\xi$ and $\widehat{\Delta\xi}$ obtained for the three mock photometric catalogues obtained from a Cox process, and for different scale ranges.

system of segments is $L_V = \lambda_s l$, where λ_s is the mean number of segments per unit volume. The density of the point process is then

$$\lambda = \lambda_l L_V = \lambda_l \lambda_s l \,,$$

where λ_l is the mean number of points per unit length of the segments. The correlation function of the point process equals the correlation function of the system of segments (Stoyan et al., 1995), which is given by

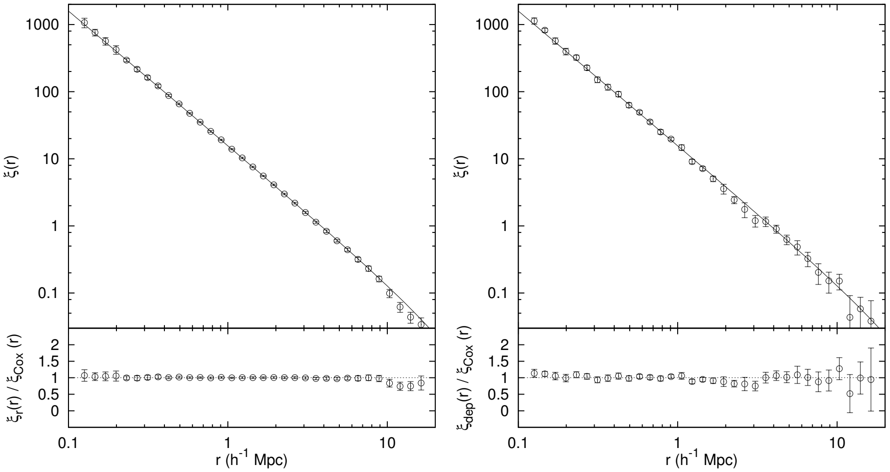
$$\xi_{\rm Cox}(r) = \begin{cases} \frac{1}{2\pi r^2 L_V} - \frac{1}{2\pi r l L_V} \,, & r \leq l \\ 0 \,, & r > l \end{cases} \,. \tag{2.8}$$

We simulated a segment Cox process in the same volume as the catalogues from the haloes simulation. The parameters we used were $l = 50\,h^{-1}\,\mathrm{Mpc}$, $\lambda_s = 2 \cdot 10^{-4}\,h^3\,\mathrm{Mpc}^{-3}$ and $\lambda_l = 4\,h\,\mathrm{Mpc}^{-1}$, which result in $L_V = 0.01\,h^2\,\mathrm{Mpc}^{-2}$ and $\lambda = 0.04\,h^3\,\mathrm{Mpc}^{-3}$. These parameters were chosen to approximately match the density of points and the behaviour of $\xi(r)$ in the haloes simulation. We considered the catalogue obtained directly from the segment Cox process as the 'real-space' catalogue. We created three mock 'photometric redshift catalogues' following the same procedure and using the same values for Δz as described in Section 2.2.2.

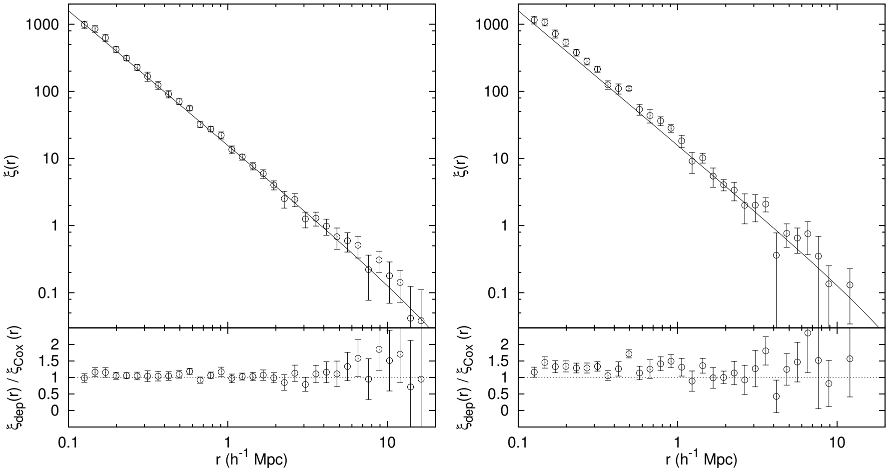
We calculated directly the correlation function for the real-space catalogue and, for the three mock 'photometric catalogues', we used the method described in Sections 2.1 and 2.3.1 to obtain the deprojected correlation function. The estimation of errors was performed using the same jackknife method as described above for the haloes simulation case. Below we show how we also used the Cox segment process to analyse the reliability of this error estimation. We show in Fig. 2.10 the comparison of our results to the analytical prediction (equation 2.8).

We quantify the quality of the recovery in the same way as we did for the haloes simulation, using the quantities $\Delta\xi$ and $\widehat{\Delta\xi}$. In this case, we define $\Delta\xi$ as the relative deviation of $\xi_{\rm dep}(r)$ from the analytical prediction $\xi_{\rm Cox}(r)$ of equation (2.8). The values obtained are shown in Table 2.3.

We recover the real-space correlation function within a 10 per cent for the $\Delta z/(1+z) = 0.005$ catalogue. In this case, however, our method starts to fail at



(a) Real space **(b)** $\Delta z = 0.005(1+z)$



(c) $\Delta z = 0.015(1+z)$ **(d)** $\Delta z = 0.05(1+z)$

Figure 2.10: The correlation function measured in the real-space and the three mock photometric catalogues obtained from a segment Cox process (open circles), compared to the analytical prediction, equation (2.8), for this process (solid line).

$r \simeq 3 - 4\,h^{-1}$ Mpc for the $\Delta z/(1+z) = 0.015$ catalogue (this is seen as a larger value of $\Delta\xi$ for this range, and as an increasing trend in Fig. 2.10). When applying the method to the $\Delta z/(1+z) = 0.05$ catalogue, $\xi_{\rm dep}(r)$ is consistently higher than $\xi_{\rm Cox}(r)$. Although this bias could be an artifact of this particular point process, it also means that the deprojection method described in this Chapter can not be fully trusted when it is applied to catalogues with large redshift errors.

Finally, we also used the segment Cox process mocks to assess the reliability of the jackknife error estimation. As realisations of this Cox process are easy and fast to generate, we can compare our error estimate to the variance and covariances obtained from several independent realisations.

The jackknife error estimation for $\xi_{\rm dep}(r)$ was performed here in the same way as for the haloes simulation catalogues, by dividing the sample in $N_{\rm jack} = 12$ subvolumes, omitting one at a time in the calculation, and using equation (2.7). We will focus here only on the diagonal terms of the covariance matrices, σ_i. This is because we expect the estimation of the full covariance matrix with only 12 subsamples to be too noisy to make a fair comparison. We compare this jackknife error estimation to the standard deviation of the results for $\xi_{\rm dep}(r)$ obtained from a set of N independent realisations of the segment Cox process. We consider in this case that these are the 'real' errors we are trying to estimate using the jackknife method.

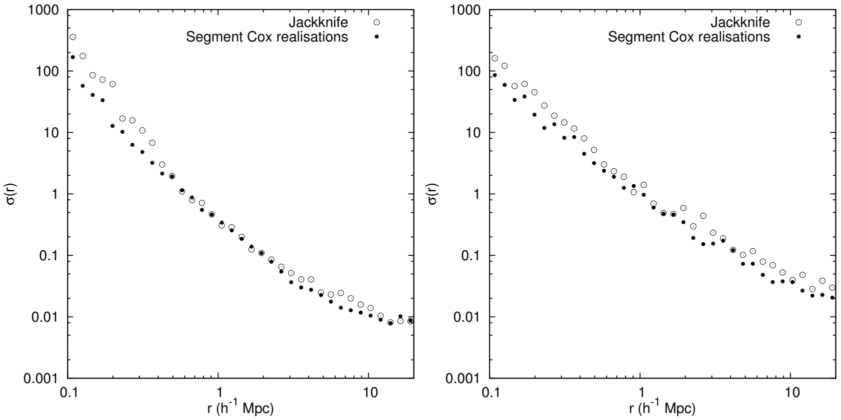
In Fig. 2.11 we show the comparison of the jackknife error estimate for one realisation to the 'real' errors obtained from $N = 12$ realisations. We see how the jackknife method recovers quite accurately the real uncertainty in this case. In some cases, specially for smaller scales or larger photometric redshift errors, the jackknife method slightly overestimate the errors, but it does not underestimate them in any case. This could explain the fact that, both for the segment Cox process and the haloes simulation, we obtain larger values for $\widehat{\Delta\xi}$ than for $\Delta\xi$ (see Tables 2.1 and 2.3).

2.5 Discussion

We have shown the reliability of recovering the real-space two-point correlation function from photometric redshift surveys. We have used light-cone simulations to produce mock catalogues that have been distorted by randomising along the line of sight the object positions following Gaussian distributions with different variances similar to the associated nominal errors of the photometric redshift surveys $\Delta z/(1+z)$.

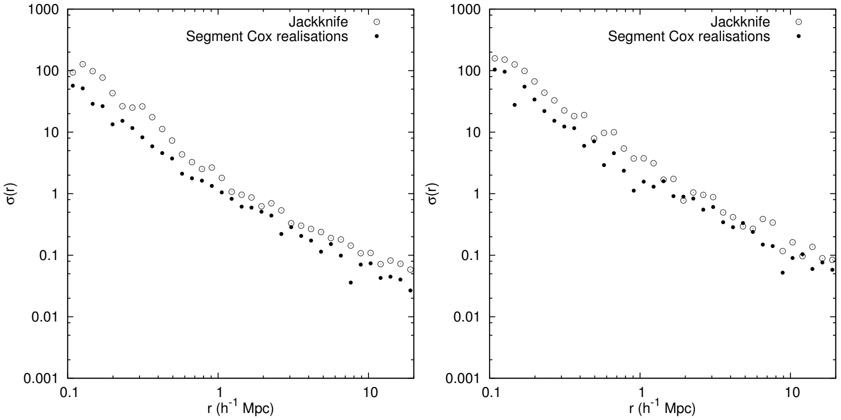
The method used to recover the real-space correlation function consists in obtaining the projected correlation function by integrating the two-dimensional correlation function along the line of sight. The projected correlation function is then deprojected assuming that redshift errors do not affect transverse distances.

The deprojection method applied on the distorted mock surveys provides quite



(a) Real space

(b) $\Delta z = 0.005(1+z)$



(c) $\Delta z = 0.015(1+z)$

(d) $\Delta z = 0.05(1+z)$

Figure 2.11: Comparison between the error estimation from the jackknife method and the variance of several realisations, for the segment Cox process mock catalogues

satisfactory results for recovering the real-space correlation function. We have quantified the quality of the recovering process as a function of the errors in the photometric redshifts. Our method was able to recover the real-space correlation function within a 5 per cent for $r < 10\,h^{-1}\,\mathrm{Mpc}$ from photometric catalogues with $\Delta z \leq 0.015(1+z)$. For larger redshift errors, the method is only valid (within a 7 per cent) for smaller scales, $r < 2\,h^{-1}\,\mathrm{Mpc}$. Hence, our method allows the extraction of useful information on the clustering of galaxies through the correlation function. That information can be used for the estimation of cosmological parameters based on data from photometric redshift surveys.

As an additional test, we used the deprojection method to recover the real-space correlation function from a realisation of a segment Cox process with simulated redshift errors. This stochastic model has a known analytical expression for the correlation function. This new catalogue has density and two-point correlation function similar to the ones in the simulations, but the higher-order correlation characteristics are very different. The results of the deprojection method are similar, indicating that it can be a general method to use independently of the details of the galaxy distribution. Finally, we checked that the jackknife method is an acceptable way to estimate the uncertainties on the deprojected correlation function in this case.

In the next chapter, we apply this method to the calculation of the real-space correlation function for samples drawn from the ALHAMBRA Survey. This shows the utility of this method for the analysis of large-scale structure using photometric redshifts.

3

The evolution of galaxy clustering since $z = 1.5$ in the ALHAMBRA Survey

The clustering of galaxies is an important tool for the study of both the large-scale distribution of matter, and its relation to galaxy formation and evolution. As these phenomena evolve with cosmic time, it is important to compare clustering measurements for local $z \simeq 0$ samples to measurements at higher redshifts.

As explained in the previous chapters, the two-point correlation function $\xi(r)$ is a simple tool that provides very useful information about the characteristics of the galaxy clustering pattern. For samples at low redshift, $\xi(r)$ is known to follow approximately a power-law behaviour, over a large range of scales, from hundreds of kpc to tens of Mpc (Totsuji and Kihara, 1969; Peebles, 1974; Martínez, 1999). However, studies with recent larger surveys have found significant deviations from this behaviour (see, e.g., Zehavi et al., 2004), which are fully consistent with the halo model of clustering.

One of the main problems (or opportunities) for the extraction of information from galaxy clustering is that galaxies are biased tracers of the matter density field (see Section 1.2). Galaxy bias depends on several properties of the galaxies (see Chapter 4). Observationally, the clearest segregation observed is that due to luminosity (Hamilton, 1988): bright galaxies are more strongly clustered than faint ones. This can be understood in the standard hierarchical structure formation theories from the fact that bright galaxies form at more massive dark matter haloes. This effect has been clearly observed in galaxy surveys of the local universe (see, e.g. Norberg et al., 2001; Tegmark et al., 2004) for galaxies with luminosities $L \gtrsim L^*$. This luminosity segregation, however, does not affect significantly galaxies with $L < L^*$.

Recently, these studies of galaxy clustering and luminosity segregation have been extended to higher redshifts, using state-of-the-art pencil-beam spectroscopic surveys, such as the VIMOS-VLT Deep Survey (VVDS, Pollo et al., 2006), the Deep Extragalactic Evolutionary Probe survey (DEEP2, Coil et al., 2006), or the zCOSMOS survey (Meneux et al., 2009). These works analysed the clustering in several galaxy samples extending up to $z \simeq 1.2$. Overall, the correlation function measurements were well fitted by power laws. Luminosity segregation is clearly detected, and in fact the dependence of clustering on luminosity is found to be stronger in these cases than in local samples.

In this chapter, we use preliminary data from the Advanced Large, Homogeneous Area Medium-Band Redshift Astronomical (ALHAMBRA) Survey (Moles et al., 2008) to study galaxy clustering for redshifts up to $z = 1.5$. ALHAMBRA is a deep photometric survey, which uses a total of 23 optical and near-infrared (NIR) bands in order to obtain accurate and reliable photometric redshifts for a large number of objects, in a total area (when finished) of $4\,\mathrm{deg}^2$. It is a survey specially suited for the study of the large-scale distribution of galaxies at high redshifts, given its photometric depth and area covered, which improve on those of similar spectroscopic surveys. Its main drawback is obviously the use of photometric redshifts, which may affect the clustering measurements. Our main aim is therefore to test the de-projection method for the recovery of the real-space correlation function presented in Chapter 2. In this way, we explore the possibilities of ALHAMBRA to study the evolution of galaxy clustering and its dependence on luminosity up to high redshifts, $z = 1.5$.

In Section 3.1, we explain in detail the characteristics of the ALHAMBRA Survey, and of the preliminary catalogue that we use here. Then, in Section 3.2, we select a catalogue suitable for galaxy clustering studies. This includes building the angular masks describing the selection function of the survey, eliminating possible stars in the catalogue, and selecting objects with a reliable measure of redshift. In Section 3.3 we present our correlation function calculations, based on the de-projection method of Chapter 2, for a set of samples selected in absolute magnitude, in three redshift bins. We fit the different correlation functions by two simple models in order to extract information regarding clustering evolution and luminosity segregation. Finally, in Section 3.4 we discuss our results, and compare them to those obtained for different surveys.

3.1 The ALHAMBRA Survey

The ALHAMBRA survey is a photometric survey which will cover a total of $4\,\mathrm{deg}^2$ in the sky, using 20 medium-band filters in the optical range, and three standard broad-band filters (J, H, and K_s) in the NIR. The aim of the survey is to study cosmic evolution in a broad sense, by providing an inventory of the contents of the Universe through a large fraction of cosmic history. The survey was designed, therefore, to provide relatively accurate redshift estimates and spectral classification for the different objects and, at the same time, to sample a statistically significant volume at different redshifts. The survey is being carried out using the 3.5-meter telescope at the *Centro Astronómico Hispano-Alemán* (CAHA)[1] in Calar Alto (Almería, Spain). The camera used for the optical observations is the Large Area Imager for Calar Alto (LAICA)[2], and Omega-2000[3] is used for the NIR observations.

The expected photometric depth of the survey is $AB \leq 25$ (for a point source with $S/N = 5$) for the optical filters bluer than $\simeq 8500$ Å, with the depth decreasing toward the red, and reaching $K_s \leq 23$ (all magnitudes are in the AB system). The use of a large set of medium-band filters places ALHAMBRA half-way between the classical spectroscopic and broad-band photometric types of surveys. It reaches deeper in magnitude than previous spectroscopic surveys, even the deepest ones as VVDS or DEEP2. This means that the density of objects is higher, and it is able to sample a fainter region of the luminosity function. Regarding broad-band photometric surveys, the main advantage of ALHAMBRA is the improved accuracy for the determination of redshift and spectral type provided by the use of a large number of filters over the optical and NIR spectral range.

Observations for the ALHAMBRA survey were carried out between 2004 and 2011, and are now completed. The preparation of the final catalogue of the survey, including the determination of photometric redshifts, is now underway, and it is expected to be finished during 2012. Therefore, in this thesis we use the preliminary data contained in the Internal Data Release 3 (IDR3), described in Section 3.1.1.

The ALHAMBRA filter system

The possibility of performing a survey using a large set of filters, in order to obtain a kind of very low resolution spectrum for each object was first discussed by Hickson et al. (1994). A similar idea was implemented in the surveys Calar Alto Deep Imaging Survey (CADIS, Meisenheimer et al., 1998), and Classifying Objects by Medium-Band Observations in 17 filters (COMBO-17, Wolf et al., 2003). These surveys used different combinations of broad-, medium- and narrow-band filters in the optical range.

The optical filter system for the ALHAMBRA survey was specifically designed to optimise the output of the survey in terms of photo-z accuracy, and number of objects with reliable z determination, as shown in Benítez et al. (2009b). The chosen system consists of a set of 20 contiguous, equal-width, medium-band filters covering the full optical spectrum, between 3500 and 9700 Å. The width of each

---

[1] http://www.caha.es

[2] http://www.caha.es/CAHA/Instruments/LAICA

[3] http://www.caha.es/CAHA/Instruments/O2000

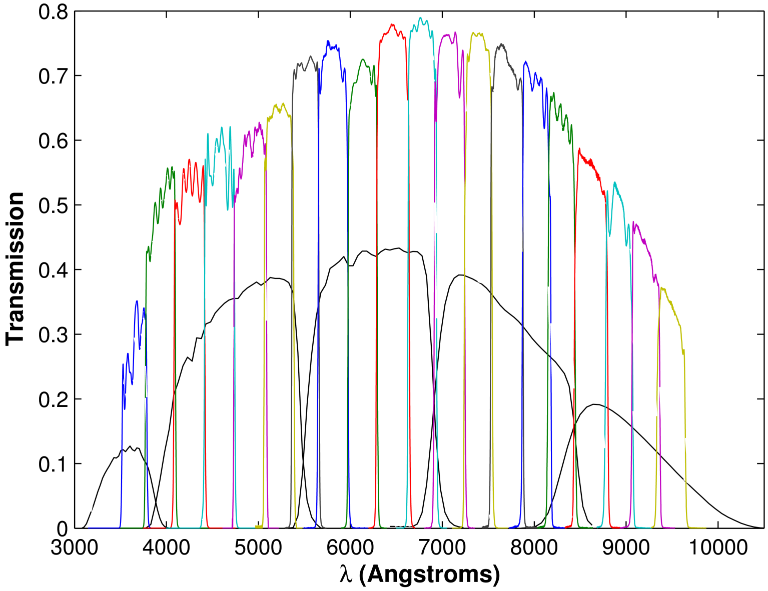


Figure 3.1: Response functions for the ALHAMBRA photometric system filters (colour lines). These response functions include the detector transmission, and atmospheric transmission at 1.2 air masses. These response functions are compared to the standard SDSS filters *ugriz* (broad-band filters in black lines). Figure from Aparicio Villegas et al. (2010).

of these filters is $FWHM \simeq 310$ Å. The transmission curves for these filters is shown in Fig. 3.1. This filter configuration also provides a homogeneous spectral coverage for a large range in wavelength. In this way, we minimise variation in the selection functions of the different objects with redshift. Aparicio Villegas et al. (2010) characterised in detail the ALHAMBRA optical photometric system, and also provided a set of transformation equations between this system and the SDSS filter system.

The survey is complemented by observations in the standard NIR filters J, H and K_s. The fact of complementing the optical observations with these three NIR filters is important for two reasons. On one side, to avoid the confusion between the Lyman and Balmer breaks that appears frequently when measuring photometric redshifts. This confusion creates a degeneracy between galaxies at redshifts $z \sim 0.5$ and $z \sim 3$, which affects the accuracy of the photo-z determinations (Moles et al., 2008). On the other side, the NIR observations provide valuable information about the old stellar population in galaxies, as opposed to the information about recent stellar formation that one can get from bluer wave-

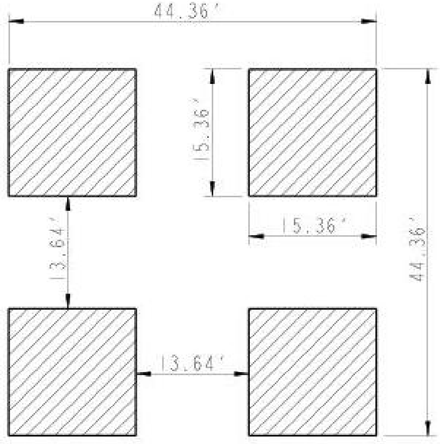


Figure 3.2: Configuration of the four CCDs of LAICA in the focal plane. Dimensions are shown in arc minutes. Figure from LAICA's webpage (http://www.caha.es/CAHA/Instruments/LAICA).

lengths. In this way, these observations provide information about parameters such as the mass of each galaxy.

Survey area and geometry

The total survey area of $4\,\mathrm{deg}^2$ is distributed over 8 widely separated fields in the sky. In this way, the effect of cosmic variance is minimised by measuring at independent volumes, while contiguous areas large enough to sample transverse scales up to tens of Mpc are still covered. The main criterion for the selection of fields was their low extinction, and they were chosen so that 7 out of the 8 fields have a significant overlap with other surveys (see Table 3.2 below).

The geometry of each of the surveyed fields is imposed by the geometry of the LAICA camera. As shown in Fig. 3.2, LAICA consists of four $4k \times 4k$ CCDs arranged in a 2×2 mosaic, with the gaps between them being approximately the same size as the CCDs. In this way, a single pointing produces images for four separated frames of $\sim 15' \times 15'$, and using four pointings a contiguous area of $1\,\mathrm{deg}^2$ can be covered. In the case of ALHAMBRA, two pointings are observed for each field. This results in 8 observed frames, corresponding to each of the LAICA CCDs, forming two contiguous strips of approximately $1^\circ \times 15'$ covered per field, adding up to the total $4\,\mathrm{deg}^2$ for the full survey. The field of view of the

Omega-2000 camera is equivalent to one of the LAICA CCDs. When referring to images corresponding to a given frame of the survey, we use the notation fA_pB_C, where $A = (1, \dots, 8)$ refers to the field, $B = (1, 2)$ refers to the pointing within this field, and $C = (1, \dots, 4)$ refers to the frame within this pointing. In total, 64 frames will be observed to complete the survey.

This field geometry will influence the way in which we measure galaxy clustering. In Section 3.2.1 below, we describe in detail the effective geometry of the survey taking this observing strategy into account.

3.1.1 The Internal Data Release 3 (IDR3)

In this work, we used the preliminary catalogues from the Internal Data Release 3 (IDR3), which contains data for frames that had been observed in the 23 bands, and fully processed by March 2011. In total, this release contains data for 39 frames, out of the total 64 planned for the full survey. These are distributed in three complete fields (fields 2, 7, and 8), and four partially completed fields (fields 3, 4, 5, and 6). Thus, the nominal area covered in IDR3 is $2.4375\,\mathrm{deg}^2$, and the total number of detected sources included in the catalogue is 597724. However, we only consider those sources for which there is a redshift determination (as explained below). This makes a total of 501868 objects in the catalogue.

The data reduction and the preparation of the catalogues was carried out by the ALHAMBRA team and is still preliminary, but it will be very similar to that for the final ALHAMBRA catalogues. The two main points relevant for our analysis are the way in which objects are detected for inclusion in the catalogue (Husillos et al., in prep.), and the way in which the photometric redshifts are estimated (Molino et al., in prep.).

Detection of objects

The fact that the ALHAMBRA survey performs observations in 23 bands (20 medium band filters in the optical, and 3 in the NIR) makes the definition of the object detection quite complex. A special strategy was defined in order to use a deep image for detection, while not biasing the detection towards any special type of objects. For each of the 39 frames, a 'deep' image was constructed as the sum of a set of individual exposures. These exposures are those observed through the optical filters with largest efficiencies (those between A457M and A829M, included), and taken under the best observational conditions. Here, these conditions are defined by an atmospheric transparency better than 50%, and a seeing better than 1.2 arcsec. As conditions change, this means that, for each frame, the 'deep' image is a combination of a different number of exposures per filter. However, on average, this 'deep' image is similar to an image using the SDSS i filter, although with an extended wing towards the blue, entering into the r filter.

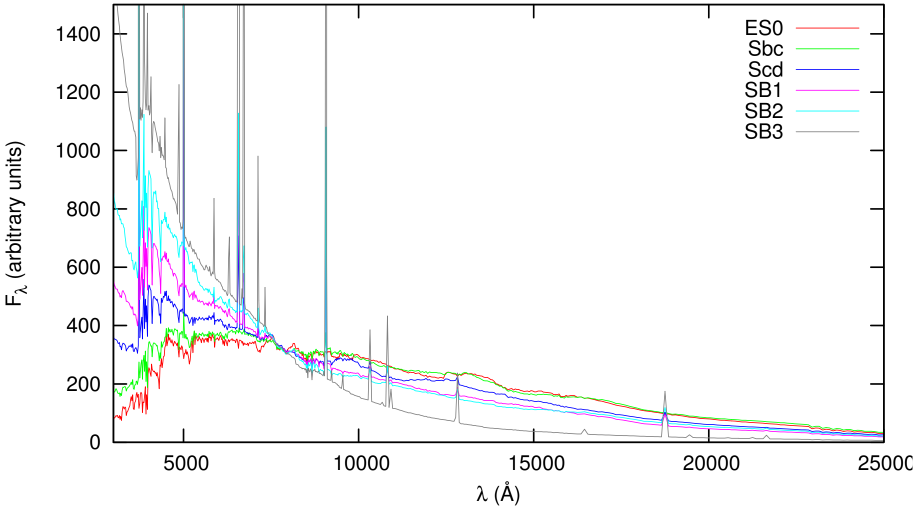


Figure 3.3: SED of the six galaxy templates used for the determination of photometric redshifts for ALHAMBRA, using the BPZ software. We plot the flux per unit wavelength, with the normalisation chosen so that all of them have the same magnitude in the I band.

Object detection is therefore performed in this artificial 'deep' image using the software SExtractor (Bertin and Arnouts, 1996), and then photometry is obtained for all objects in each of the filters in the standard way. The average depth obtained (for 3σ detections) in this catalogue is $AB \simeq 24.5$ for most of the optical filters, and $AB \simeq 22$ for the NIR filters (Stefanon, 2011).

Redshift determination

Photometric redshifts (photo-z) for the objects in the IDR3 catalogue were obtained using an updated version of the Bayesian Photometric Redshift (BPZ) software (Benítez, 2000). This method fits the observed photometry of the object to a library of template spectral energy distributions (SEDs) corresponding to different types of galaxies. In this way, it finds the z and type that best match the observations for each object. A prior probability on the expected distribution of types and redshifts as function of apparent magnitude is taken into account, in order to improve the accuracy of the redshift determination, and to minimise the number of 'catastrophic errors'. In the analysis used for IDR3, the template library is an interpolation between six model SEDs corresponding to E/S0, Sbc, Scd, and three starburst types of galaxies. Hence, the best-fit spectrum for a given object can be a weighted average of two neighbouring templates. The SED of the six templates used are shown in Fig. 3.3.

In order to improve the accuracy of the photo-z, the photometric zero-points in each band can be re-calibrated using the output of the photometric redshift determination (Coe et al., 2006). This process is carried out independently in each of the frames. In the cases in which enough spectra are available (11 out of the 39 frames), the spectroscopic redshifts $z_{\rm spec}$ of these objects are used for the re-calibration. In the rest of the cases, the re-calibration is done using purely the photometric redshifts $z_{\rm phot}$ obtained. The IDR3 catalogue contains the redshifts obtained with the best zero-point calibration available in each case. A preliminary comparison with the ~ 3600 spectra available in the area covered by IDR3 shows a typical error of $\Delta z \simeq 0.012(1+z)$ when the re-calibration is performed using $z_{\rm spec}$, and of $\simeq 0.014(1+z)$ when using only $z_{\rm phot}$ (A. Molino, priv. comm.).

The Bayesian framework used allows us to not only obtain the best-fit redshift and galaxy type for each object, but also to obtain the full redshift probability distribution function $p(z)$. It is however not feasible to store and use the full $p(z)$ of all galaxies in the analysis, so the catalogue contains only a set of parameters to characterise it. The first of these parameters is the mode of the distribution, which is the best estimate for the redshift of the object, and we call simply z.

The 'odds' parameter $p_{\rm odds}$ gives the probability that the redshift is contained within a distance $\pm 0.2(1+z)$ from the best estimation. This parameter is therefore a reliable estimation of the quality of the redshift determination, as it gives a measure of how concentrated the obtained $p(z)$ is. In general, selecting objects with values of $p_{\rm odds}$ close to 1 corresponds to selecting objects with high-quality redshifts. The parameter $p_{\rm odds}$ is a better estimator of the quality of the redshift determination than the usual χ^2 of the fit of an objects's photometry to the template spectra. This is because the model used (the set of template spectra) does not represent in detail the variety of galaxy spectra in the real world. Therefore, it often happens that the fit to the model obtained for a galaxy is bad (it has a very large value of χ^2), but the determination of its redshifts is actually accurate (i.e., $p(z)$ has a high peak at the right redshift). This is usually the case for bright galaxies, where the photometry errors are small, and hence the differences between the model spectrum and the actual galaxy spectrum are significant. This is the reason why we do not use the χ^2 parameter as indicator of photometric redshift quality.

Finally, the parameters $z_{\rm max}$ and $z_{\rm min}$ define an interval around z containing 68% of the probability. In the case in which $p(z)$ is a Gaussian, therefore, its standard deviation is given by

$$\sigma_z = \frac{z_{\rm max} - z_{\rm min}}{2} \,. \tag{3.1}$$

Although $p(z)$ is not Gaussian in general, this can be a good approximation when we restrict the analysis to high-quality photo-z (see e.g. Fernández-Soto et al., 2001; Ilbert et al., 2009). Hence, we will use σ_z defined above as measure to characterise the accuracy of the redshift determination in the different samples used below in our analysis.

3.2 Selection of a catalogue for large-scale structure studies

Based on the IDR3 data, we prepared a galaxy catalogue suitable for the study of the large-scale structures (LSS). The reason for this is twofold. On one side, the original IDR3 catalogue contains all the objects which passed basic detection criteria in the 'deep' images. This means that a large fraction of these are spurious objects due to noise in regions with low exposure times, or to imaging defects (e.g. fragmentation of bright saturated stars). Moreover, from the real objects of the catalogue, we should then select those objects which are useful for the LSS analysis, i.e., galaxies with a good photometric redshift determination, thus eliminating the stars and galaxies with low quality photo-z. On the other side, in order to make LSS studies, such as the calculation of the two-point correlation function, we need to characterise the selection function of the survey.

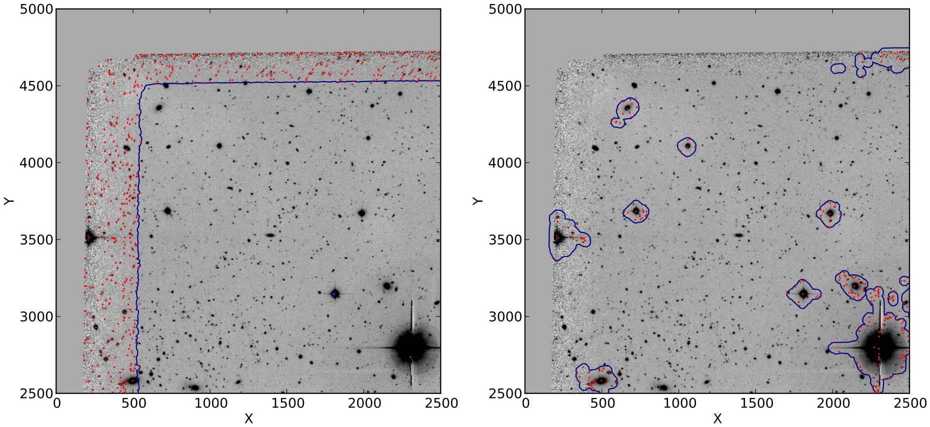
In the first place, we describe how we characterised the angular selection function of the survey, and eliminate the objects outside of it. We do not consider here the radial selection function of the survey, but will model it for each particular sample used in the analysis, as described in Section 3.3.2. We also describe how we perform the star-galaxy separation in the catalogue, and our criterion for the selection of objects with good redshift determination.

3.2.1 Angular selection function and survey masks

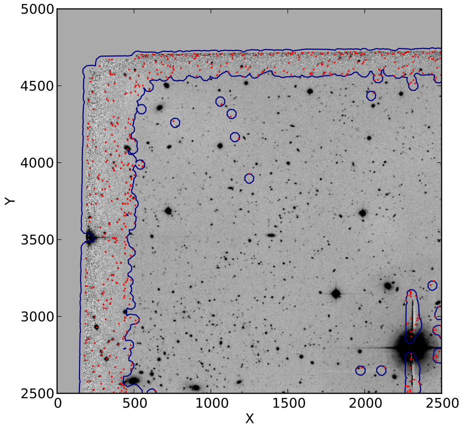
We performed a basic characterisation of the angular selection function of the survey based on the object detection procedure. We use here a basic selection function, in the form of an angular mask which describes only which areas in the sky have been reliably observed by the survey and which have not. However, we do not study variations in completeness between different areas inside the mask.

Our approach is to focus in the object detection procedure (see Section 3.1.1 above), and to identify areas of the images in which there are potential problems for a correct detection of extra-galactic objects. Those identified areas are left out of the survey mask, and those objects outside of the mask are eliminated from the catalogue. Although most of these objects are spurious or not very reliable detections, a fraction of them are actually 'good' objects. Hence, a compromise is needed when choosing the actual parameters that define the mask.

The actual criteria used for the definition of the mask are of two types. The first is related to the elimination of areas of the 'deep' image with low exposure time, and the second to the identification of defects in the image, or extended



(a) ‘Weights mask’ **(b)** ‘Bad objects mask’



(c) ‘Negative mask’

Figure 3.4: A region of the ‘deep’ image corresponding to the frame ‘f02p01_1’, illustrating the process to define the angular mask of the survey. It corresponds to a quarter of the CCD, covering $\sim 8 \times 8$arcmin. In each of the images, we show one of the partial masks combined to get our final mask: the mask based on the ‘weight image’ (a), the mask for objects which are either saturated or too large (b), and the mask for regions with negative values (c). In all three cases, the border of the mask is the blue line, and the red dots are objects in the catalogue which are excluded by the mask.

Pixel weight threshold	0.75
Object area threshold	$3000\,\mathrm{pix}^2 = 147.59\,\mathrm{arcsec}^2$
Object length threshold	$400\,\mathrm{pix} = 88.72\,\mathrm{arcsec}$
Extra border	$60\,\mathrm{pix} = 13.31\,\mathrm{arcsec}$

Table 3.1: Parameters used in the definition of angular masks (see the text for details).

objects, which may affect the detection of near objects. We illustrate the different steps in the construction of our angular mask for a region of the frame 'f02p01_1' in Fig. 3.4.

The first criterion is based on the 'weight image' associated to each of the 'deep' images. This image contains a weight for each pixel which is proportional to the total exposure time at this pixel of the combination used to build the 'deep' image. We normalise the weights in each frame to the maximum of the weight image. We then select only the regions of the image where this pixel weight is larger than 0.75. That is, we do not consider in our masks those regions which were observed (for the 'deep' image used for object detection) for less than the 75% of the maximum exposure time. The main effect is that, due to the dithering in the observation of the different exposures, we eliminate the areas around the borders of the image (see Fig. 3.4a).

In order to apply the criteria related to particular defects or objects, we used the SExtractor software. We performed a standard object detection in each of the 'deep' images (with the default parameters) using SExtractor. We then selected a series of objects that may cause spurious or incorrect object detections. We masked out all objects which were flagged as saturated by SExtractor, and also those with an area larger than $3000\,\mathrm{pix}^2$, or with a major axis larger than 400 pix (see Fig. 3.4b). Objects with large area may correspond to bright stars, but also to nearby galaxies. In either case, they would affect the detection of far extra-galactic objects in the region around them. The very elongated objects would normally correspond to defects in the image (e.g. cosmic rays not corrected for), or spikes created by bright saturated stars. Finally, we also masked all objects detected in the inverted image, which correspond to regions of the image with negative values found next to some saturated stars, or near the border of the image (see Fig. 3.4c). In order to eliminate the effect of these 'bad' objects in the neighbour regions, we also masked out an extra border of 60 pix around each of them.

Table 3.1 contains a summary of the parameters used when defining our survey angular mask for ALHAMBRA, and Fig. 3.5 illustrates the final mask considered, for the same region shown in Fig. 3.4. Clearly, the choice of these parameters is somehow arbitrary, and it depends in part on the visual inspection of the final masks and 'deep' images, as in Figs. 3.4 and 3.5. For example, we see that the border areas eliminated using the 'weights mask' (Fig. 3.4a) and 'negative mask' (Fig. 3.4c) are similar. This indicates that our choice of the pixel weight threshold

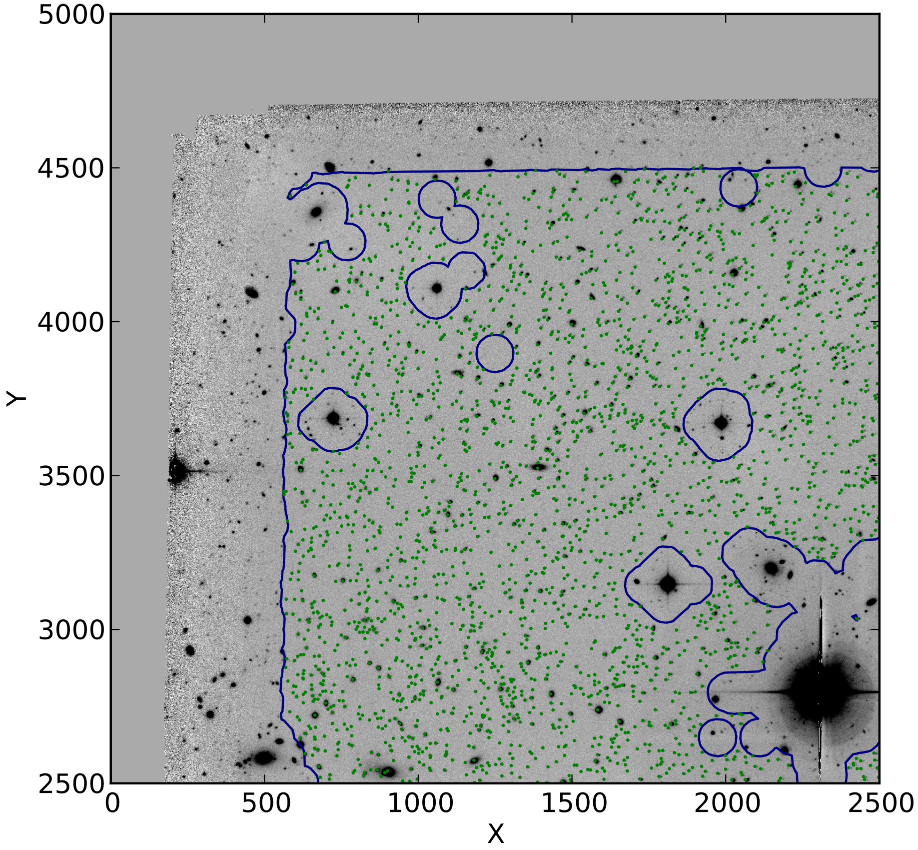


Figure 3.5: We show the same region of the 'deep' image for 'f02p01_1' as in Fig. 3.4, together with the final angular mask used. The border of the mask is the blue line, and the green dots are the positions of the objects inside the mask.

Field	Other survey	No. of frames	A (deg^2)	$N_{\rm mask}$	$N_{\rm mask}/N_{\rm tot}$
ALHAMBRA-2	DEEP2	8	0.389	76618	76.2%
ALHAMBRA-3	SDSS	4	0.198	37113	74.6%
ALHAMBRA-4	COSMOS	4	0.190	40827	81.6%
ALHAMBRA-5	HDF-N	4	0.208	41176	79.5%
ALHAMBRA-6	GROTH	3	0.155	28379	78.2%
ALHAMBRA-7	ELAIS-N1	8	0.377	77440	73.1%
ALHAMBRA-8	SDSS	8	0.360	74622	69.3%
Total		39	1.877	376175	75.0%

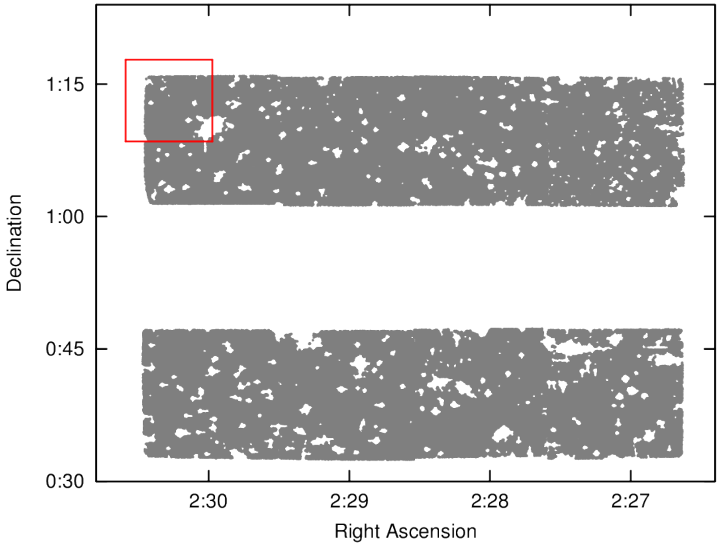
Table 3.2: Properties of the angular masks obtained for the different ALHAMBRA fields. For each field, we list the number of frames used, the effective area of our mask A, the number of objects we select inside the mask $N_{\rm mask}$, and the fraction of these selected objects with respect to the total number of objects in the catalogue $N_{\rm tot}$. We also list other surveys which have overlap with each of the fields, see Moles et al. (2008) for details.

(0.75) is reasonable, as it corresponds to areas where the noise is important and results in a large number of pixels with negative values.

This set of parameters is a conservative choice, driven by our interest on only using regions with a high level of completeness. Using a more relaxed choice would increase the effective area used. For example, using 0.70, 4000 pix^2, 500 pix, and 40 pix, respectively, for the parameters listed in Table 3.1, the final effective area is increased by a $\sim$ 10%. In any case, the details of the construction of the angular mask will be revised when using the final ALHAMBRA catalogues, depending on their characteristics.

Once we had defined the angular masks for each of the frames, we converted them to the 'polygon' format of the MANGLE software (Swanson et al., 2008). This software allows an easy manipulation of the masks, and also provides useful routines to perform tasks such as calculating the effective area covered by the masks, or generating random Poisson catalogues of points inside them. We combined the masks of the different frames into 7 masks, one for each of the fields. In the case of the three completed fields in IDR3, the resulting mask is composed by 2 strips of approximately $1^\circ \times 15'$ each, while in the rest of the fields (with just one LAICA pointing completed), the frames have gaps in between them. We avoided overlaps between frames by simply cutting the masks where appropriate, but the overlap regions were minimal anyway. As an illustration, Fig. 3.6 shows the final masks used for a completed field (field 2), and a field with only one completed pointing (field 3).

The final effective area of the survey mask for IDR3 is 1.877 deg^2, which means that we are using a $\sim$ 77% of the nominal area of 2.438 deg^2 for the 39 frames. We eliminated from our catalogue those objects located outside of our mask, the



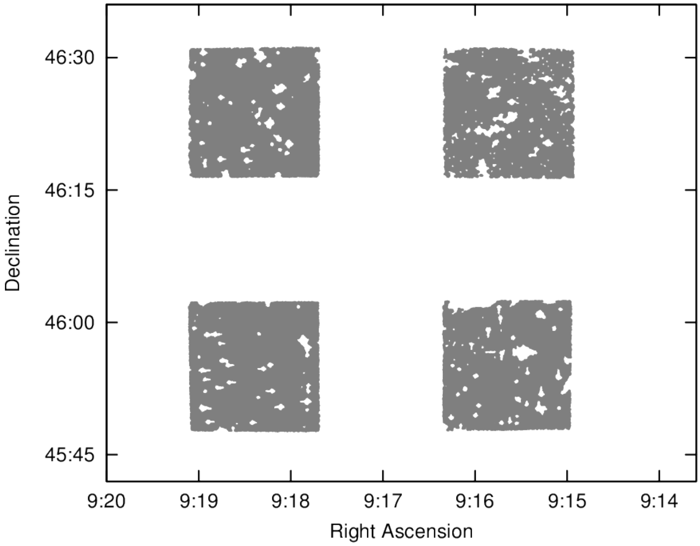


Figure 3.6: Angular masks for the ALHAMBRA fields 2 (top) and 3 (bottom), which are illustrative of the masks obtained for fields with two or one completed pointings, respectively. The shaded area corresponds to the regions of the survey that are included in the calculations. The red rectangle marks the region which is shown in Figs. 3.4 and 3.5.

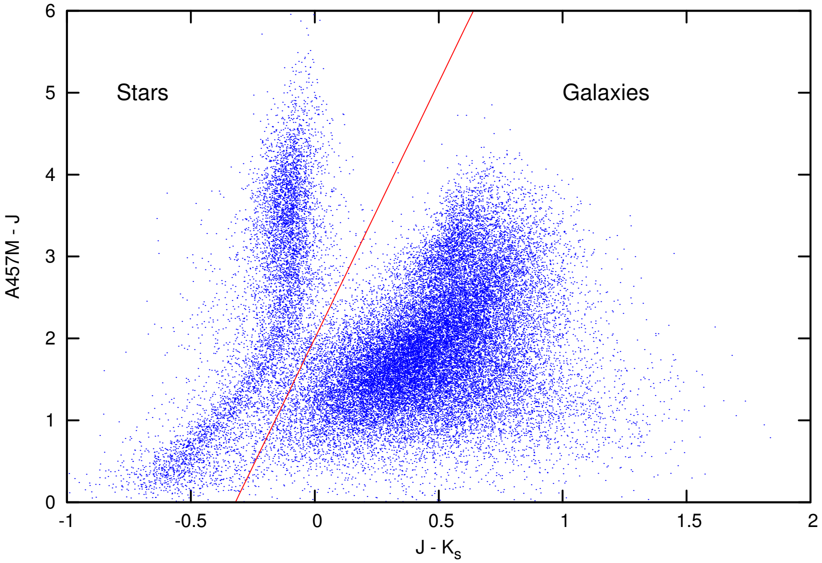


Figure 3.7: Colour-colour diagram used for our star-galaxy separation in the ALHAMBRA catalogue. For clarity, only objects with magnitude errors smaller than 0.1 are shown. The line corresponds to the cut we use for the separation, equation (3.2).

25.0% of the total. This left us with 376175 objects inside the mask. The mask area and number of objects in each of the fields is listed in Table 3.2.

3.2.2 Star-galaxy separation

In order to build a catalogue for LSS studies, we need to eliminate the stars present in the original catalogue. In order to do so, we used a separation in colour, similar to the usual BzK method (Daddi et al., 2004). Adopting this method to the ALHAMBRA photometric system, we made a cut in the (A457M $-$ J) vs. (J $-$ K_s) diagram, shown in Fig. 3.7. In particular, we classified as stars objects with

$$(J - K_s) \leq 0.16(\text{A457M} - J) - 0.32\,. \tag{3.2}$$

This cut has been shown to work well to separate stars from galaxies in ALHAMBRA data by Stefanon (2011). This cut eliminates 86704 stars from our sample (the 23.0%), hence we are left with 289471 galaxies in our catalogue.

The distribution of redshifts for the selected galaxies is shown in Fig. 3.8 (in grey). The mean of this distribution is $\bar{z} = 1.143$, and the median is $z_m = 0.972$.

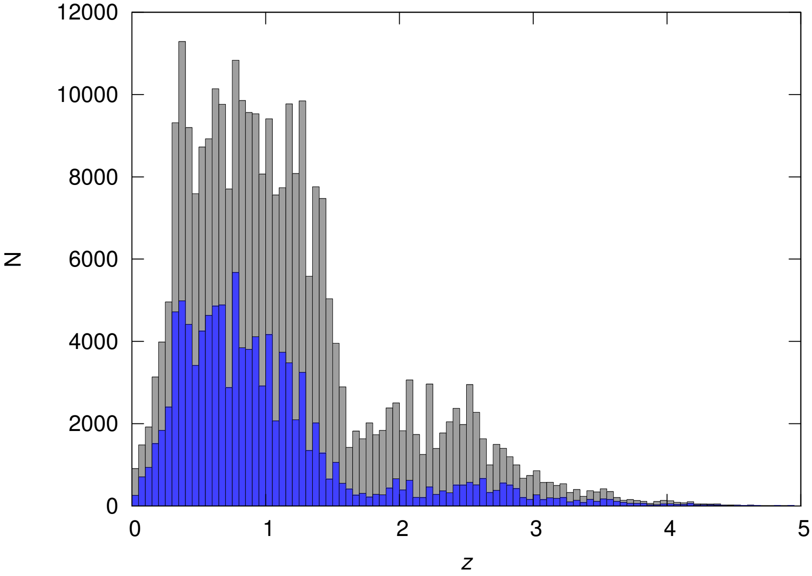


Figure 3.8: Histogram showing the distribution of the measured redshifts for the 289471 galaxies selected using relation (3.2) (grey), and for the 106713 objects in our final LSS catalogue, after making the redshift quality selection $p_{\rm odds} > 0.85$ (blue). The width of the bins is 0.05 units. The mean of the distributions for the full galaxy catalogue is $\bar{z} = 1.143$, and the median is $z_m = 0.972$, while for the final LSS catalogue, these values are $\bar{z} = 0.997$ and $z_m = 0.810$.

Although the bulk of this population is located approximately in the region $0.3 < z < 1.5$, there is a large secondary bump at $1.5 \lesssim z \lesssim 3$.

3.2.3 Redshift quality selection

In order to study the LSS using photometric redshift catalogues, it is essential to select samples with high quality redshift determination. This means both a small average error of the redshifts, and a small number of outliers. An example of this need are the results obtained in Chapter 2 for the de-projection of the correlation function.

In the case of redshifts determined by BPZ, the parameter $p_{\rm odds}$ provides a reliable way to select redshifts of high quality (Benítez, 2000). The distribution of this parameter in our catalogue (after the star-galaxy separation above) is shown in Fig. 3.9. The choice of the threshold value to make a selection should be a compromise between getting the objects with the best redshift quality, while

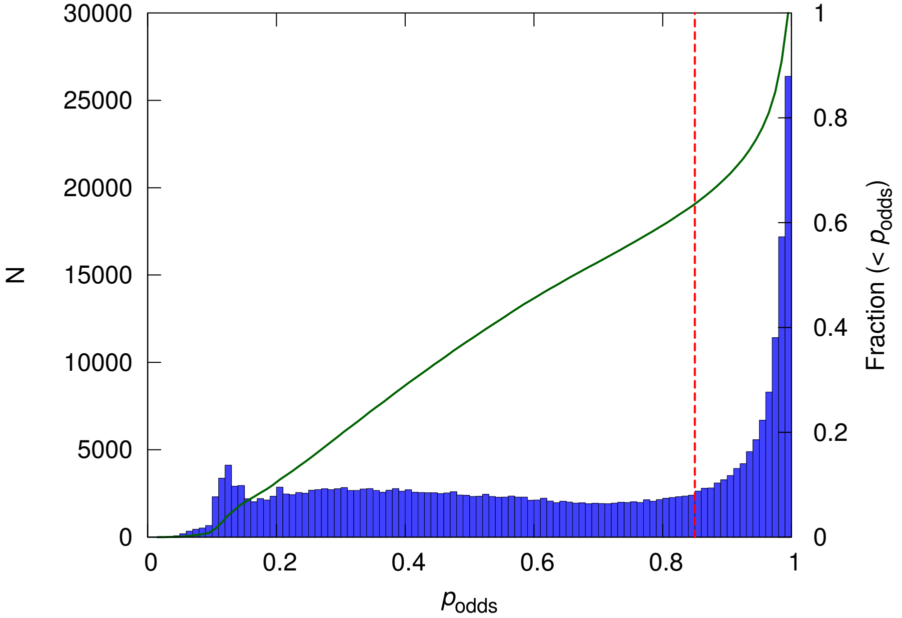


Figure 3.9: Histogram showing the distribution of the parameter $p_{\rm odds}$ in the catalogue. The width of the bins is 0.01 units, and the histogram contains data for a total of 289471 galaxies, after the star-galaxy separation. The green continuous line gives the cumulative distribution, to be read from the right vertical axis. The red vertical dashed line corresponds to the threshold $p_{\rm odds} = 0.85$ used for our selection.

retaining a large enough number of objects in our sample. However, the fact that the histogram has a pronounced peak for large values of $p_{\rm odds}$ means that the result will not depend much on the exact threshold chosen.

We use σ_z defined as in equation (3.1) as a measure of the standard deviation of the posterior probability distribution of z, and therefore as an estimate of the error on z. We plot, in Fig. 3.10, the σ_z obtained for each object as function of the $p_{\rm odds}$ parameter. As expected, σ_z decreases, on average, with $p_{\rm odds}$. Moreover, we see that there is a population of objects that are far from the main locus, with values as high as $\sigma_z \simeq 2$. These correspond to objects whose redshift determination is very uncertain, due to their $p(z)$ being very wide or having more than one peak. A large fraction of them will be outliers, in the sense that the best z estimation from BPZ will be far from their real redshift. In order to avoid this type of objects, we select only objects with $p_{\rm odds} > 0.85$. Making this selection, we have, for all

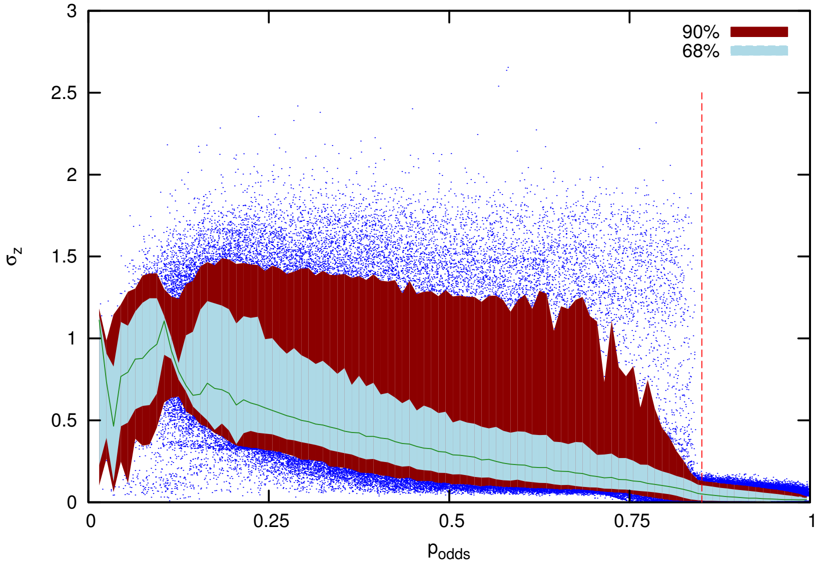


Figure 3.10: Distribution of the σ_z obtained according to equation (3.1), as a function of the $p_{\rm odds}$ parameter. The green continuous line shows the median σ_z for each value of $p_{\rm odds}$, and the shaded areas the symmetric regions containing 68% and 90% of the galaxies, as indicated. The remaining 10% of galaxies are plotted as blue dots. The red vertical dashed line corresponds to the threshold $p_{\rm odds} = 0.85$ used for our selection.

objects, $\sigma_z \leq 0.2$, and thus we minimise the possibility of having outliers in our sample.

Making this selection based on the $p_{\rm odds}$ parameter, we eliminate 63.1% of the objects previously selected, and we are left with 106713 objects with high redshift quality in our final LSS catalogue. In Fig. 3.8, we can see the effect of this selection on the redshift distribution. The main effect, apart from a global reduction on the number of objects, is that the secondary bump at $1.5 \lesssim z \lesssim 3$ is reduced greatly. This indicates that this bump at relatively high redshift was mostly due to confusion between different spectral features in the photometric redshift algorithm, but these problematic objects are correctly eliminated using our $p_{\rm odds}$ cut.

In Fig. 3.11, we study how our redshift quality selection affects the photometric accuracy of the sample. We show the distribution of errors in the magnitude A802M, corresponding to an ALHAMBRA filter close to the central wavelength of the i filter, and which we use below for the calculation of absolute magnitudes.

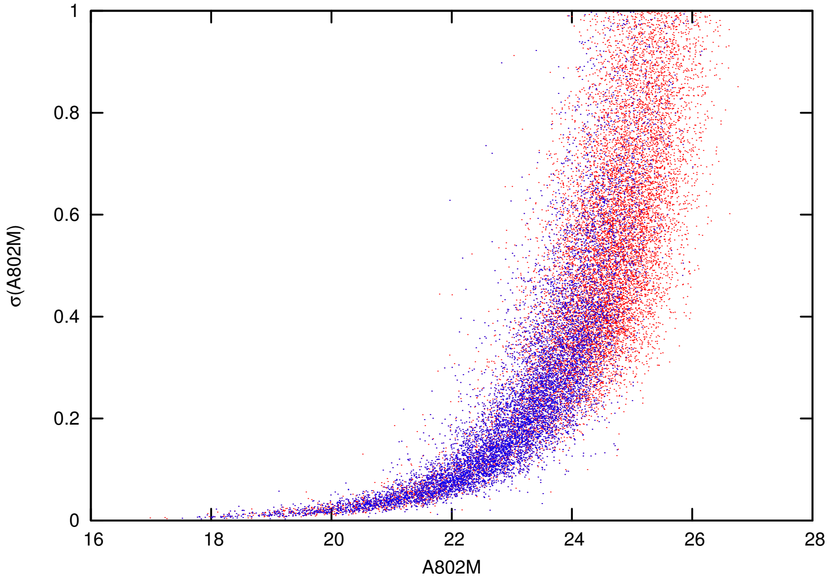


Figure 3.11: We show the magnitude vs. magnitude error diagram corresponding to the ALHAMBRA filter A802M, which is close to the central wavelength of the i filter. We show in the diagram the galaxies selected using relation (3.2) (red), and those in our final LSS catalogue, after the redshift quality selection $p_{\mathrm{odds}} > 0.85$ (blue). For clarity, only 10% of the points are shown in the plot.

We see how the our cut in p_{odds} eliminates most of the galaxies with the largest photometric errors, $\sigma(A802M) \gtrsim 0.3$.

3.2.4 Characteristics of the final LSS catalogue used

The final catalogue contains 106713 objects after the three selection steps explained above. The redshift distribution of the objects is shown in Fig. 3.8. The mean redshift of the catalogue is $\bar{z} = 0.997$, and the median redshift is $z_m = 0.810$. The bulk of the objects is located at $z < 1.5$, although there is still a significant population at $z \in [2, 3]$ and a small number of objects at higher z, up to the maximum $z_{\mathrm{max}} = 5.38$.

As we have already made a severe selection in order to restrict the catalogue to galaxies with high quality photometric redshifts, we can use σ_z, as defined by equation (3.1), as a measure of the redshift uncertainty for each object. We consider the relative uncertainty, $\frac{\sigma_z}{1+z}$, and show its distribution in the catalogue

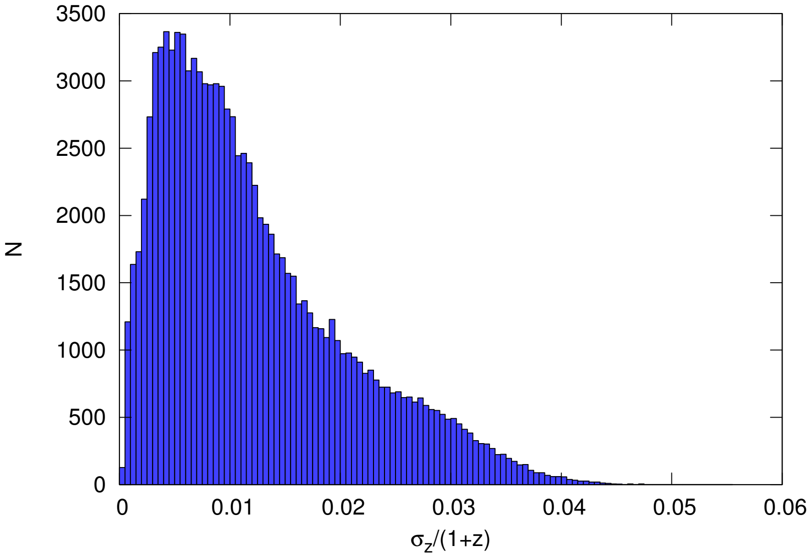


Figure 3.12: Histogram showing the distribution of the relative redshift uncertainties calculated from equation (3.1) for the 106713 objects in our final LSS catalogue. The width of the bins is 0.0005 units. The mean of this distribution is $\overline{\left(\frac{\sigma_z}{1+z}\right)} = 0.012$, and the median value is $\left(\frac{\sigma_z}{1+z}\right)_m = 0.010$.

in Fig. 3.12. The mean value for the final catalogue is $\overline{\left(\frac{\sigma_z}{1+z}\right)} = 0.012$, and the median value is $\left(\frac{\sigma_z}{1+z}\right)_m = 0.010$. These results are in line with the predictions for the redshift accuracy of the survey (Moles et al., 2008; Benítez et al., 2009b), and with the preliminary results obtained in comparisons with spectroscopic redshifts.

Distances, K-corrections and absolute magnitudes

We added to this final catalogue two quantities that we need later for the sample selection and LSS studies: the distance and absolute magnitude of each object.

We calculated the co-moving distance to each object using the best-fit photometric redshift z in each case, and assuming a flat fiducial cosmology with $\Omega_M = 0.27$ and $\Omega_\Lambda = 0.73$, as given by the WMAP 7-year results (Komatsu et al., 2011). We express all quantities in terms of h, e.g. we measure distances in units of h^{-1} Mpc, so we do not need an explicit value for it.

We also calculate, for each object, the corresponding absolute magnitude in the B band, M_B. We use here the B band as it allows a more direct comparison with results from other surveys. Moreover, the region of the spectrum corresponding to the B band in the rest-frame is well sampled by the set of filters used by ALHAMBRA (including the NIR filters) for redshifts as high as $z \sim 2$. The same procedure used here could be used, anyway, to calculate the absolute magnitude in any given filter. We calculate M_B in terms of the apparent magnitude m_A in one of the ALHAMBRA filters as (Hogg et al., 2002)

$$M_B = m_A - DM - K_{BA}\,, \tag{3.3}$$

where DM is the distance modulus, defined in terms of the luminosity distance d_L as

$$DM = 5\log_{10}\left(\frac{d_L}{10\mathrm{pc}}\right)\,, \tag{3.4}$$

and K_{BA} is the K-correction term (Humason et al., 1956; Oke and Sandage, 1968). This term accounts for the difference between the bandpass of the chosen ALHAMBRA filter in the observer's frame, and the bandpass of the B filter in the object's rest-frame. In the general case, K_{BA} is given by

$$K_{BA} = -2.5\log_{10}\left[\frac{1}{(1+z)}\frac{\int \mathrm{d}\lambda_0\lambda_0 L_\lambda\left(\frac{\lambda_0}{1+z}\right)T_A(\lambda_0)\int \mathrm{d}\lambda_e\lambda_e g_\lambda(\lambda_e)T_B(\lambda_e)}{\int \mathrm{d}\lambda_0\lambda_0 g_\lambda(\lambda_0)T_A(\lambda_0)\int \mathrm{d}\lambda_e\lambda_e L_\lambda(\lambda_e)T_B(\lambda_e)}\right]\,, \tag{3.5}$$

where $L_\lambda(\lambda)$ is the emitted-frame luminosity of the object per unit wavelength, $T_X(\lambda)$ is the transmission of filter X, and λ_e and λ_0 refer to wavelengths in the rest-frame and in the observer's frame, respectively. The function $g_\lambda(\lambda)$ gives the flux of the standard source defining the zero-point for the magnitude system. We use in all cases the AB magnitude system, so that $g_\lambda(\lambda) \propto \lambda^{-2}$.

One can only calculate exactly the K-correction when knowing the real emitted spectrum of the object, $L_\lambda(\lambda)$. We approximate it in our case using the best-fit spectrum determined by BPZ for each object during the photometric redshift determination (see Section 3.1.1). However, we can choose the ALHAMBRA filter m_A used for the calculation in a way that minimises the dependence of K_{BA} on the actual spectrum of the object. To this end, we choose $m_A \equiv$ A802M. This filter has an effective wavelength of $\lambda_{\mathrm{eff}} = 802\,\mathrm{nm}$ (Aparicio Villegas et al., 2010), which corresponds approximately to the effective wavelength of the B filter redshifted to the median redshift of the catalogue, $z_m = 0.810$. As explained above, this best-fit spectrum is in each case a combination of two neighbouring templates in the library. We recover this spectrum in this same way in each case, for the calculation of K_{BA} using equation (3.5). Fig. 3.13 shows the K-correction term for our choice of filters and for the six template spectra used. We see that, although our choice of the filter A802M minimises the dependence of K_{AB} on the selected template at $z \sim 0.8$, we see that there is still an important dependence for different redshifts, going up to maximum differences of ~ 1.5 mag at $z \sim 0.3$ or $z \sim 1.5$.

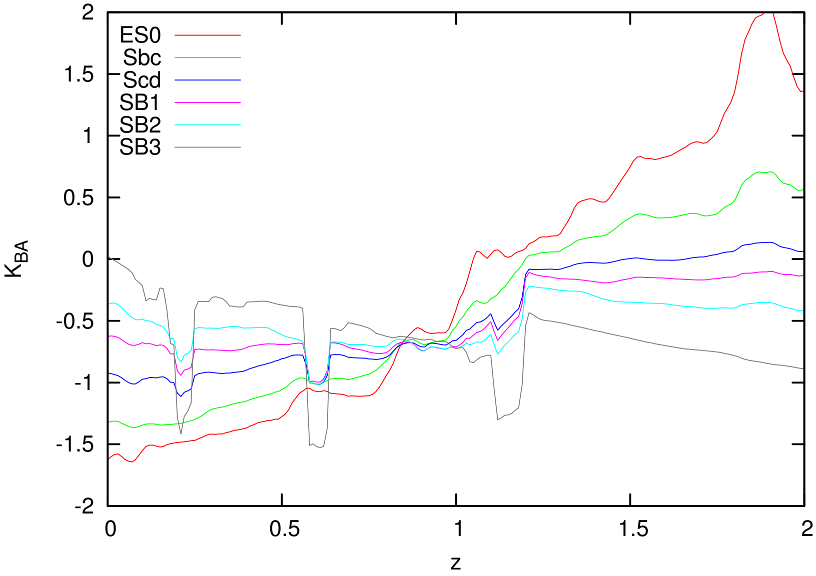


Figure 3.13: K-correction terms, as defined by equation (3.5), for the transformation from filter A802M in the observer's frame to filter B in the object's rest frame. They are calculated for the six BPZ template spectra shown in Fig. 3.3.

Therefore, using equations (3.3, 3.4, 3.5), we calculated the absolute magnitude M_B for the objects in the catalogue. We had to exclude 448 objects from the catalogue (the 0.4%) which do not have a measured value of the apparent magnitude A802M. As we use all the distances in units of h^{-1} Mpc, we actually calculate the quantity $M_B - 5\log_{10} h$.[4] We always refer to this quantity in the rest of this chapter, even when we drop the second term for simplicity.

3.3 Correlation functions for ALHAMBRA catalogues

The aim of this Section is to test the method for the calculation of the real-space correlation function described in Chapter 2 to real data from the ALHAMBRA survey, and to use the results obtained to study the evolution and luminosity dependence of galaxy clustering.

---

[4] For reference, $5\log_{10} h = -0.775$ for $h = 0.7$.

3.3.1 Selection of samples

In order to be able to study the dependence of clustering properties on both luminosity and cosmic time, we built a series of subsamples from our catalogue, by making a selection in redshift and absolute magnitude.

We selected subsamples in non-overlapping bins in redshift. The size of these bins is limited by the requirement that they are much larger than the distance we will integrate in the radial direction, $r_{\parallel,\mathrm{max}}$. We showed in Chapter 2 that using smaller bins may introduce systematic effects in the correlation functions we want to measure. Taking this fact into account, and the limitations in volume covered and galaxy density, we decided to use three redshift bins. The low redshift bin (abbreviated as 'L' in sample names) corresponds to $z \in [0.3, 0.6]$, the medium redshift ('M') one to $z \in [0.6, 1.0]$, and finally the high redshift bin ('H') covers redshifts $z \in [1.0, 1.5]$. We note that our 'H' redshift bin gets deeper than any previous correlation function study based on spectroscopic surveys (Coil et al., 2006; Meneux et al., 2009; Abbas et al., 2010). We do not consider redshifts $z < 0.3$, as the volume covered is too small, and the geometry of the survey would reduce our analysis to the smallest scales. The three redshift bins chosen contain similar numbers of galaxies in total. From Fig. 3.8, we see that it may be possible to extend our analysis to higher redshift, although with a much lower density. We do not explore this possibility here.

On top of the redshift selection, we also apply a set of cuts in absolute magnitude. We use 'threshold samples', meaning that we will impose a faint luminosity threshold, but not a bright limit. In this way, we obtain approximately volume-limited samples, but also we can study the luminosity dependence of clustering, and its evolution. Following Meneux et al. (2009) and Abbas et al. (2010), we apply an absolute magnitude threshold depending linearly on redshift as

$$M_B^{\mathrm{th}}(z) = M_B^{\mathrm{th}}(0) + Az\,, \tag{3.6}$$

in order to follow the evolution of samples corresponding approximately to the same galaxy population. The value of the constant A characterises the typical luminosity evolution of the galaxies in the catalogue. It can be derived from the evolution of the luminosity function (LF) parameter M^*. Here, we take the value $A = -1$, which is similar to the LF evolution observed in different samples at these redshifts (Ilbert et al., 2005; Zucca et al., 2009). However, we could refine it by measuring the LF using this same ALHAMBRA catalogue. We characterise each sample by the corresponding threshold at $z = 0$, $M_B^{\mathrm{th}}(0)$, and select the galaxies at each redshift z requiring $M_B < M_B^{\mathrm{th}}(z)$, as given by equation (3.6). We made five absolute magnitude cuts at the low redshift bin, and only kept the more luminous samples at higher redshifts. This allows us to study in detail the luminosity dependence of clustering, but we should bear in mind that these samples (for a given redshift bin) are not independent from each other.

Sample	z range	$M_B^{\rm th}(0)$	N	$V(h^{-3}\,{\rm Mpc}^3)$	$\bar{n}(h^3\,{\rm Mpc}^{-3})$	$\bar{z}$	$M_B^{\rm med}$	$L^{\rm med}/L^*$	$\bar{\sigma}_z$	$r(\bar{\sigma}_z)(h^{-1}\,{\rm Mpc})$	$r_{\parallel,{\rm max}}(h^{-1}\,{\rm Mpc})$
L15.5B	$0.3-0.6$	-15.5	24104	0.62×10^6	39.19×10^{-3}	0.452	-17.83	0.10	0.0132	31.8	130
L16.5B	$0.3-0.6$	-16.5	18139	0.62×10^6	29.79×10^{-3}	0.459	-18.28	0.16	0.0116	27.8	110
L17.5B	$0.3-0.6$	-17.5	11179	0.62×10^6	18.18×10^{-3}	0.461	-18.90	0.28	0.0099	23.6	95
L18.5B	$0.3-0.6$	-18.5	5274	0.62×10^6	8.58×10^{-3}	0.458	-19.62	0.54	0.0089	21.3	85
L19.5B	$0.3-0.6$	-19.5	1643	0.62×10^6	2.67×10^{-3}	0.455	-20.34	1.05	0.0084	20.1	80
M16.5B	$0.6-1.0$	-16.5	30742	1.76×10^6	17.45×10^{-3}	0.793	-19.11	0.26	0.0173	34.2	135
M17.5B	$0.6-1.0$	-17.5	23990	1.76×10^6	13.62×10^{-3}	0.802	-19.42	0.34	0.0154	30.3	120
M18.5B	$0.6-1.0$	-18.5	13240	1.76×10^6	7.51×10^{-3}	0.813	-19.86	0.51	0.0128	25.1	100
M19.5B	$0.6-1.0$	-19.5	3849	1.76×10^6	2.18×10^{-3}	0.816	-20.69	1.10	0.0109	21.2	85
H17.5B	$1.0-1.5$	-17.5	22116	3.34×10^6	6.63×10^{-3}	1.203	-20.08	0.46	0.0283	44.7	180
H18.5B	$1.0-1.5$	-18.5	15108	3.34×10^6	4.53×10^{-3}	1.216	-20.50	0.67	0.0258	40.4	160
H19.5B	$1.0-1.5$	-19.5	4482	3.34×10^6	1.34×10^{-3}	1.228	-21.06	1.11	0.0222	34.4	140

Table 3.3: Properties of the subsamples used in the analysis. For each sample, we give the redshift bin; the absolute B magnitude threshold at $z = 0$, $M_B^{\rm th}(0)$, used for the luminosity cut following equation (3.6); the number N of galaxies in the sample; the volume V covered by the redshift bin; the mean galaxy density for this sample $\bar{n}$; the mean redshift $\bar{z}$; the median M_B of the galaxies in the sample; the median luminosity of the galaxies in terms of the L^* at this redshift; the average redshift uncertainty $\bar{\sigma}_z$ calculated according to equation (3.1); the radial distance interval Δr corresponding to this $\bar{\sigma}_z$ at $\bar{z}$; and finally the maximum used for the integration along the line-of-sight in our calculation, $r_{\parallel,{\rm max}}$.

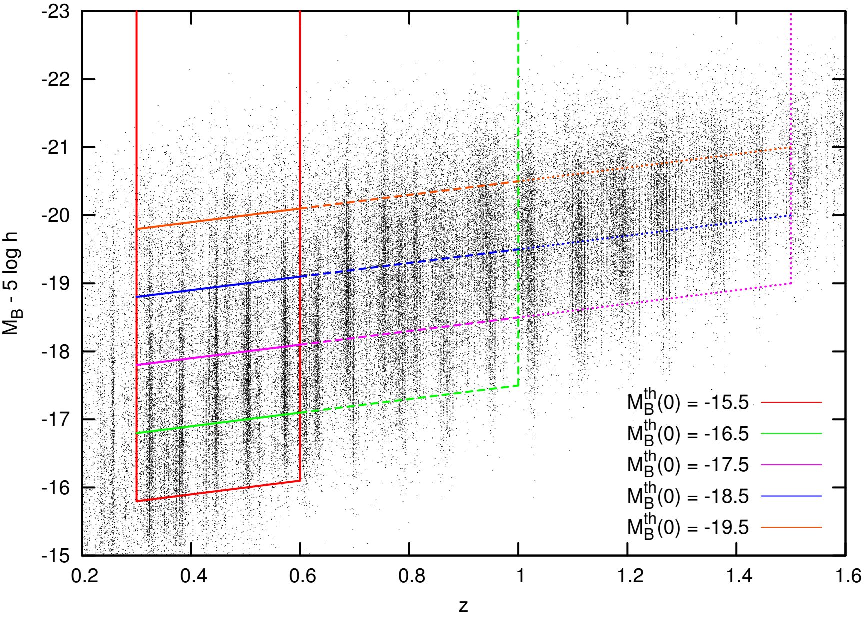


Figure 3.14: Absolute B-band magnitude M_B vs. redshift z for our LSS catalogue. The different lines show the boundaries of the samples we select for our analysis.

We show in Fig. 3.14 the actual cuts made in the redshift – absolute magnitude plane to define our samples, and give details of the properties of each sample in Table 3.3. The name used for each sample comes from the letter 'L', 'M' or 'H' to denote the redshift bin, the absolute value of the $M_B^{\rm th}(0)$ used for the cut, and a 'B' denoting the band used for the absolute magnitudes. For each sample we give the basic properties such as number of galaxies, volume and density, but also the characteristic redshift $\bar{z}$ and B band absolute magnitude $M_B^{\rm med}$ of the galaxies in the sample. We compare the $M_B^{\rm med}$ to the characteristic absolute magnitude M_B^* at $\bar{z}$ for each of the samples. We calculate $M_B^*(z)$ for this comparison using the data from Abbas et al. (2010), who used the LF obtained by Ilbert et al. (2005) for the VVDS. From the obtained luminosity ratios $L^{\rm med}/L^*$, we see that we are considering here samples with $L \lesssim L^*$.

As the accuracy of the photometric redshifts depends on magnitude and redshift, it varies among our samples. We calculated the mean estimated error on the redshift determination, $\bar{\sigma}_z$, calculated according to equation (3.1). We use this parameter below to determine the maximum line-of-sight distance $r_{\parallel,\max}$ for our integration of $\xi(r_\perp, r_\parallel)$.

3.3.2 The calculation in practice

We measured both the projected correlation function $w(r_\perp)$, and the de-projected real-space correlation function $\xi_{\mathrm{dep}}(r)$ using the method described in detail in the previous chapter, and specifically in Sections 2.1 and 2.3.1. In summary, we separate the distance between pairs of galaxies in two components, one along the line-of-sight ($r_\parallel$), and one in the transverse direction ($r_\perp$), following equation (2.1). We estimate the two-point correlation function $\xi(r_\perp, r_\parallel)$ using the Landy and Szalay (1993) (LS) estimator, and then integrate along the line-of-sight (equation 2.5) to obtain $w(r_\perp)$. Finally, we use the relation (2.4) between the real-space correlation function $\xi_{\mathrm{r}}(r)$ and $w(r_\perp)$ to obtain our de-projected correlation function, $\xi_{\mathrm{dep}}(r)$, which should be a good estimate of $\xi_{\mathrm{r}}(r)$. The actual calculation is made based on the discrete values calculated for $w(r_\perp)$ using equation (2.6).

In the case of ALHAMBRA, we are dealing with 7 fields widely separated in the sky. Correlations between different fields are therefore unimportant for our calculations. However, correlations between pairs of objects located in a different frame or strip but in the same field are actually important for the study of correlations at scales up to a few tens of Mpc. We therefore make the full calculation of $w(r_\perp)$ and $\xi_{\mathrm{dep}}(r)$ for each of the fields separately, and we obtain our final result by averaging the results obtained for the different fields. When making this average, we weight each of the fields proportionally to the effective area covered in each of them (see Table 3.2).

In order to calculate $\xi(r_\perp, r_\parallel)$ using the LS estimator, we need an auxiliary un-clustered catalogue. This should be a realisation of an homogeneous Poisson random process, following the same selection function as the real data for the sample considered. We model the angular selection function of the survey using the mask created in Section 3.2.1, and assuming a homogeneous completeness inside it. In order to model the radial selection function, we calculate the density along the radial direction, $\rho(r)$, for each of the samples, averaging over the different fields. We smooth $\rho(r)$ with a kernel of width $30\,h^{-1}\,\mathrm{Mpc}$, and use the resulting function as our estimate of the radial selection function. The $\rho(r)$ obtained for each of the samples is shown in Fig. 3.15. Despite our M_B selection, the samples are not exactly volume-limited, as can be seen by the global trend of $\rho(r)$ to decrease with r. This trend is more clear for the high-redshift bin. Moreover, we obtain a series of peaks at approximately regular intervals in $\rho(r)$, also seen in Fig. 3.14. We consider these peaks as part of the selection function. We generate the corresponding Poisson catalogues for each sample and field, containing $N_R = 20 N_D$ points in each case to reduce the effects of shot noise in our results.

As we are averaging over 7 separate fields, the peaks observed in $\rho(r)$ can not correspond to real structures in our sample, but should be the result of the data processing or redshift determination. In fact, when studying the locus of galaxy spectra in multi-colour space as function of redshifts, it is common to obtain that some values of z are favoured over neighbouring ones. This could be the

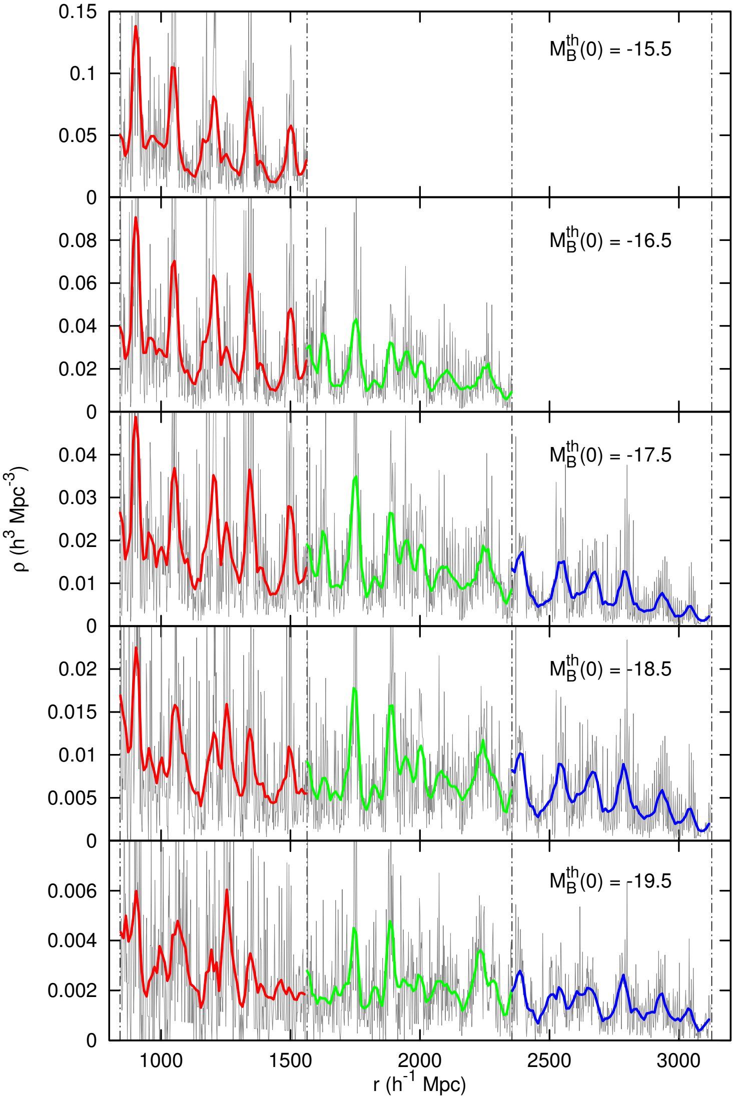


Figure 3.15: Number density of the different samples used, as function of radial distance. The different samples are identified by the corresponding absolute magnitude threshold at $z = 0$, $M_B^{\rm th}(0)$. The grey lines show the original density profile, and the solid lines (red for the low-, green for the medium-, and blue for the high-redshift bin) show the actual density functions we use for our calculations after smoothing with a $30\,h^{-1}$ Mpc kernel. The limits between redshift bins are marked by vertical dot-dashed lines.

explanation for these observed peaks. These peaks could in principle distort our galaxy clustering analysis. However, they would only affect the measurement of $r_{\parallel}$ of the pairs (at scales smaller or of the order of our integration limit $r_{\parallel,\mathrm{max}}$), but not of $r_{\perp}$. Given that most of our clustering signal comes from the transverse separations $r_{\perp}$, the possible influence of these peaks in our results must be small, and it should be reduced by considering the peaks as part of our selection function. In any case, this issue will be revised when preparing the final ALHAMBRA catalogues.

When integrating $\xi(r_{\perp}, r_{\parallel})$ along the line of sight to obtain $w(r_{\perp})$ in practice, we need to set an upper limit for the integration, $r_{\parallel,\mathrm{max}}$. We showed in Section 2.4, using our simulated halo catalogues, that an optimal choice for this value is $r_{\parallel,\mathrm{max}} \simeq 4r(\Delta z)$, where $r(\Delta z)$ is the radial distance corresponding to the typical redshift uncertainty Δz of the sample. In the case of ALHAMBRA data, we estimate Δz as the mean of the σ_z values (equation 3.1) of the galaxies in the sample, and use this same choice for $r_{\parallel,\mathrm{max}}$. This is consistent with the definition of Δz in the simulated catalogues used in Section 2.4. The actual values used for the calculations using the different samples are shown in Table 3.3.

Regarding the upper limit $r_{\perp,\mathrm{max}}$ in the integration of $w(r_{\perp})$ to obtain $\xi_{\mathrm{dep}}(r)$, we use the values $r_{\perp,\mathrm{max}} = 20, 30, 40\,h^{-1}$ Mpc, respectively, for samples in the low-, medium-, and high-redshift bins. These values correspond, approximately, to the transverse co-moving distance subtended by an angle of 45 arcmin (three times the side of a LAICA CCD) at the maximum redshift of each of the bins. In all our calculations, we use logarithmic bins in both $r_{\perp}$ and $r_{\parallel}$, with a spacing of $\Delta \log_{10} r = 0.15$.

Error estimation

In order to estimate the error and full covariance matrix of our results, both for $w(r_{\perp})$ and $\xi_{\mathrm{dep}}(r)$, we used the jackknife method (see e.g. Norberg et al., 2009). In this case, we used as 'jackknife regions' the $N_{\mathrm{jack}} = 39$ frames in the catalogue, so the jackknife samples used were obtained omitting one of the frames at a time. We repeated the full calculation for each jackknife sample, taking into account the change in weighting when averaging over the fields. Our estimate for the covariance matrix between bins for $w(r_{\perp})$ is then

$$C_{ij} = \frac{N_{\mathrm{jack}} - 1}{N_{\mathrm{jack}}} \sum_{k=1}^{N_{\mathrm{jack}}} (w_i^k - \overline{w_i})(w_j^k - \overline{w_j})\,, \tag{3.7}$$

where w_i^k is the value obtained for the jackknife sample k in the $r_{\perp}$ bin i, and $\overline{w_i}$ is the average of the values obtained for bin i. The same method was used to estimate the covariance matrix in $\xi_{\mathrm{dep}}(r)$. The standard error estimate for a single bin is given by $\sigma_i = \sqrt{C_{ii}}$.

An alternative method for estimating the errors would be to use directly the variance between the values observed in the 7 different fields. As the fields are separated in the sky, they can be considered as totally independent, and thus this can be a direct measure of the uncertainty in the mean value obtained. The results for σ_i obtained using our jackknife method are similar to those obtained from the variance between fields, but the jackknife results are less noisy. This is an additional indication that our jackknife estimation is correctly taking into account not only the shot noise, but also the cosmic variance contribution to the uncertainty. Moreover, the fact of using 39 realisations instead of only 7 fields means that the results are more robust, and that we can estimate the full covariance matrix, and not just its diagonal terms.

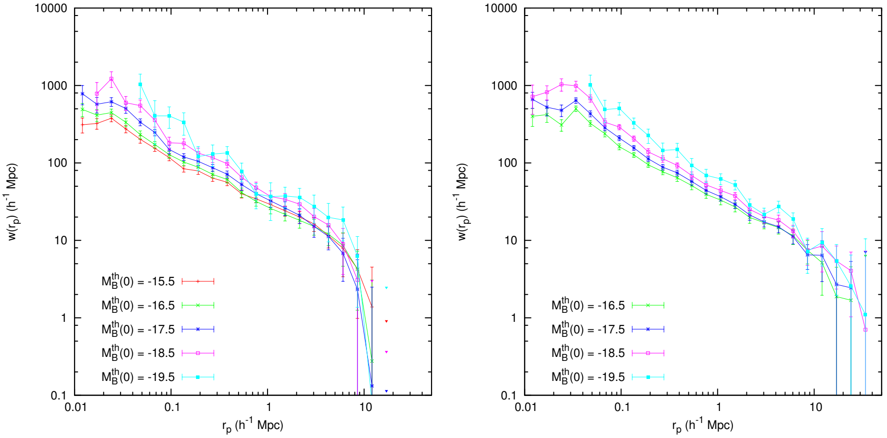
3.3.3 Results

Following the same procedure as in Chapter 2, with the details explained above, we calculated both the projected correlation function $w(r_\perp)$, and the de-projected real-space correlation function $\xi_{\mathrm{dep}}(r)$ for our samples.

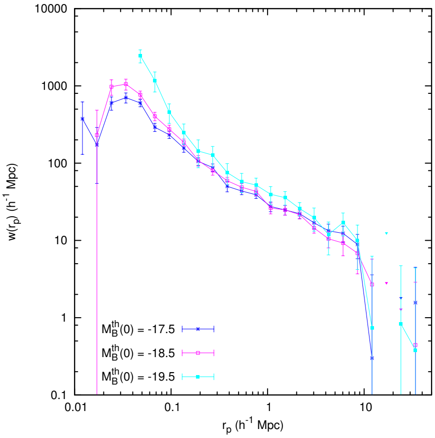
Results for $w(r_\perp)$ are shown in Fig. 3.16. We see that we can reliably measure $w(r_\perp)$ in our different samples from scales as low as $r_\perp \simeq 0.02\,h^{-1}\,\mathrm{Mpc}$ (except for the high-redshift bin), up to scales of $r_\perp \simeq 10-30\,h^{-1}\,\mathrm{Mpc}$, depending on the sample. Deep spectroscopic surveys are typically able to measure the correlation function only for scales $r \gtrsim 0.1\,h^{-1}\,\mathrm{Mpc}$ due to the sparsity of the samples used. Using ALHAMBRA data, being a photometric survey, we can explore smaller scales due to the larger number density of galaxies in the samples. The correlation function at these scales $r \lesssim 0.1\,h^{-1}\,\mathrm{Mpc}$ is related to the merger rate in the galaxy population studied (Masjedi et al., 2006). We ignore these scales for the rest of this analysis, but this shows that ALHAMBRA can be a good tool to study the history of galaxy merger rates up to high redshifts (López-Sanjuan et al., 2011; de Ravel et al., 2011).

Over the full range of scales studied, the obtained $w(r_\perp)$ shows qualitatively the expected behaviour. It follows approximately a power law, although with some deviations or changes in slope, most noticeable in the high-redshift samples. Moreover, there is a clear segregation with luminosity, as samples containing more luminous galaxies exhibit stronger clustering. The dependence of clustering on redshift is less evident here. We make more detailed comparisons in our analysis below.

The results for the de-projected correlation functions $\xi_{\mathrm{dep}}(r)$ of the different samples are shown in Fig. 3.17. Qualitatively, we can draw here the same conclusions as from the $w(r_\perp)$ results above: the correlation function is close to a power law for nearly three decades in separation, and it shows the effect of luminosity segregation. However, the results we obtain here are noisier, and with larger estimated errors, than those for $w(r_\perp)$, as is expected for any de-convolution process.

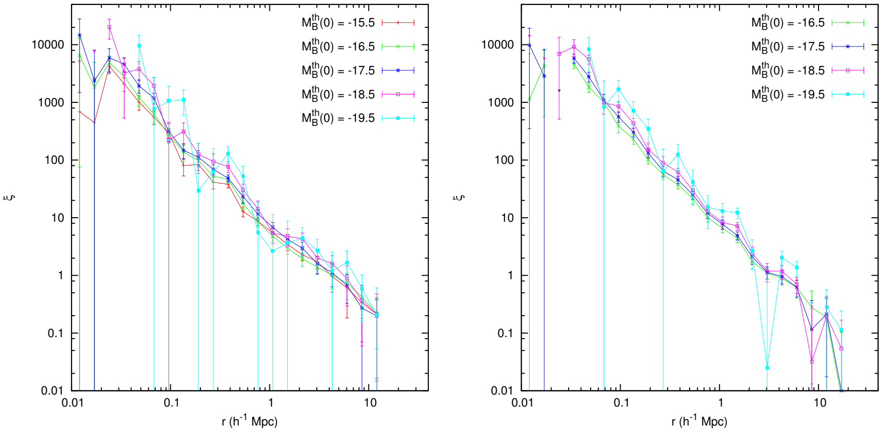


(a) $z \in [0.3, 0.6]$ **(b)** $z \in [0.6, 1.0]$

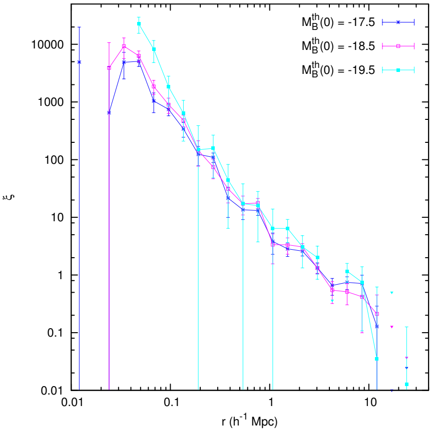


(c) $z \in [1.0, 1.5]$

Figure 3.16: Projected correlation function $w(r_\perp)$ obtained for the different samples used in our analysis. Each of the panel shows the results for the luminosity-selected samples in a given redshift bin. The errors shown in each case are calculated using the jackknife method as described in the text.



(a) $z \in [0.3, 0.6]$ **(b)** $z \in [0.6, 1.0]$



(c) $z \in [1.0, 1.5]$

Figure 3.17: Same as Fig. 3.16, but now for the de-projected real-space correlation function $\xi_{\text{dep}}(r)$.

Hence, we will use only the $w(r_\perp)$ results for further analysis, as it is in general possible to transform a model for $\xi(r)$ into a model for $w(r_\perp)$ using equation (2.3).

Modelling of the correlation function: dependence of clustering on luminosity and redshift

In order to study the change of the clustering properties with luminosity and redshift, we fit the obtained projected correlation function $w(r_\perp)$ of each sample using two models.

The first of these models is a power law in the real-space correlation function $\xi(r)$, expressed as

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

When transforming this model, using equation (2.3), to a model for $w(r_\perp)$, we also obtain a power law which, expressed in terms of the parameters r_0 and γ above is given by

$$w(r_\perp) = r_\perp \left(\frac{r_0}{r_\perp}\right)^{\gamma} \frac{\Gamma(1/2)\Gamma\left[(\gamma-1)/2\right]}{\Gamma(\gamma/2)} , \tag{3.8}$$

where $\Gamma(\cdot)$ is the gamma function. This is a simple phenomenological model, but it has been shown to fit with reasonable accuracy the correlation function of galaxies of very different properties over a large range of scales (e.g. Hawkins et al., 2003; Zehavi et al., 2005a; Coil et al., 2006). A recent analysis with the SDSS catalogue has found slight but significant deviations from the power-law behaviour at the range of scales studied here (Zehavi et al., 2004). It would be interesting to study whether these same deviations are found using high redshift samples as those used here. Fitting the power-law model of equation (3.8) to our observed data, we can study the change of both the slope γ, and the correlation length r_0 with the properties of each sample.

The second model we use has a more direct physical meaning, although it is still a simplified model. We use a model for the clustering of matter based on the standard ΛCDM perturbation theory, using the software CAMB[5] (Lewis et al., 2000). We include the non-linear corrections of HALOFIT (Smith et al., 2003), as implemented in CAMB itself. We Fourier-transform the power spectrum calculated by CAMB to obtain the matter correlation function, and then transform it using equation (2.3) to obtain the matter projected correlation function $w_m(r_\perp)$ for this model. We assume a very simplified relation between the clustering of matter and galaxies, using a constant bias parameter b not depending on scale, and therefore our model for the galaxy projected correlation function is simply

$$w(r_\perp) = b^2 w_m(r_\perp) . \tag{3.9}$$

---

[5] http://camb.info

For a more appropriate physical model, taking into account the details of the galaxy formation inside haloes, one should use an Halo Occupation Distribution (HOD) fit. We leave this for a later work with the final ALHAMBRA catalogue. In any case, this model is useful because it provides a separation between the overall clustering of matter and the details of the different tracers used. In this sense, we can factor out the overall redshift evolution of clustering, which depends mainly on the growth function $D_+(z)$, and which is included in $w_m(r_\perp)$, from the bias b of the different samples. We will keep the cosmological parameters for the calculation of $w_m(r_\perp)$ fixed at the standard WMAP7 values (Komatsu et al., 2011), and calculate it at the mean redshift of each sample. Hence, the only parameter we fit in this case is b.

We fit our observed $w(r_\perp)$ to these two models using the generalised χ^2 minimisation method. We calculate the χ^2 value on a grid of parameter values $\boldsymbol{\theta}$ (r_0, γ for the power-law model, b for the ΛCDM model), and obtain from it the expected values and confidence regions for the parameters in the standard way (see e.g. Press et al., 2002). The χ^2 function is given by

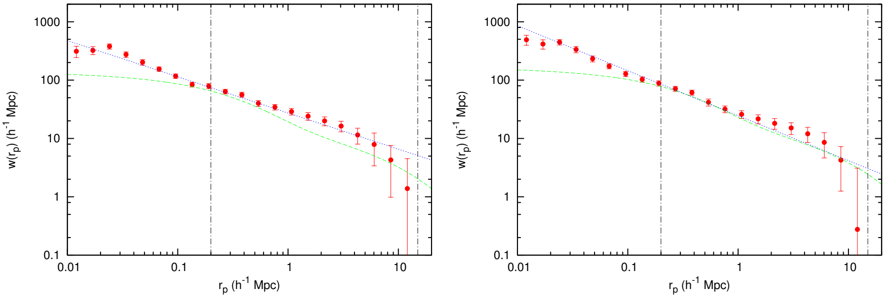
$$\chi^2(\boldsymbol{\theta}) = \sum_{i=0}^{N_{\rm bins}} \sum_{j=0}^{N_{\rm bins}} \left[w(r_{\perp,i}) - w_{\rm model}(r_{\perp,i}, \boldsymbol{\theta})\right] (C^{-1})_{ij} \left[w(r_{\perp,j}) - w_{\rm model}(r_{\perp,j}, \boldsymbol{\theta})\right] ,$$

where $N_{\rm bins}$ is the number of bins in $r_\perp$ considered, $w_{\rm model}(r_\perp, \boldsymbol{\theta})$ is the model calculated for the values $\boldsymbol{\theta}$ of the parameters, and C^{-1} is the inverse of the covariance matrix between the observations at different bins, which we calculate using the jackknife method (equation 3.7).

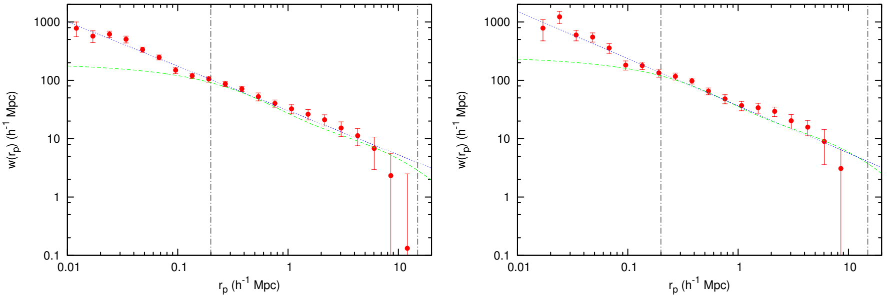
We use in all cases the distance bins in the range $r_\perp \in [0.2, 15]\,h^{-1}\,$Mpc. The lower limit is imposed in order to avoid the scales where the effect of mergers become important. This is seen, specially for the high-redshift samples, as a change in the slope of $w(r_\perp)$ around $r_\perp \simeq 0.1\,h^{-1}\,$Mpc. The upper limit is set to avoid larger scales where the error in w is large, due to the geometry of the fields. Therefore, we are using 12 $r_\perp$ bins for the fit, this leaves us with $\nu = 11$ for the ΛCDM fit, and $\nu = 10$ for the power-law fit when we allow both r_0 and γ to change. We also made a fit to a power law, but keeping the slope γ fixed to the value $\gamma = 1.7$. In this case, again, $\nu = 11$.

We compare the observed $w(r_\perp)$ to the best-fit models in Figs. 3.18, 3.19 and 3.20. Details of the resulting fits are listed in Table 3.4. From the different plots, and the $\chi^2_{\rm min}$ values in the Table, we see that both models give reasonable fits for most samples, without significant differences between the two models. This means that, using this simple approach, we do not detect deviations from the power-law behaviour of the correlation function. However, a more detailed fit using HOD could significantly improve these fits, at least for some of the samples.

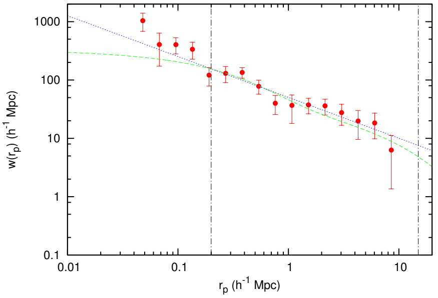
Using the power-law fits, we can study the change in clustering properties for the different samples. In Fig. 3.21, we show the dependence of the fitted parameters r_0 and γ on the median absolute magnitude of the different samples, for each



(a) L15.5B **(b)** L16.5B



(c) L17.5B **(d)** L18.5B

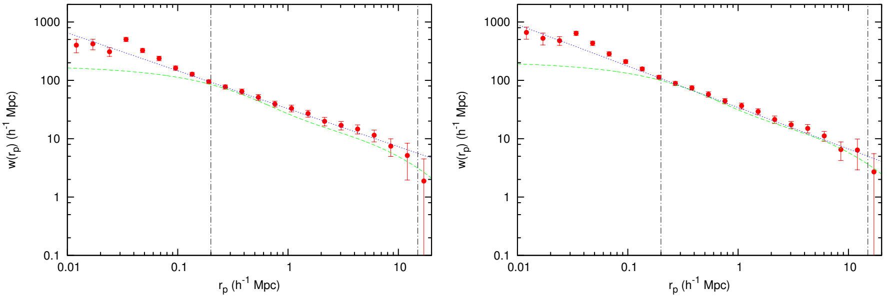


(e) L19.5B

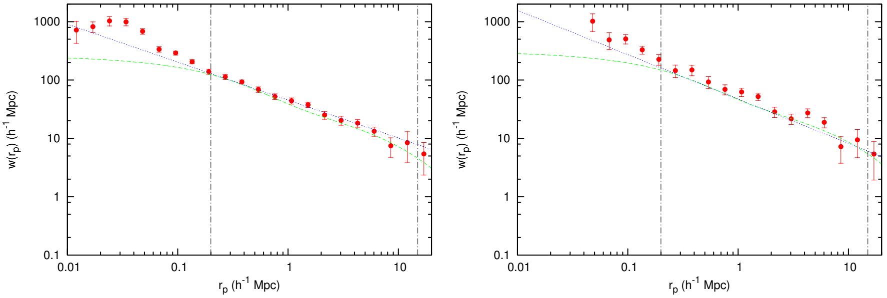
Figure 3.18: Comparison between the observed $w(r_\perp)$ (red points with error bars) and the best fit power-law (blue dotted line) and ΛCDM (green dashed line) models. Each panel shows the results for one of the luminosity-selected samples in the low-redshift bin ($z \in [0.3, 0.6]$). The vertical dot-dashed lines indicate the range which we are using for the different fits, $r_\perp \in [0.2, 15]\, h^{-1}\,\mathrm{Mpc}$.

	Power-law fit			Power-law fit with $\gamma = 1.7$		ΛCDM fit	
Sample	$r_0(h^{-1}\,\mathrm{Mpc})$	γ	$\chi^2_{\min}$	$r_0(h^{-1}\,\mathrm{Mpc})$	$\chi^2_{\min}$	b	$\chi^2_{\min}$
L15.5B	3.07 ± 0.18	1.65 ± 0.05	15.70	2.84 ± 0.11	18.25	0.76 ± 0.02	31.85
L16.5B	$2.88^{+0.21}_{-0.23}$	$1.77^{+0.08}_{-0.06}$	10.11	3.06 ± 0.13	11.26	$0.83^{+0.02}_{-0.03}$	11.00
L17.5B	$3.25^{+0.19}_{-0.22}$	$1.76^{+0.05}_{-0.04}$	10.41	3.44 ± 0.14	12.59	$0.90^{+0.02}_{-0.03}$	12.39
L18.5B	3.6 ± 0.3	$1.81^{+0.09}_{-0.07}$	14.82	3.91 ± 0.21	17.82	$1.03^{+0.04}_{-0.05}$	14.95
L19.5B	4.4 ± 0.4	1.70 ± 0.08	16.53	4.4 ± 0.4	16.53	$1.17^{+0.08}_{-0.10}$	19.36
M16.5B	$3.42^{+0.19}_{-0.24}$	$1.65^{+0.06}_{-0.04}$	11.86	$3.24^{+0.10}_{-0.11}$	12.57	$1.10^{+0.02}_{-0.03}$	14.12
M17.5B	$3.50^{+0.16}_{-0.18}$	$1.71^{+0.05}_{-0.04}$	8.21	3.52 ± 0.11	8.28	$1.19^{+0.03}_{-0.02}$	9.20
M18.5B	$4.18^{+0.21}_{-0.20}$	1.65 ± 0.04	16.46	$4.00^{+0.15}_{-0.14}$	17.92	1.33 ± 0.04	21.14
M19.5B	4.2 ± 0.3	1.76 ± 0.08	40.31	4.2 ± 0.3	40.84	$1.45^{+0.08}_{-0.10}$	38.42
H17.5B	$3.15^{+0.19}_{-0.18}$	$1.64^{+0.06}_{-0.05}$	25.49	$3.00^{+0.13}_{-0.12}$	26.63	1.37 ± 0.04	23.88
H18.5B	$3.41^{+0.19}_{-0.18}$	$1.71^{+0.05}_{-0.04}$	24.39	3.43 ± 0.11	24.46	1.51 ± 0.04	28.71
H19.5B	$3.84^{+0.22}_{-0.23}$	$1.67^{+0.11}_{-0.09}$	10.31	3.80 ± 0.18	10.39	$1.64^{+0.06}_{-0.07}$	11.36

Table 3.4: Parameters obtained from the fits to $w(r_\perp)$ for the different samples using the power-law and ΛCDM fits. In the case of the power-law fits, we show the results when we allow both r_0 and γ to vary, and when we only vary r_0, and keep the slope fixed at $\gamma = 1.7$. The uncertainties quoted are 1σ marginalised uncertainties in all cases. We also show the χ^2 value at the minimum in each case. The number of degrees of freedom is $\nu = 10$ for the power-law fit with two free parameters, and $\nu = 11$ for the ΛCDM fit, and the for power-law fit with γ fixed.



(a) M16.5B **(b)** M17.5B

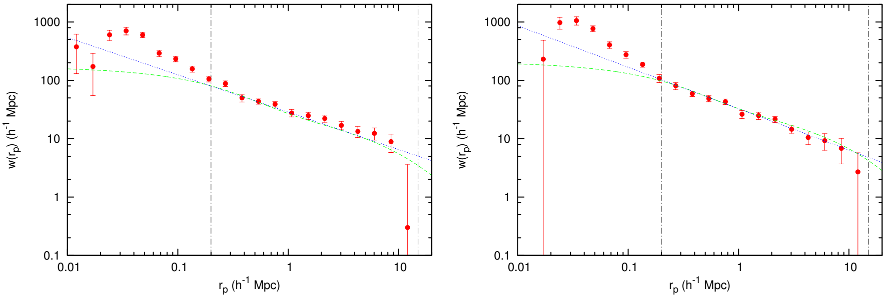


(c) M18.5B **(d)** M19.5B

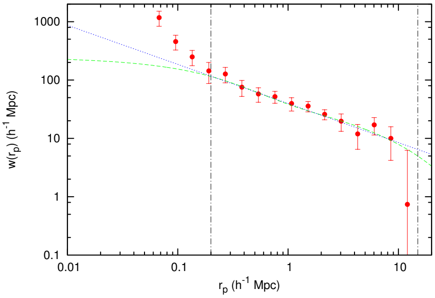
Figure 3.19: Same as Fig. 3.18, for the samples in the medium-redshift bin ($z \in [0.6, 1.0]$).

of the redshift bins. We compare our results with those obtained for the DEEP2 survey by Coil et al. (2006), whose samples have mean redshifts in the range $z \in [0.87, 0.99]$, and are therefore in between our medium- and high-redshift bins. Our results are in overall agreement with the DEEP2 data. Moreover, this comparison shows the potential of ALHAMBRA to analyse the clustering properties for relatively faint samples, while spectroscopic surveys are restricted only to the brightest samples. In contrast, our actual constraints on the parameters (r_0, γ) for individual samples are weaker, due mainly to the uncertainties introduced by the photometric redshift measurements.

In Fig. 3.21a, we study the dependence of the slope parameter γ with the redshift and median M_B of the samples. In this case, any change in γ is of the order of our errors from the fit, so we are not able to measure any significant change in the slope of the correlation function between our samples. On the other hand, in Fig. 3.21b, we see clearly an increase in the correlation length r_0 with



(a) H17.5B **(b)** H18.5B



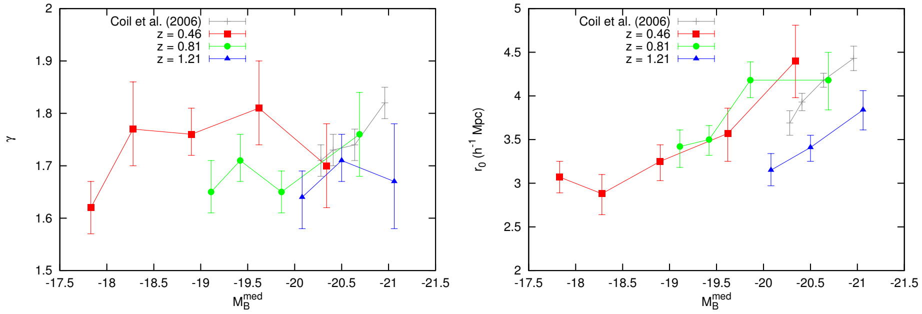
(c) H19.5B

Figure 3.20: Same as Fig. 3.18, for the samples in the high-redshift bin ($z \in [1.0, 1.5]$).

the median M_B of the sample. This is a consequence of luminosity segregation in our data, as brighter galaxies tend to cluster more strongly than fainter ones. This phenomenon is present in the three redshift bins studied.

However, as the values of r_0 and γ resulting from the fits are strongly correlated, the r_0 we obtain from this fit is a not clean measure of the overall strength of the clustering for different samples. An alternative is to repeat the power-law fit, but this time keeping the value of γ fixed. As we have not seen a change in γ for different samples, it makes sense to set $\gamma = 1.7$ in all cases, and fit for r_0 only. We show the constraints on r_0 obtained in this case in Fig. 3.22. Here, we see more clearly a monotonic increase of r_0 with luminosity in all cases. Moreover, we see a sign of the evolution of clustering in the fact that the strength of clustering (measured through r_0) decreases with redshift. This is qualitatively the expected behaviour from the process of structure formation.

We show the dependence of the bias parameter b defined in equation (3.9) on



(a) Slope γ

(b) Correlation length r_0

Figure 3.21: Parameters r_0 and γ obtained from the power-law fits for the different samples, as function of the median absolute magnitude M_B of the samples, for each of the redshift bins. For comparison, we show the results obtained for DEEP2 data by Coil et al. (2006), for samples at redshift $z \simeq 0.9 - 1.0$.

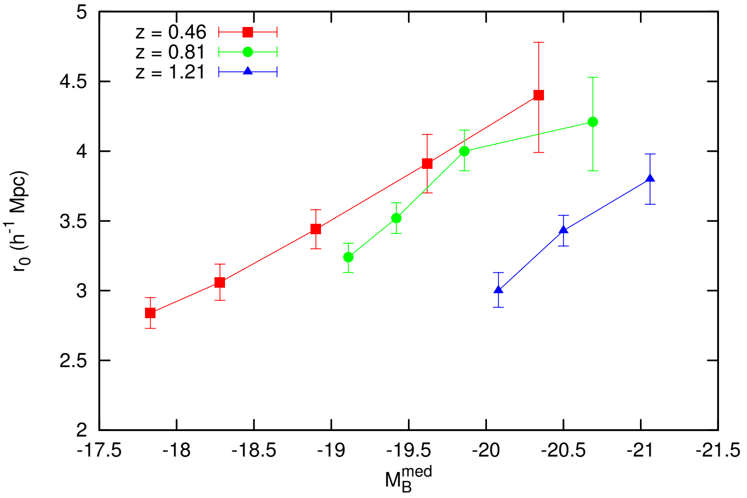


Figure 3.22: Constraints on the parameter r_0 obtained from the power-law fits when the slope is kept fixed at $\gamma = 1.7$, as function of the median absolute magnitude M_B of the samples, for each of the redshift bins.

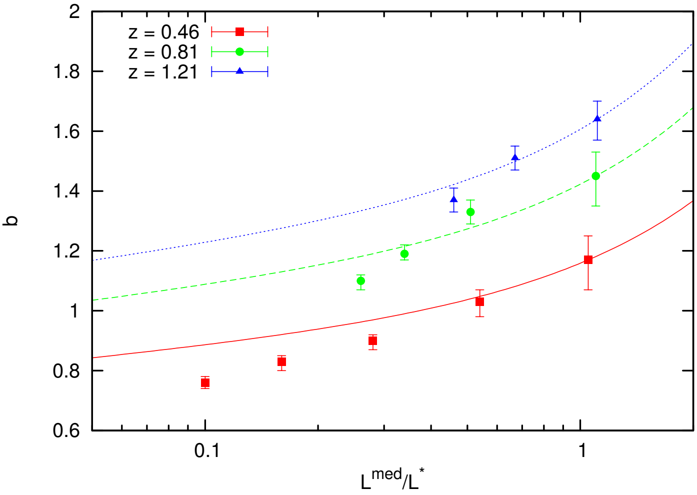


Figure 3.23: Bias parameter b obtained from our Camb fit for the different samples, as function of the median absolute luminosity L/L^* of the samples, for each of the redshift bins. The lines are the fits obtained by Tegmark et al. (2004) for SDSS data at $z \simeq 0.1$ (equation 3.10), with b^* chosen in each case to match the observed value of b for our brightest sample.

luminosity in Fig. 3.23. We measure in each case the median luminosity of the samples in terms of the L^* at the appropriate redshift (as shown in Table 3.3), in order to be able to compare to results at $z \sim 0$. In this case, the effect of the growth of structures with cosmic time is included in $w_m(r_\perp)$, so here we study the absolute bias of the galaxy population studied with respect to the matter distribution.[6] We see that the bias is increasing with redshift, even when we tried to select samples corresponding to the same type of objects at different redshift. This is qualitatively consistent with the fact that, for dark matter haloes of a given fixed mass, the bias increases with redshift (Mo and White, 1996; Sheth and Tormen, 1999).

Regarding the dependence of b on L, we see a clear positive trend, in line with our results for r_0 above. This means that luminosity segregation is present in our samples, even for $L < L^*$. We compare our results with those obtained by Tegmark et al. (2004) using data from SDSS at $z \simeq 0.1$. They found (see their Fig. 30) that the dependence of b on luminosity is well fitted in that case by the

---

[6] Obviously, using different cosmological parameters for the calculation of $w_m(r_\perp)$ would give slightly different values of b, but the qualitative behaviour should be similar for reasonable values of these parameters.

function

$$b/b^* = 0.85 + 0.15\frac{L}{L^*} - 0.04(M - M^*) = 0.85 + 0.15\frac{L}{L^*} + 0.1\log_{10}\frac{L}{L^*}\,. \quad (3.10)$$

A similar fit was found by Norberg et al. (2001) using data from the 2dFGRS. We compare our results to this fit (shown as lines in Fig. 3.23), by choosing the appropriate value of b^* at each redshift bin to match our brightest sample. We see a reasonable agreement for $L \gtrsim 0.4L^*$. However, for the faintest samples, we obtain a stronger dependence of b on luminosity than observed at low redshift. A similar effect, although for brighter samples, was already found by Coil et al. (2006) at $z \sim 1$.

3.4 Discussion

In this chapter, we used preliminary data from ALHAMBRA to study the clustering of galaxies at high redshift, in the region between $z = 0.3$ and $z = 1.5$. The aim of this chapter was mainly exploratory, in order to show the capabilities of ALHAMBRA for this kind of analysis, and to test the method for the calculation of the correlation function from photometric surveys (described in Chapter 2) in real data.

In the first place, we described the selection procedure needed to select, from the full ALHAMBRA preliminary Data Release (IDR3), a galaxy catalogue which is appropriate for large-scale structure studies. To this end, we built the angular masks describing the survey's angular selection function, performed a colour-based star-galaxy separation, and selected objects with high quality redshift determinations. Although this selection is preliminary, and some details may change to adapt it to the final data release, it describes the basic catalogue that will be used for LSS studies with ALHAMBRA. Moreover, it describes the type of selections needed to use data from any photometric survey, specially those with high quality redshift determinations, for clustering studies. However, details will vary in function of the characteristics of the survey, and of the analysis to be performed.

In the second place, we selected, based on redshift and absolute magnitude, 12 different samples from our catalogue. We calculated, using the method described in the previous chapter adapted to the characteristics of the ALHAMBRA survey, both the projected correlation function $w(r_\perp)$ and the de-projected real-space correlation function $\xi_{\rm dep}(r)$ for these samples. We showed that our method allows us to measure the correlation function reliably in this case, as shown using simulations in the previous chapter.

In order to study the dependence of clustering properties on luminosity and redshift, we fit $w(r_\perp)$ of the different samples by two models: a power-law correlation function, and a ΛCDM model for the matter correlation function multiplied by a constant bias factor b. We did not find any significant deviation from a power law in the range of scales studied, $r_\perp \in [0.2, 15]\, h^{-1}$ Mpc, comparing to the alternative ΛCDM model. Studying the dependence of the fitted parameters r_0 (correlation length) and b (bias) with the luminosity of the samples, we see a clear indication of segregation: more luminous galaxies tend to cluster more strongly than fainter ones. Given that our samples have typical luminosities $L \lesssim L^*$, this behaviour is different to that observed at $z \simeq 0$. For local samples, as first noted by Hamilton (1988), luminosity segregation is weak for faint galaxies, and is only clear for galaxies at the bright end of the luminosity function, with $L > L^*$. We show this discrepancy explicitly in Fig. 3.23, when we compare our bias results to those from SDSS. When studying the dependence of the slope γ with luminosity, we are not able to measure any significant change.

When comparing our results with those from deep spectroscopic surveys, and specifically DEEP2 (Coil et al., 2006), we see the differences between both approaches. The constraints obtained in DEEP2 for a given sample are tighter than those obtained by ALHAMBRA, due to the much higher precision in the determination of redshifts. However, ALHAMBRA allows us to explore a much larger range of redshift and of absolute magnitude. Moreover given the large density of the ALHAMBRA samples, we can measure the correlation function to relatively small scales. Although not explored here, the study of clustering at scales $r_\perp \in [10, 100]\, h^{-1}$ kpc with ALHAMBRA could give important information regarding the process of galaxy merging over a large range of redshifts.

The work presented here is anyhow preliminary, and the main purpose was to test our clustering analysis machinery. When the final ALHAMBRA catalogue is available, this work will be adapted to the characteristics of the data, and a more detailed analysis will be made, in particular using a HOD fit to the data, and, if possible, extending the analysis to higher redshifts. Also, an analysis of segregation, depending not only on luminosity but also on other galaxy properties, such as colour, spectral type, morphology, or stellar mass, will be made. In Chapter 4 we review some of the more advanced statistical methods that could be used for this purpose.

4

Measuring galaxy segregation with mark statistics

It is a well known fact that different types of galaxies show different clustering properties. This phenomenon is called segregation. These clustering properties depend on several intrinsic characteristics of the galaxies, some of them directly observable, such as luminosity, colour, spectral type or morphology. Elliptical galaxies, for example, are preferentially located in high density environments, such as the centres of rich galaxy clusters, while the dominant population of the field are mainly spiral galaxies (Davis and Geller, 1976; Dressler, 1980; Postman and Geller, 1984; Goto et al., 2003). Segregation is a consequence of the process of structure and galaxy formation through cosmic time. In the standard ΛCDM model it can be explained considering the joint evolution of galaxies and the dark matter haloes that host them, and the process of assembly and mergers that they undergo.

The most used methods to quantify clustering segregation between galaxies with different properties have been standard two point statistics, such as the correlation function $\xi(r)$ (see Section 1.2). Normally, galaxy samples are divided into two or more classes by a given property, and the differences in the corresponding correlation functions give information about segregation between these classes. One can identify broadly two types of galaxies. Early-type galaxies show elliptical (or bulge-dominated) morphologies, red colours and low star formation rates. Late-type galaxies, on the other side, show spiral (or disk-dominated) morphologies, blue colours and ongoing star formation. The correlation function of early-type galaxies shows at small scales steeper slopes and larger amplitudes than that of late-type galaxies (Loveday et al., 1995; Hermit et al., 1996; Guzzo et al., 1997). On top of the galaxy type, the clustering properties also depend on galaxy luminosity, with brighter galaxies showing larger clustering amplitude than fainter ones (Phillipps and Shanks, 1987; Hamilton, 1988; Davis et al., 1988). Recent studies have made use of the last generation of galaxy redshift surveys to analyse these properties in more detail. The results using data from the Two-degree Field Galaxy Redshift Survey (2dFGRS, Norberg et al., 2001, 2002; Madgwick et al., 2003b), and from the Sloan Digital Sky Survey (SDSS, Zehavi et al., 2002, 2005b, 2011), broadly confirm earlier results when studying galaxy clustering as a function of galaxy luminosity, colour or spectral type. In order to quantify the luminosity or morphological segregation, clustering measures other than two point statistics have also been used, such as multifractals (Dominguez-Tenreiro and Martinez, 1989; Dominguez-Tenreiro et al., 1994), void probability functions (Vogeley et al., 1991; Croton et al., 2004), or distributions of the distances to the nearest neighbours (Salzer et al., 1990).

As explained in Section 1.2, in order to study the properties of galaxy clustering, we consider the galaxy distribution as a realisation of a homogeneous random point process, and apply the statistical machinery associated with them. To analyse segregation as a function of a given galaxy property, we can generalise this approach, and consider this distribution as a realisation of a marked point process (Stoyan and Stoyan, 1994; Martínez and Saar, 2002; Illian et al., 2008). In this kind of processes, each point carries a given characteristic or ‘mark’ associated with it, and the joint properties of the spatial distribution and of the associated mark distribution can be studied. The mark can be either a continuous variable describing quantitatively a given property associated to the points, or a discrete variable defining qualitatively different categories or types of points. This approach has been applied in different fields, such as geology, material science or ecology, to the study of several phenomena. For example, this approach has been used to study the dependence of the spatial distribution of a given plant community on the plant stem diameter, or to study the relations between different species through their spatial distributions (Stoyan and Penttinen, 2000). For a review on current statistical methods applied to these mark point processes, see Illian et al. (2008, chap. 5).

Recently, methods of marked clustering statistics have been introduced for the study of the galaxy distribution. Beisbart and Kerscher (2000) used several statistics to analyse segregation by luminosity and morphological type in the Southern Sky Redshift Survey 2 (SSRS2, da Costa et al., 1998) catalogue. More recently, Skibba et al. (2009) used mark statistics to analyse SDSS data as function of the morphological properties obtained by the Galaxy Zoo[1] project (Lintott et al., 2008).

The aim of this chapter is to review some of the statistical methods associated with marked point processes in the context of the study of galaxy segregation, and to compare the different information they provide. We consider both the cases of

---

[1] http://www.galaxyzoo.org

continuous and discrete marks. In the former case, we use the normalised mark correlation function $k_{mm}(r)$ and the mark variogram $\gamma(r)$. In the case of qualitative marks, we use the standard partial correlation functions $\xi_{ij}(r)$ and introduce, for the first time in a cosmological context, the mark connection functions $p_{ij}(r)$. We illustrate the use of these methods using a volume-limited sample with a simple geometry drawn from the 2dFGRS. We use as a continuous mark the spectral type parameter η defined by Madgwick et al. (2002), and use this same parameter to divide our sample into two categories of 'early type' and 'late type' galaxies. In this way, we can cross-check our results for continuous and discrete marks. This work has been published as Martínez et al. (2010).

4.1 Clustering analysis methods for marked point processes

We will consider the galaxy distribution as a realisation of a marked point process X^M. This is a sequence of random marked points $X^M = \{(\mathbf{x}_i, m_i)\}$, where m_i is the mark associated to the point located at position $\mathbf{x}_i$. Typically, the objective of the analysis of this point process will be to determine whether there is a relation between the spatial clustering of the points and their marks, and to understand this relation.

The mark m_i can be a continuous variable, giving the value of a given physical property of the point. In the case of galaxies, this would typically be an observable property such as luminosity, colour, or some kind of spectral parameter. However, it could also be a property inferred indirectly, such as mass or star formation rate. Marks can also be discrete values, describing qualitatively different classes, or populations, of points. In the case of the galaxy distribution, these are typically 'early type' and 'late type' populations of galaxies, divided either by morphology, spectral characteristics, or colour. One could think also of different applications, e.g. studying the clustering properties of QSOs and galaxies. Obviously, one can construct qualitative marks derived from continuous ones, by binning the continuous marks into discrete categories. When dealing with discrete marks, we consider here only a *bi-variate* process with only two populations, but this can be easily extended to the *multivariate* case.

As a natural extension of the work with the two point correlation function, we focus on second-order statistics in the case of marked point processes. As in the unmarked case, we consider only homogeneous and isotropic point processes. The definition of the different statistics is based on the second-order intensity function $\lambda_2(r)$ for the unmarked point process. $\lambda_2(r)$ was introduced in Section 1.2. We remind here its definition, and extend it to the marked case below. Given two infinitesimal spheres of volume $\mathrm{d}V_1$ and $\mathrm{d}V_2$ separated by a distance r, the joint probability of finding a point of the process in each of the spheres is

$$\mathrm{d}P = \lambda_2(r)\mathrm{d}V_1\mathrm{d}V_2\,. \tag{4.1}$$

This intensity function is therefore related to the standard two-point correlation function as

$$\lambda_2(r) = n^2[1+\xi(r)]\,,$$

where n is the intensity of the point process.

4.1.1 Statistics for continuous marks

In the case of continuous marks, we use here two complementary statistics: the normalised mark correlation function $k_{mm}(r)$, and the mark variogram $\gamma(r)$. Let us first define, in analogy to equation (4.1), the quantity

$$\lambda_2^M[(\mathbf{x}_1,m_1),(\mathbf{x}_2,m_2)]\mathrm{d}V_1\mathrm{d}m_1\mathrm{d}V_2\mathrm{d}m_2$$

as the joint probability that in the volume element $\mathrm{d}V_1$ lies a galaxy with the mark in the range $[m_1, m_1+\mathrm{d}m_1]$ and another galaxy lies in $\mathrm{d}V_2$ with the mark in $[m_2, m_2+\mathrm{d}m_2]$ (Martínez and Saar, 2002).

Starting from λ_2^M, a number of second order statistics can be constructed, using different 'test functions' of the marks, $t(m_1,m_2)$, as

$$\kappa_t(r) = \frac{1}{\lambda_2(r)}\int\int t(m_1,m_2)\lambda_2^M[(\mathbf{x}_1,m_1),(\mathbf{x}_2,m_2)]\mathrm{d}m_1\mathrm{d}m_2\,. \tag{4.2}$$

Functions such as $\kappa_t(r)$ can be interpreted as a conditional quantity. It represents the expected value of the function $t(m_1,m_2)$ under the condition that there are two points in the process with marks m_1 and m_2 separated by a distance r. Equivalently, this is the value of $t(m_1,m_2)$ averaged over all pairs of points present in the process with separation in $[r, r+\mathrm{d}r]$ (Stoyan and Stoyan, 1994; Illian et al., 2008). One should define the statistic $\kappa_t(r)$ to use in each case depending on the physical problem, and on the function $t(m_1,m_2)$ of interest. We define below the most usual statistics, which can be applied to a large range of problems and, in particular, to the study of galaxy segregation.

The normalised mark correlation function (or, simply, 'mark correlation function') $k_{mm}(r)$ uses the test function

$$t(m_1,m_2) = \frac{m_1 m_2}{\bar{m}^2}\,,$$

where $\bar{m}$ is the mean of the marks. It is therefore defined as (Stoyan, 1984; Stoyan and Stoyan, 1994)

$$k_{mm}(r) = \frac{1}{\bar{m}^2\lambda_2(r)}\int\int m_1 m_2\lambda_2^M[(\mathbf{x}_1,m_1),(\mathbf{x}_2,m_2)]\mathrm{d}m_1\mathrm{d}m_2 \tag{4.3}$$

for $\lambda_2 \neq 0$. From this definition, $k_{mm}(r)$ represent the expected value of the normalised product of the marks for pairs of points separated by a distance r.

Values $k_{mm}(r) < 1$ indicate inhibition of the marks, i.e., pairs of points at a distance r tend to have smaller values of the marks than the average. On the other side, values $k_{mm}(r) > 1$ indicate that pairs of points at distance r tend to have larger values. In cases in which the marks are independent of the clustering pattern, $k_{mm}(r) = 1$. For example, in even-aged forests it is typically found that trees with larger stem diameter (mark) tend to be isolated, since they make use of much more ground and sun-light resources than smaller trees. In this case, therefore, we would have $k_{mm}(r) < 1$ at small scales, with $k_{mm}(r)$ increasing with distance.

In the case of the galaxy distribution, Beisbart and Kerscher (2000) studied the behaviour of $k_{mm}(r)$ using luminosity as the mark. They found $k_{mm}(r) > 1$ at small scales, implying stronger clustering of brighter galaxies at small separations, as expected.

Despite its name, the mark correlation function is not strictly a correlation function (Schlather, 2001). A true mark correlation function would be obtained by using the test function

$$t(m_1, m_2) = \frac{(m_1 - \bar{m})(m_2 - \bar{m})}{\sigma_m^2}\,,$$

where σ_m^2 is the variance of the mark distribution. However, $k_{mm}(r)$ is more commonly used and, as described above, provides important information about mark correlations.

The mark variogram $\gamma(r)$ (Wälder and Stoyan, 1996) is constructed analogously, using the test function

$$t(m_1, m_2) = \frac{1}{2}(m_1 - m_2)^2\,.$$

Its definition is therefore given by

$$\gamma(r) = \frac{1}{2\lambda_2(r)} \int\int (m_1 - m_2)^2 \lambda_2^M[(\mathbf{x}_1, m_1), (\mathbf{x}_2, m_2)]\mathrm{d}m_1\mathrm{d}m_2\,. \tag{4.4}$$

The mark variogram $\gamma(r)$ is a measure of the similarity of the marks depending on the separation between points. When the marks are independent of the clustering pattern, then $\gamma(r)$ is constant and equals the variance σ_m^2. In the presence of segregation, $\gamma(r) > \sigma_m^2$ indicates that pairs of points with separation r tend to have a larger difference between marks than the average, while $\gamma(r) < \sigma_m^2$ indicates that they tend to have similar marks.

Both statistics provide therefore complementary information. While the mark correlation function $k_{mm}(r)$ gives information about the values of the marks present in pairs at a given scale, the mark variogram $\gamma(r)$ is affected by the differences between those values in each pair.

4.1.2 Statistics for discrete marks

In the case in which the sample is divided in several discrete populations, we studied first the partial two-point correlation functions, and introduced also the mark connection functions.

The partial two-point correlation functions are simply the standard correlation function calculated for pairs of points in the different populations (Peebles, 1980). More formally, we define, analogously to equation (4.1), the quantity

$$\lambda_{ij}(r)\mathrm{d}V_1\mathrm{d}V_2$$

as the probability of finding in the volume element $\mathrm{d}V_1$ a point belonging to population i, and in the volume element $\mathrm{d}V_2$ a point belonging to population j. Then, the partial correlations functions are given by

$$\xi_{ij} = \frac{\lambda_{ij}(r)}{n_i n_j} - 1 \,, \tag{4.5}$$

where n_i is the number density of population i.

We have, therefore, that for $i = j$ the partial correlation functions are the correlation functions of the different populations calculated independently. For $i \neq j$, $\xi_{ij}(r)$ are the cross-correlations between the corresponding populations. The meaning of these function was already explained in Section 1.2. They measure the excess probability, with respect to a homogeneous Poisson process, of finding pairs of points with separation r. They are therefore a good measure of the clustering of the points in the different populations independently (for $i = j$), and of the correlations between points in different populations (for $i \neq j$).

In the case in which (discrete) marks are independent of the clustering pattern, the correlation functions will be similar for all the populations, so that

$$\xi_{ij}(r) \simeq \xi(r) \quad \forall i, j \,,$$

where $\xi(r)$ is the correlation function for the full sample. This follows from the invariance of the correlation function under thinning.

The mark connection functions $p_{ij}(r)$ measure conditional probabilities of the relative clustering between marks. In this sense, they are similar to the mark correlation function $k_{mm}(r)$ and mark variogram $\gamma(r)$ defined above for the case of continuous marks or, in fact, to any statistic $\kappa_t(r)$ defined similarly (as in equation 4.2). $p_{ij}(r)$ represents the conditional probability to find two galaxies of type i and type j at positions separated by a distance r, under the condition that at these positions there are indeed galaxies. Formally, we can define them as

$$p_{ij}(r) = \frac{\lambda_{ij}(r)}{\lambda_2(r)} \,,$$

where $\lambda_2(r)$ is the second order intensity function of the full unmarked process. They are therefore related to the partial correlation functions as

$$p_{ij}(r) = p_i p_j \frac{1+\xi_{ij}(r)}{1+\xi(r)} \,,$$

where $p_i = n_i/n$ is the probability that a randomly chosen point in the process is of the type i.

The mark connection functions yield information different to that from the partial correlation functions $\xi_{ij}(r)$ as shown, e.g., in Illian et al. (2008). By their definition they give the relative frequencies of mark pairs (i, j) at a distance r. While $\xi_{ij}(r)$ takes high values if there are many (i, j) pairs at a distance r, $p_{ij}(r)$ is high if the proportion of (i, j) pairs relative to the total number of pairs at that distance is high. So it may happen that, for some r, $\xi_{ij}(r)$ has a minimum, but $p_{ij}(r)$ has a maximum, if there is only a small number of point pairs at a distance r in the whole pattern, but many of them are exactly (i, j) pairs. Experience in other fields shows that often $p_{ij}(r)$ is able to find finer structures in point patterns than $\xi_{ij}(r)$ because of the nature of $p_{ij}(r)$ as a conditional probability.

If the marks are independent of the clustering pattern, then all $p_{ij}(r)$ are constant, and they take the values

$$p_{ij}(r) = \begin{cases} 2p_i p_j & \text{if} \quad i \neq j \\ p_i^2 & \text{if} \quad i = j \end{cases} \,. \tag{4.6}$$

In the presence of segregation, this is the expected behaviour of $p_{ij}(r)$ for $r \to \infty$.

4.1.3 Estimators used

We computed the different statistics based on the smooth estimation[2] of the second order intensity function $\lambda_2(r)$ presented in Stoyan and Stoyan (1994). This estimator works well for the correlation function at small scales (Pons-Bordería et al., 1999), and it is easy to extend for the estimation of the different statistics used here. It is given by

$$\widehat{\lambda}_2(r) = \frac{1}{4\pi r^2} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b\neq a}}^{N} \frac{k(r - |\mathbf{x}_a - \mathbf{x}_b|)}{V(W \cap W_{\mathbf{x}_a - \mathbf{x}_b})} \,. \tag{4.7}$$

Here, N is the total number of points in the sample, $\mathbf{x}_a$ are the positions of the points, and $k(\cdot)$ is a kernel function. Conventionally, W represents the observation window, i.e., the region of space in which we observed our sample, and $W_\mathbf{r}$ is this same region shifted by the vector $\mathbf{r}$, $W_\mathbf{r} = \{\mathbf{y} : \mathbf{y} = \mathbf{x} + \mathbf{r}, \mathbf{x} \in W\}$. The term

---

[2] As usual, we denote the estimators of the different quantities by putting a hat ($\widehat{}$) on top of the corresponding symbol.

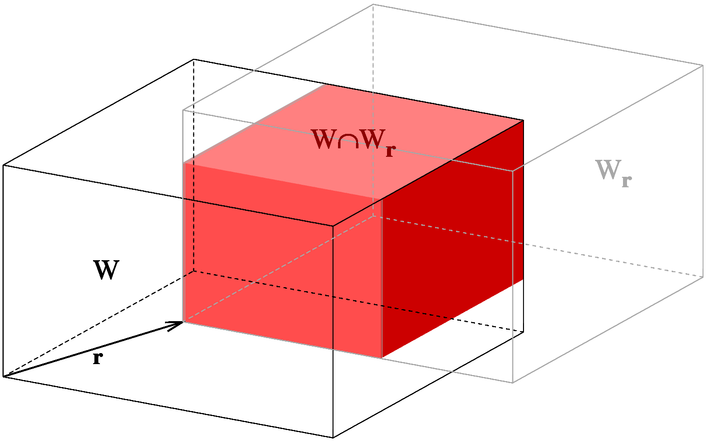


Figure 4.1: Diagram illustrating the definition of $V(W \cap W_{\mathbf{r}})$ used in our estimators. The shaded region represents $W \cap W_{\mathbf{r}}$, and its volume is given by equation (4.8).

$V(W \cap W_{\mathbf{r}})$ appearing in the denominator of equation (4.7) is the volume of the intersection of W with $W_{\mathbf{r}}$ (see Fig. 4.1). In the simple case in which the window W is a parallelepiped of dimensions L_x, L_y, L_z, this volume is given by

$$V(W \cap W_{\mathbf{r}}) = (L_x - |x|) \cdot (L_y - |y|) \cdot (L_z - |z|)\,, \tag{4.8}$$

with $\mathbf{r} = (x, y, z)$.

In our calculations, we used the Epanechnikov kernel, given by

$$k(x) = \begin{cases} \frac{3}{4w}\left(1 + \frac{x^2}{w^2}\right) & \text{for} - w \le x \le w \\ 0 & \text{otherwise} \end{cases}\,, \tag{4.9}$$

and characterised by the bandwidth w. This compact kernel is very well suited for correlation analysis (Pons-Bordería et al., 1999). We note however that the choice of a given kernel is not crucial, while the choice of the bandwidth w is more important and plays the role of the binning in the standard calculation of correlation functions, where a top-hat kernel is typically used as a default (see Section 1.2).

From this basic estimator, we can construct the different estimators for the mark statistics described above. In the case of the normalised correlation function, from equations (4.3) and (4.7) we obtain the estimator

$$\widehat{k}_{mm}(r) = \frac{1}{4\pi r^2 \bar{m}^2 \widehat{\lambda}_2(r)} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \frac{m_a m_b k(r - |\mathbf{x}_a - \mathbf{x}_b|)}{V(W \cap W_{\mathbf{x}_a - \mathbf{x}_b})}\,. \tag{4.10}$$

In a similar way, combining now equation (4.4) and (4.7), we obtain the estimator of the mark variogram $\gamma(r)$,

$$\widehat{\gamma}(r) = \frac{1}{8\pi r^2 \widehat{\lambda}_2(r)} \sum_{a=1}^{N} \sum_{\substack{b=1 \\ b \neq a}}^{N} \frac{(m_a - m_b)^2 k(r - |\mathbf{x}_a - \mathbf{x}_b|)}{V(W \cap W_{\mathbf{x}_a - \mathbf{x}_b})} . \tag{4.11}$$

In the case of discrete marks, from equations (4.5) and (4.7), we can estimate the partial correlations functions $\xi_{ij}(r)$ as

$$\widehat{\xi}_{ij}(r) = \frac{1}{4\pi r^2 \widehat{n}_i \widehat{n}_j} \sum_{a=1}^{N_i} \sum_{b=1}^{N_j} \frac{k(r - |\mathbf{x}_a^{(i)} - \mathbf{x}_b^{(j)}|)}{V(W \cap W_{\mathbf{x}_a^{(i)} - \mathbf{x}_b^{(j)}})} - 1 , \tag{4.12}$$

where $\mathbf{x}_a^{(i)}$ are the positions of points of the population i, N_i is the total number of points of population i, and the number densities of the different populations are estimated as $\widehat{n}_i = N_i/V$. The mark connection functions can be estimated from the partial correlation functions and the correlation function of the whole sample as

$$\widehat{p}_{ij}(r) = \widehat{p}_i \widehat{p}_j \frac{1 + \widehat{\xi}_{ij}(r)}{1 + \widehat{\xi}(r)} . \tag{4.13}$$

Here, the partial probabilities p_i are estimated as

$$\widehat{p}_i = \frac{N_i}{N} ,$$

and the global correlation function as

$$\widehat{\xi} = \frac{\widehat{\lambda}_2(r)}{\widehat{n}^2} - 1 .$$

4.1.4 Null test for segregation

It should be possible, in principle, to model the relation between the marks and the clustering properties of the point process considered. Some basic models are used in the statistics literature (Illian et al., 2008), such as the *random field model*, or the *random superposition model*. In the *random field model* the marks are assigned to each point from a continuous random field, which is independent of the non-marked point process which determines the points' positions. The *random superposition model* applies to discrete marks cases, and it assumes that each population is distributed according to a given point process, totally independent of the other populations. These kind of models are hardly justified when studying clustering segregation in galaxies, but one could in principle be able to make predictions based on galaxy formation models (Sheth et al., 2005).

In our case, however, we will not attempt a complete modelling of galaxy segregation. We will only consider a 'null model', in which points are *randomly marked*. That means that marks are drawn from a given probability distribution function, and are assigned randomly to the points in the sample, independently of their positions. This model has the advantage that it is very easy and fast to simulate. Given a marked sample, one can construct realisations for this null model by keeping fixed the positions of the points, but redistributing randomly the marks between them. These are called *random relabelling* simulations. If the results obtained for the original sample are significantly different from those obtained in the random relabelling simulations, this will indicate the presence of segregation between the marks, i.e., that marks are not independent of the clustering properties of the points' spatial distribution.

4.2 Data sample used

To illustrate the use of different mark clustering measures, we used a nearly volume-limited sample drawn from the 2dFGRS (Colless et al., 2001). This sample was prepared by the 2dFGRS team (Croton et al., 2004), and contains galaxies with absolute magnitudes in the range $-20 < M_{b_J} < -19$ at redshifts $z < 0.13$. In this chapter, we are interested in making a comparison between different methods, but not in a detailed modelling of galaxy clustering segregation. For this reason, we decided to use a simple geometry which would allow us to use the simple and fast estimators based on equation (4.7). Therefore, we selected galaxies inside a rectangular parallelepiped (with maximal volume) inscribed in the North slice of 2dFGRS. The dimensions of this parallelepiped are $254 \times 133 \times 31\,h^{-1}$ Mpc. This allows us to use the simple analytical expression for $V(W \cap W_{\mathbf{r}})$ given by equation (4.8). The final sample contains $N = 7741$ galaxies and covers a volume of $V \simeq 10^6\,h^{-3}\,\mathrm{Mpc}^3$. For all conversions from redshifts to distances, we used a flat fiducial cosmological model with $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$.

In order to assign marks to the galaxies, we characterised them using the spectral classification parameter η defined by Madgwick et al. (2002). This parameter was obtained from a principal-components analysis (PCA) of the galaxy spectra, and provides a continuous parametrisation of the importance of the absorption/emission features in them. Low values of η correspond to spectra with strong absorption features, while high values indicate strong emission lines. This η parameter is therefore related to physical characteristics of the galaxies, such as morphology or star formation rate.

Madgwick et al. (2002) studied the relation of their spectral classification to morphology using a sample of galaxies from the Kennicutt Atlas (Kennicutt, 1992). They found a clear correlation between the η parameter and morphological type, shown in Fig. 4.2. The early type (elliptical and lenticular) galaxies have low values of η, while the late type (spiral) galaxies have higher ones. Although

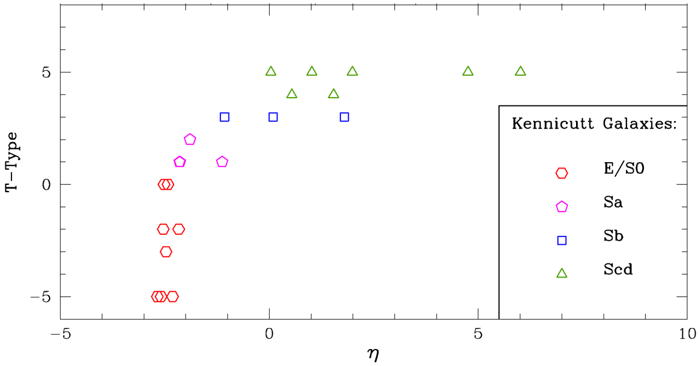


Figure 4.2: Relation between the spectral classification η and morphology for the galaxies in the Kennicutt (1992) sample. Figure from Madgwick et al. (2002).

this correlation allows a clear separation between early- and late-type galaxies, and even between different types of spirals, the η parameter can not differentiate between different elliptical types, as all of them have values $\eta \in [-3, -2]$.

The relation of the η parameter to the star formation rate of galaxies was studied by Madgwick et al. (2003a). They parametrised the star formation in galaxies using the birth-rate parameter b_S (Scalo, 1986), defined as the ratio of the present-day star formation rate in a galaxy to the mean of this rate through the galaxy's history:

$$b_S = \frac{SFR_{\mathrm{present}}}{\langle SFR \rangle_{\mathrm{past}}} \,. \tag{4.14}$$

Using mock catalogues built from stellar population synthesis models, they found a direct relation between η and b_S, shown in Fig. 4.3. Galaxies with low values of η are passive galaxies, with low rates of star formation, while galaxies with high η are actively forming stars.

We show the distribution of η values in our sample in Fig. 4.4. This distribution is very similar to the one obtained by Madgwick et al. (2002) for the full 2dFGRS sample. We see the presence of two 'bumps' in the distribution, corresponding to the two basic galaxy populations. The first population has a narrow distribution peaking at $\eta \simeq -3$, and corresponds to early type, passive galaxies. The second population has a distribution peaking at $\eta \simeq 0$ and with a large tail towards larger values of the parameter. It corresponds to late type, active galaxies.

It makes sense, therefore, to divide our sample into two populations, in order to illustrate the use of mark clustering methods associated to discrete marks. We follow Madgwick et al. (2002) and Madgwick et al. (2003b), and use a threshold

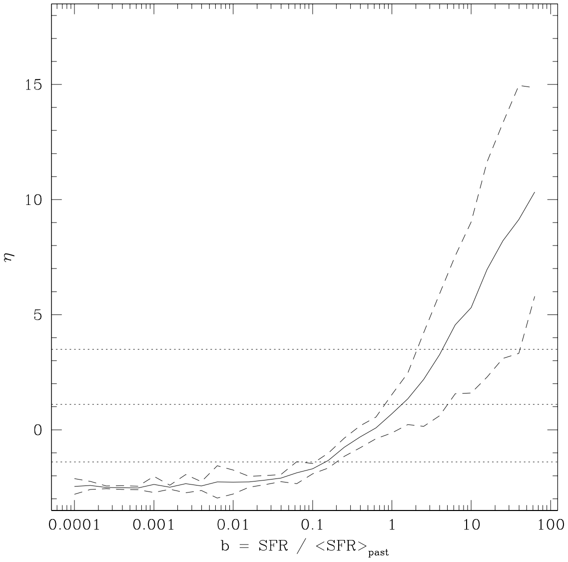


Figure 4.3: Relation between the spectral classification η and star-formation rate, parametrised by the parameter b_S (see equation 4.14). The solid and dashed lines show the mean and 1σ intervals of the values obtained for a mock catalogue built using a stellar population synthesis model. The horizontal dotted lines show the cuts in η used by Madgwick et al. (2002). Here we only use the lowest one, $\eta_{\rm thres} = -1.4$, which from the plot corresponds to $b_{\rm thres} \simeq 0.1$. Figure from Madgwick et al. (2003a).

of $\eta_{\rm thres} = -1.4$ to make this separation. We define our population '1' (early type, passive galaxies) as all galaxies in the sample with $\eta \leq -1.4$, and our population '2' (late type, active galaxies) with $\eta > -1.4$. These sub-samples contain $N_1 = 3828$ and $N_2 = 3913$ galaxies, respectively. Madgwick et al. (2003b) used this same division between populations to study clustering segregation in the 2dFGRS catalogue. Hence, using these populations, we can make comparisons to their work. However, we should have in mind that we are using a volume limited sample, with a narrow range in luminosities, while they use the complete magnitude limited sample from 2dFGRS. Fig. 4.5 shows the full sample we used, showing with different colours the two discrete populations.

When applying the methods associated to continuous marks, we used as our mark $m = \eta + 10$, in order to avoid negative values of the mark. This shift leaves invariant the value of the mark variogram $\gamma(r)$ (see equation 4.4), but it does affect our results for the normalised mark correlation function $k_{mm}(r)$ (equation 4.3).

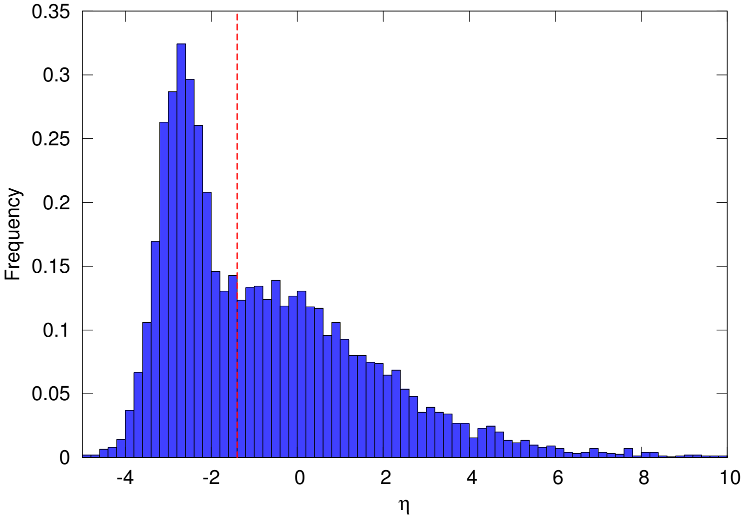


Figure 4.4: Distribution of the values of the spectral classification parameter η in the 2dFGRS sample used in this chapter. The vertical line corresponds to the value $\eta_{\text{thres}} = -1.4$ which we use to separate our two discrete populations of early type ($\eta < -1.4$) and late type $\eta > -1.4$ galaxies.

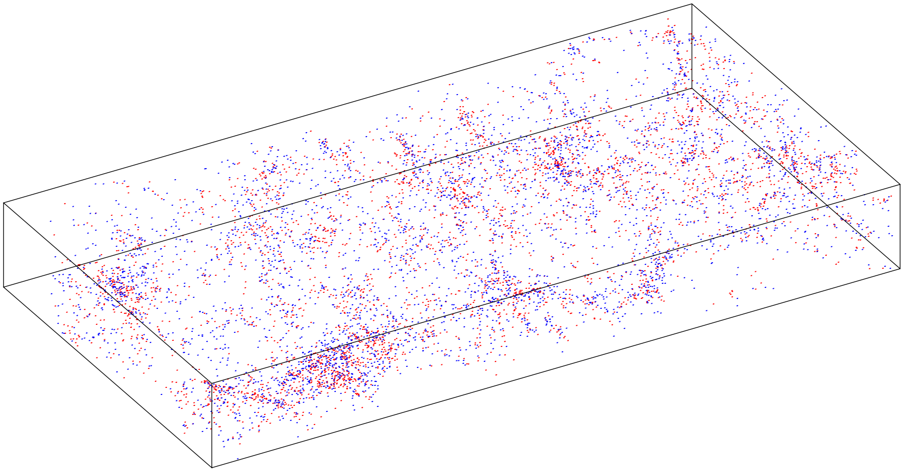


Figure 4.5: Three-dimensional plot of the galaxy sample used. Red dots correspond to early-type galaxies (population '1'), and blue dots to late-type galaxies (population '2'). The length of the parallelepiped sides are $254 \times 133 \times 31\, h^{-1}$ Mpc.

In this way, we follow the applications found in the mark statistics literature, where marks are almost always defined to be non-negative (Stoyan et al., 1995; Schlather, 2001; Illian et al., 2008). This allows us to interpret the different statistics, and $k_{mm}(r)$ in particular, in the way described in Section 4.1.1. As the absolute values of η have no physical meaning, but only its relative values, this shift does not affect our analysis. We should bear in mind that the actual values we obtain for $k_{mm}(r)$ do not have a direct physical meaning. However, the dependence of $k_{mm}(r)$ on scale and the comparison with the random relabelling simulations, do provide useful information about segregation, as explained below.

4.3 Results: segregation in 2dFGRS

In order to study galaxy segregation in 2dFGRS, we calculated the different statistics described in Section 4.1 for our sample. In all cases, we used a bandwidth of $w = 1\,h^{-1}$ Mpc for the Epanechnikov kernel (equation 4.9), and sampled the different functions with a step in r of $0.5\,h^{-1}$ Mpc.

In order to perform our null test for the existence of clustering segregation, we simulated $N_{\rm real} = 200$ random relabelling simulations, as described in Section 4.1.4, and obtained their maximum and minimum values as a function of the distance r for each statistic. Deviations of the observed statistics from this range of values correspond to a rejection of the mark-independent clustering model at a point-wise significance of $1 - \frac{2}{N_{\rm real}+1} \simeq 99\%$ (Illian et al., 2008).

We note that, in all our calculations in this chapter, we work in redshift space, and do not try to recover the real space clustering, as we did in Chapter 2. Therefore, assume that redshift distortions affect equally all galaxies, regardless of their mark. This is not strictly true (Madgwick et al., 2003b), but can be a good approximation for our purpose of illustrating the use of the different methods.

4.3.1 Results for continuous marks

We estimated the normalised mark correlation function $k_{mm}(r)$ for our sample according to equation (4.10). We show the result in Fig. 4.6. For small scales, the curve for $k_{mm}(r)$ shows a weak negative correlation, or spatial inhibition, $k_{mm}(r) < 1$. The range of correlation is about $20\,h^{-1}$ Mpc, where $k_{mm}(r)$ gets values close to 1, compatible with mark independent clustering. This means that galaxy pairs at small distances must have lower values of the mark η, i.e. have lower star formation rate than the average. It is interesting to compare this result with the $k_{mm}(r)$ function shown in Beisbart and Kerscher (2000) using the galaxy luminosity as the mark. They obtain a decreasing behaviour of $k_{mm}(r)$ at small scales, with $k_{mm}(r) > 1$ for $r < 12\,h^{-1}$ Mpc, showing that bright galaxies exhibit stronger correlation at these scales than faint ones.

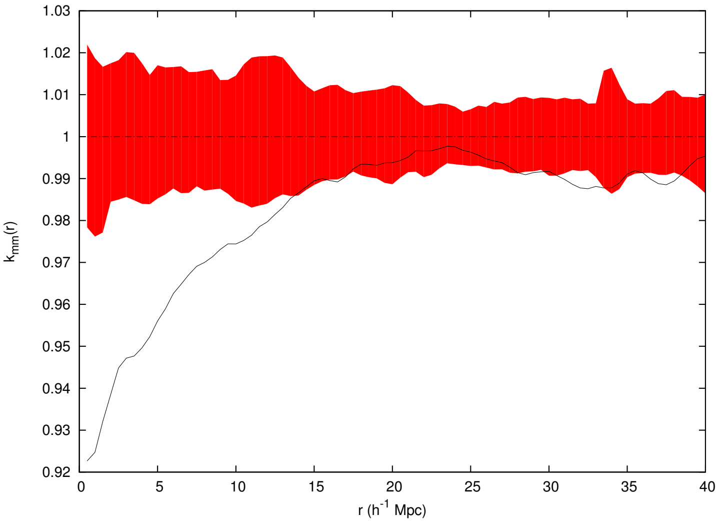


Figure 4.6: Normalised mark correlation function $k_{mm}(r)$ for our sample (continuous line). The shaded band shows the minimum and maximum values for the 200 realisations of the random relabelling simulation, while the dot-dashed line corresponds to the value for the case with no segregation, $k_{mm}(r) = 1$.

We also computed the mark variogram for our sample, using the estimator of equation (4.11). The result is shown in Fig. 4.7. This function is monotonously increasing, showing small values for small r. The range of correlation in this case is $r \simeq 10\,h^{-1}$ Mpc, at which scale $\gamma(r)$ becomes roughly constant and compatible with the mark independent clustering value, $\gamma(r) = \sigma_m^2$. The interpretation of this result is that, at small scales, galaxy pairs tend to have more similar marks, i.e. spectral type, than the average.

We can see from these results how the two statistics used, $k_{mm}(r)$ and $\gamma(r)$, provide complementary information useful to understand the interaction of the marks with the spatial correlations. If we had only the result for $k_{mm}(r)$ (Fig. 4.6), we would not know whether galaxy pairs at small distances were predominantly formed by two galaxies with relatively low values of η, or by one galaxy with a very low value of η and any other galaxy. Similarly, considering only the $\gamma(r)$ result (Fig. 4.7), we know that pairs with small separation have similar values of the marks, but we do not know whether these marks are low or high.

Combining both results, we get a clear picture. At the smallest scales, corresponding to the inner regions of clusters and groups, we have mostly galaxy pairs formed by two galaxies which both show low values of η. This effect decreases with

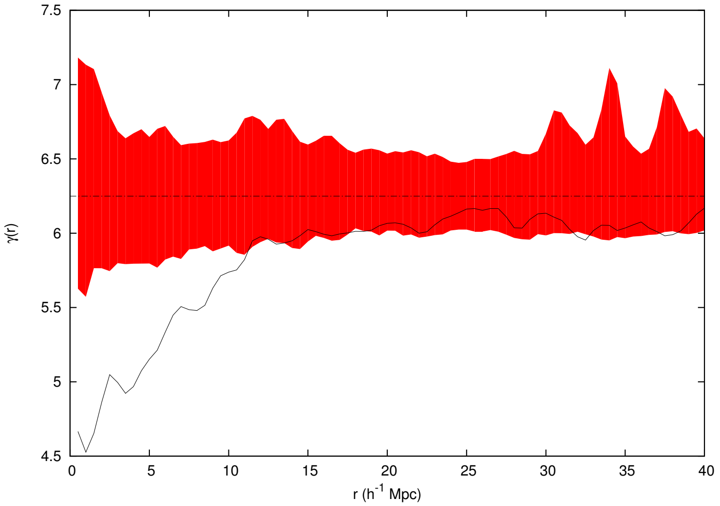


Figure 4.7: Mark variogram $\gamma(r)$ for our sample (continuous line). The shaded band shows the minimum and maximum values for the 200 realisations of the random relabelling simulation, while the dot-dashed line corresponds to the value for the case with no segregation, $\gamma(r) = \sigma_m^2 = 6.25$.

the scale, towards the asymptotic behaviour at the largest scales, at which spatial clustering is independent of the marks. The fact that the range of the correlation for $\gamma(r)$ ($r \simeq 10\,h^{-1}$ Mpc) is smaller than that for $k_{mm}(r)$ ($r \simeq 20\,h^{-1}$ Mpc) may hint at the existence of an intermediate regime at these scales. In this case, we would have a significant fraction of pairs with relatively different marks (so that $\gamma(r) \simeq \sigma_m^2$), but with one of the galaxies having a relatively low value of η (so that $k_{mm}(r) < 1$). However, the significance of the latter result is not very high. We compare below the information we obtained from $k_{mm}(r)$ and $\gamma(r)$ to that obtained when dividing the sample in two discrete population, and using the corresponding statistical methods.

4.3.2 Results for qualitative marks

We estimated the redshift-space partial correlation functions $\xi_{ij}(r)$ for our two discrete populations using the estimator in equation (4.12). As in this case we are measuring overall clustering properties of the samples, we also estimated the error of the measured correlation functions with the jackknife method (see e.g. Norberg et al., 2009). We divided the data volume in 32 equal, nearly cubic, sub-volumes. We generated the corresponding "mock" data sets omitting one of these sub-volumes at a time, and calculated the correlation functions for these.

The jackknife errors for each scale, $\sigma_{ij}(r)$ are then obtained as

$$\sigma_{ij}^2(r) = \frac{N_k - 1}{N_k} \sum_{k=1}^{N_k} \left[\xi_{ij}^k(r) - \bar{\xi}_{ij}(r)\right]^2 ,$$

where $\xi_{ij}^k(r)$ is the partial correlation function $\xi_{ij}(r)$ of the "mock" data set k, $\bar{\xi}_{ij}(r)$ is the value averaged over these data sets, and $N_k = 32$.

We show our results in Fig. 4.8, comparing each of the partial correlation functions to the correlation function $\xi(r)$ of the full sample. The three functions show the high degree of clustering present in the sample, specially at small scales, with correlation lengths[3] varying between $r = 6\,h^{-1}$ Mpc and $r = 10\,h^{-1}$ Mpc. We can see that the correlation function for galaxies of type '1' (early type and passive) has a larger amplitude for scales $r \lesssim 20\,h^{-1}$ Mpc than that for galaxies of type '2' (late type and active). The cross-correlation function $\xi_{12}(r)$, corresponding to mixed galaxy pairs, lies in between $\xi_{11}(r)$ and $\xi_{22}(r)$. This result is similar to that found by Madgwick et al. (2003b), although they used a larger, magnitude limited, sample, and worked in real space.

Finally, we also computed the mark connection functions for our two discrete populations, using the estimator defined in equation (4.13). We show the resulting $p_{ij}(r)$, together with the results form our 200 relabelling simulations, in Fig. 4.9. The nature of the mark connection functions as conditional probabilities make them filter away the overall clustering pattern of the sample. Therefore, we can study directly the relative clustering properties for the different populations, or discrete marks.

From Fig. 4.9, we see that the three mark connection functions show significant deviations from the results expected in the mark-independent clustering model. The function $p_{11}(r)$ shows that galaxy pairs of type '1' (passive) are more frequent than the average at small scales, up to a separation of $r \simeq 20\,h^{-1}$ Mpc, where $p_{11}(r)$ becomes flat. $p_{22}(r)$ shows the opposite behaviour, meaning that galaxy pairs of type '2' (active) are less frequent for these scales. In the crossed pairs case, $p_{12}(r)$ gets low values for small scales, but becomes flat for scales $r \gtrsim 10\,h^{-1}$ Mpc.

Overall, the mark dependence of clustering we obtain from the analysis of the $p_{ij}(r)$ is consistent with the results obtained using $k_{mm}(r)$ and $\gamma(r)$ with the continuous marks. Actually, the behaviour of the crossed mark connection function $p_{12}(r)$, showing a range of correlation shorter than that of $p_{11}(r)$ and $p_{22}(r)$ indicates more clearly the existence of an intermediate regime of clustering, approximately at scales $r \in [10, 20]\,h^{-1}$ Mpc. At these scales, there is a difference between the frequency of passive-passive and active-active galaxy pairs, but the frequency of mixed pairs corresponds to the overall mean in the sample. This shows the usefulness of the mark connection functions, which show directly the interactions between the marks. In contrast, in the case of the pair correlation

---

[3] We define here the correlation length as the scale r at which $\xi_{ij}(r) = 1$

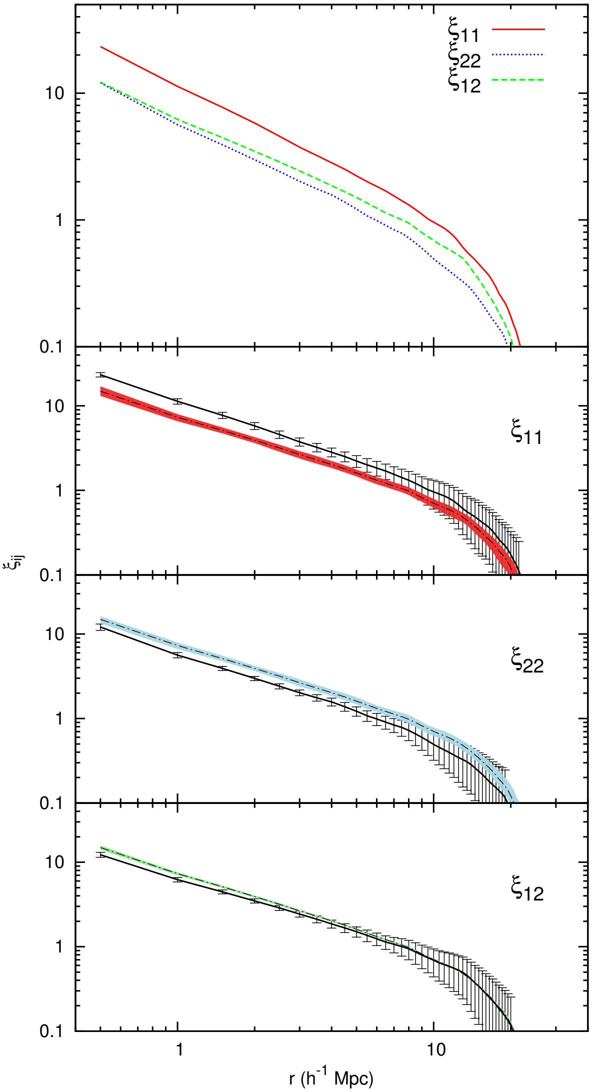


Figure 4.8: The partial two-point correlation functions, $\xi_{ij}(r)$, estimated for population '1' (early-type) and population '2' (late-type) galaxies in our sample. The top panel shows the three functions together. The three lower panels show each of them separately (solid lines with error bars estimated using the jackknife method), together with a shaded band showing the minimum and maximum values for the 200 realisations of the random relabelling simulation. The dot-dashed lines correspond to $\xi(r)$ for the full sample, which is the expected value of all $\xi_{ij}(r)$ in the absence of segregation

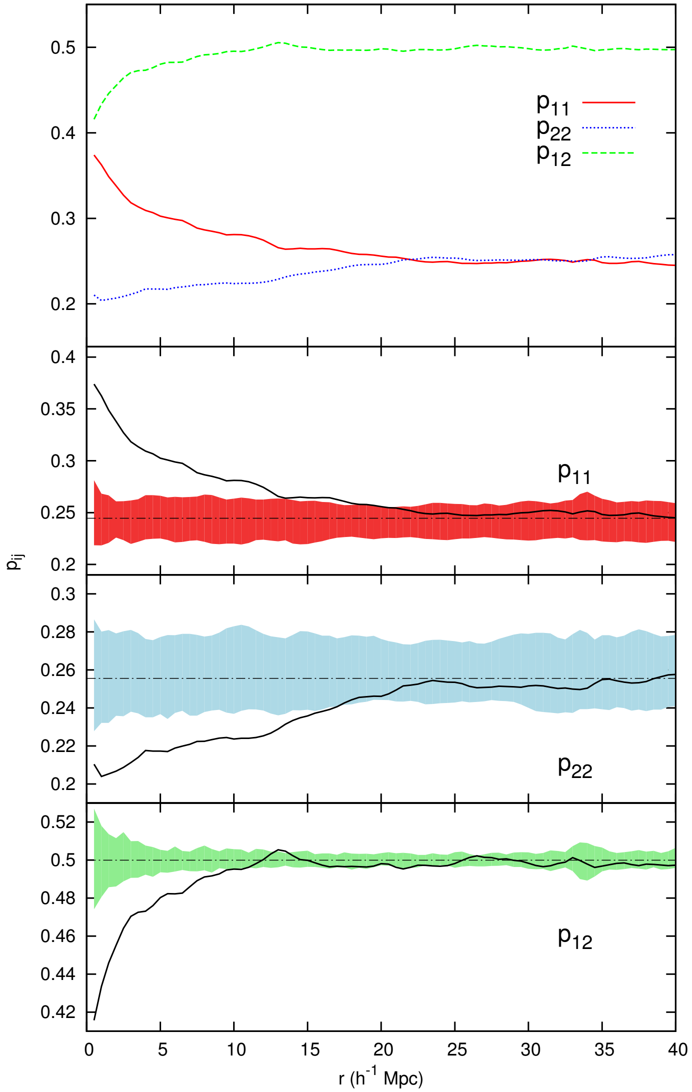


Figure 4.9: Mark connection functions, $p_{ij}(r)$ obtained for 'early-type' (population '1') and 'late-type' (population '2') galaxies in our sample. The top panel shows the three functions together. The three bottom panels show $p_{11}(r)$, $p_{22}(r)$, and $p_{12}(r)$ separately (solid lines), together with the shaded band showing the minimum and maximum values for the 200 realisations of the random relabelling simulation. The dot-dashed lines correspond to the expected values for the random labels case according to equation (4.6)

functions, the main effects come from the overall clustering pattern in the sample, and the relation of it with the marks becomes a second order effect.

4.4 Discussion

In this chapter, we have reviewed some of the statistical methods used to analyse clustering in marked point processes, and have applied them to the study of galaxy segregation. The methods used were the normalised mark correlation function $k_{mm}(r)$ and the mark variogram $\gamma(r)$ in the case of marks taking continuous values. For the case of qualitative marks, we studied the partial correlation functions $\xi_{ij}(r)$, and introduced, for the first time in the field of Cosmology, the mark connection functions $p_{ij}(r)$.

We tested the different methods for the analysis of galaxy clustering segregation using a volume limited sample drawn from the 2dFGRS. We used as continuous mark the spectral parameter η defined by Madgwick et al. (2002). This parameter can be directly related to the morphology and star formation rates of the galaxies. Moreover, using this parameter alone, one can clearly differentiate two populations, roughly corresponding to early- (low η) and late-type (high η) galaxies. We used these two populations as qualitative marks in our analysis. The use of discrete marks coming from a binning of η allows us to compare in a direct way the kind of information we extract using our statistics for continuous and discrete marks.

Overall, our galaxy sample forms a highly clustered pattern, as shown by the correlation function $\xi(r)$ of the full sample, with a correlation length of $r \simeq 8\,h^{-1}$ Mpc, and by the partial correlation functions $\xi_{ij}(r)$. The behaviour of these $\xi_{ij}(r)$ confirms qualitatively the results of (Madgwick et al., 2003b). Using the mark statistics, we can then disentangle the influence of the marks from this overall clustering.

When studying the continuous marks, we see that, at scales $r \lesssim 10\,h^{-1}$ Mpc corresponding to the inner regions of groups and clusters, galaxy segregation favours pairs in which both galaxies have small values of η. There is a hint of an intermediate regime for $r \in [10, 20]\,h^{-1}$ Mpc, in which there is still a tendency for low values of the marks, but not necessarily for both galaxies in the pair. This could correspond to correlations between galaxies inside groups and field galaxies. Finally, for the largest scales, $r \gtrsim 20\,h^{-1}$ Mpc, galaxy segregation has no effect, and clustering is independent of the marks. In order to get this detailed picture, it is important to use several mark statistics, as the mark correlation function $k_{mm}(r)$ and variogram $\gamma(r)$ in our case, which provide complementary information about the mark relations as function of pair separation.

When we consider the relation between the two populations as discrete marks, we need to use the mark connection functions $p_{ij}(r)$ in order to obtain similar information as above about the relative clustering of the marks. From the functions

$p_{11}(r)$ and $p_{22}(r)$ we see that, for $r \lesssim 20\,h^{-1}$ Mpc, pairs formed by two early-type galaxies are more frequent than the average, and the opposite occurs for pairs formed by two late-type galaxies. For crossed pairs, from $p_{12}(r)$, we see that they are less frequent than the average for $r \lesssim 10\,h^{-1}$ Mpc, but become abundant for larger scales. At scales $r \gtrsim 20\,h^{-1}$ Mpc all three functions are compatible with the clustering being independent of the marks. These details about the influence of the marks are very difficult to interpret from the partial correlation functions $\xi_{ij}(r)$, as they are dominated by the overall clustering signal.

We have shown, therefore, that the mark connection functions, being a conditional measure of the relation between marks, prove to be a useful tool to study galaxy clustering segregation. This tool is the equivalent to statistics such as the mark correlation function or variogram for the case of discrete populations. It can be applied, as in this case, to populations derived from some continuous mark characterising the galaxy, but also to populations defined by intrinsically discrete properties. The latter can be the case, for example, of galaxy populations defined by the morphology, or corresponding to completely different objects, such as galaxies and QSOs.

A possible application of the methods presented in this chapter is the study of the Butcher-Oemler (BO) effect. The BO effect (Butcher and Oemler, 1978, 1984) is the tendency of galaxy clusters at high redshift to have a larger fraction of blue galaxies than similar clusters at low redshift. This fact can be explained by the existence of mergers in the galaxy formation process (Baugh et al., 1996). Extending the BO effect to the overall galaxy distribution, not only to galaxy clusters, we expect it to have an effect on the segregaion of galaxies as function of redshift. In this way, we would expect that the segregation signal, as measured by the mark statistics methods presented in this chapter, is stronger for low redshift samples than for those at high redshifts. Therefore, studying segregation as function of redshift in deep surveys, such as ALHAMBRA, it should be possible to quantify the BO effect, and thus to obtain additional information about the process of galaxy evolution.

Part II

Clustering at large scales and baryon acoustic oscillations

5

Reliability of the detection of the baryon acoustic peak in the correlation function

In this chapter, we study the reliability of the detection of baryon acoustic oscillations (BAO) in the galaxy distribution. To this end we measure the correlation function at the appropriate scales for different samples drawn from the Sloan Digital Sky Survey (SDSS) and Two-degree Field Galaxy Redshift Survey (2dFGRS). We make a general introduction to BAO in Section 5.1. There, we explain their importance for the determination of cosmological parameters, and the status of previous detections in the galaxy distribution. Then, we describe the selection of our samples in Section 5.2, and the method we used to estimate the correlation function and its error in Section 5.3. Finally, we present our results in Section 5.4, and a brief discussion, including comparison to later results, in Section 5.5. The work presented in Sections 5.2 to 5.4 of this chapter was published as Martínez et al. (2009).

5.1 Baryon acoustic oscillations

We study in this chapter and the next one the signature left by baryon acoustic oscillations in the very large-scale distribution of galaxies at low redshift (Peebles and Yu, 1970; Hu et al., 1997; Eisenstein and Hu, 1998). This signature is interesting for several reasons. On the one side, the detection of this signature meant the confirmation of a prediction of the ΛCDM model. This feature links the present-day distribution of galaxies to the physics of the early universe, before recombination. Moreover, the fact of studying very large scales, around $\sim 100\,h^{-1}$ Mpc co-moving, means that the physics involved is very approximately linear, and thus this link is direct. On the other side, as we explain below, this feature is a useful tool to study the expansion history of the Universe, giving important information to constraint different cosmological parameters, and in particular the properties of dark energy. See Bassett and Hlozek (2010) for a recent review on the subject.

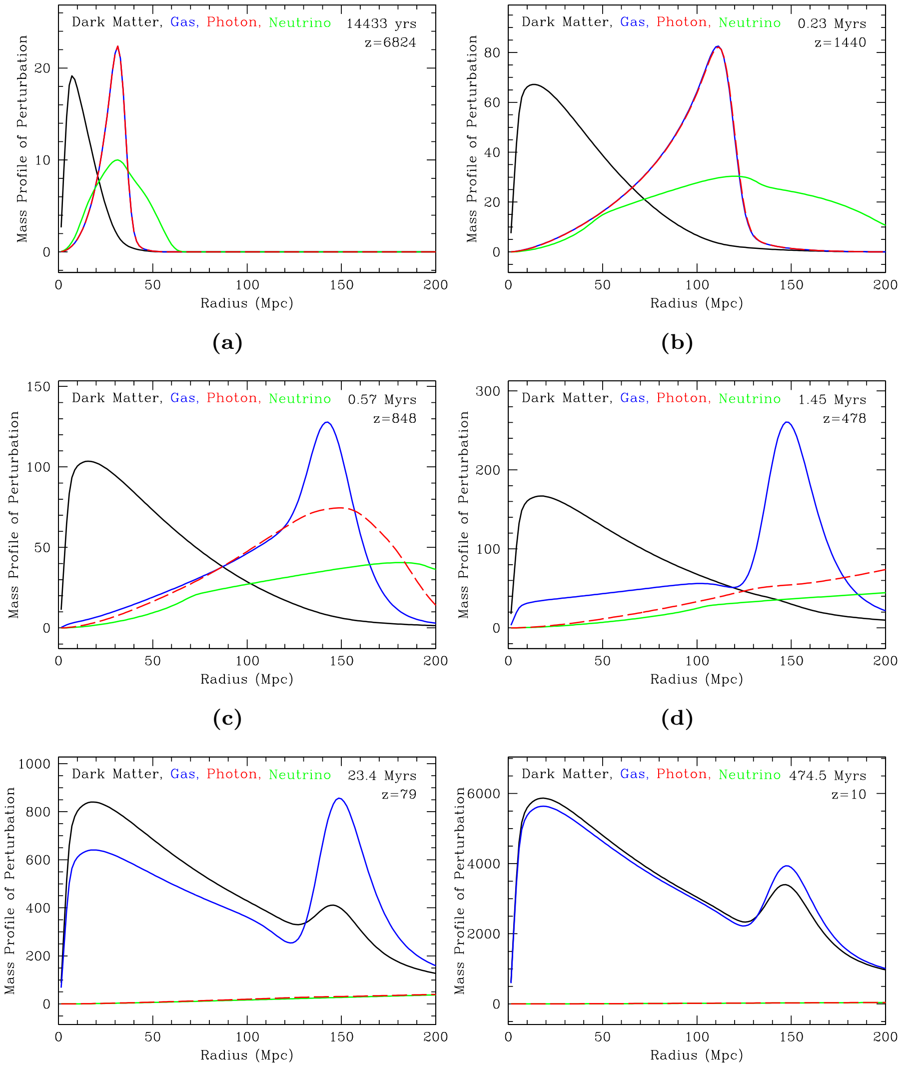
5.1.1 Physics of the baryon acoustic oscillations

For explaining the Physics describing the BAO in the early Universe, and their signature in the galaxy distribution, we follow Eisenstein et al. (2007), and work in configuration space. Although the description is totally equivalent to that in Fourier space, the explanation is slightly more intuitive.

Before recombination, the energy of photons is high enough to avoid the formation of neutral hydrogen atoms. This means that baryons and photons are coupled through Compton scattering and electromagnetic interaction between protons and electrons, forming a plasma. In this fluid two phenomena act in opposite directions: gravitational forces tend to compress the plasma around high-density regions, while radiation pressure tends to dilute any such over-density. The combination of both in the presence of any initial perturbation give rise to acoustic waves propagating in the baryon-photon plasma. This phenomenon ends abruptly at the epoch of recombination, when the temperature drops enough as to allow hydrogen atoms to form, and therefore the baryons decouple from radiation. During all this time, both cold dark matter and neutrinos evolve independently, as they are only coupled gravitationally to other species.

In order to understand what is the expected signature of BAO in the large-scale matter distribution, we can study the evolution of the density of the different species around a point-like initial perturbation. We represent some steps in this evolution in Fig. 5.1, reproduced from Eisenstein et al. (2007). The different panels correspond to different redshifts. In each case, the figure shows the mass profile for each of the species, as function of the co-moving distance r from the position of the initial perturbation. The mass profile plotted corresponds to the density multiplied by a factor of r^2, to avoid the volume effect one expects for an expanding spherical wave. We assume that this perturbation is adiabatic, and therefore the over-density is initially present in all the species.

The initial radiation over-density means that the pressure in this region is also higher than average. To compensate this overpressure, a spherical acoustic wave travelling outwards is produced. This wave drags both the photons and the baryons, as they are all coupled at this stage, overcoming gravity which tends to make all matter collapse towards the centre of the perturbation. This effect can be seen in Figs. 5.1a and 5.1b. This wave travels at the sound speed c_s in the



(a)

(b)

(c)

(d)

(e)

(f)

Figure 5.1: Evolution of the mass profile from an adiabatic, spherically symmetric, point-like initial perturbation. Each panel shows the mass profile at a given redshift (shown in the top right of the panel, together with the corresponding age of the Universe) for four different species: dark matter (black), baryons (blue), radiation (red), and neutrinos (green). The mass profile is equal to $r^2\rho(r)$, where $\rho(r)$ is the density profile in each case, and r is given in co-moving units. $\rho(r)$ are calculated as the Fourier transform of the corresponding transfer function, which was computed using the software CMBFAST (Seljak and Zaldarriaga, 1996; Zaldarriaga and Seljak, 2000). Figure taken from Eisenstein et al. (2007).

plasma, which is given approximately by (Eisenstein and Hu, 1998)

$$c_s = \frac{c}{\sqrt{3(1+R)}}\,,$$

where c is the light speed, and $R \equiv 3\rho_b/4\rho_\gamma$ is the baryon-to-photon momentum density ratio.[1] Meanwhile, both cold dark matter and neutrinos are de-coupled from the acoustic wave. Cold dark matter (CDM) feels no pressure, and therefore its density simply grows around the initial over-density by gravitational accretion. On the other side, neutrinos are (nearly) mass-less and therefore they are too fast and stream away from the perturbation.

This acoustic wave continues travelling outwards until the epoch of recombination. At $z_{\rm rec} \sim 1100$ (corresponding to $t_{\rm rec} \sim 380000\,{\rm yr}$), the photon energy has dropped enough to allow the formation of hydrogen atoms. This means that the baryons decouple from the photons, and they are not affected anymore by the radiation pressure. Therefore, as seen in Figs. 5.1c and 5.1d, the baryons that travelled out of the perturbation stop, and they now form a spherical shell surrounding the initial perturbation. Photons, meanwhile, stream away as free radiation. These free photons are the radiation that we observe nowadays in the form of the cosmic microwave background (CMB).

We are left with two matter over-densities, one near the centre which is mostly formed by CDM, and a shell with a co-moving radius of $\sim 150\,{\rm Mpc}$ around it, formed mainly by baryons. After recombination, these overdensities attract both baryons and CDM to them (see Figs. 5.1e and 5.1f). Therefore, at low redshift both overdensities have grown by gravitational infall, and the baryon fraction in both of them tends to the average. The final result in terms of the matter distribution is a large bump located at the position of the initial perturbation, and a faint shell surrounding it at a radius given by the sound horizon at recombination, i.e., the maximum distance the acoustic wave could travel between $t = 0$ and recombination,

$$r_s = \int_0^{t_{\rm rec}} c_s(t)(1+z){\rm d}t\,.$$

This radius depends on a complex way on the different cosmological parameters. However, a useful approximation is given, in co-moving units, by Eisenstein and Hu (1998) as

$$r_s = \frac{44.5\ln\left(\frac{9.83}{\Omega_m h^2}\right)}{\sqrt{1+10\left(\Omega_b h^2\right)^{3/4}}}\,{\rm Mpc} \simeq 150.7\,{\rm Mpc}\,,$$

where the numerical value was obtained for $\Omega_m = 0.27$, $\Omega_b = 0.05$ and $h = 0.7$. Hence, in the usual units used for LSS studies, $r_s = 105.5\,h^{-1}\,{\rm Mpc}$ in this case. The density ratio between the two overdensities depends basically on the baryon fraction, and it is on the order of ~ 0.01.

---

[1] As this stage corresponds to the radiation domination era, $R \ll 1$.

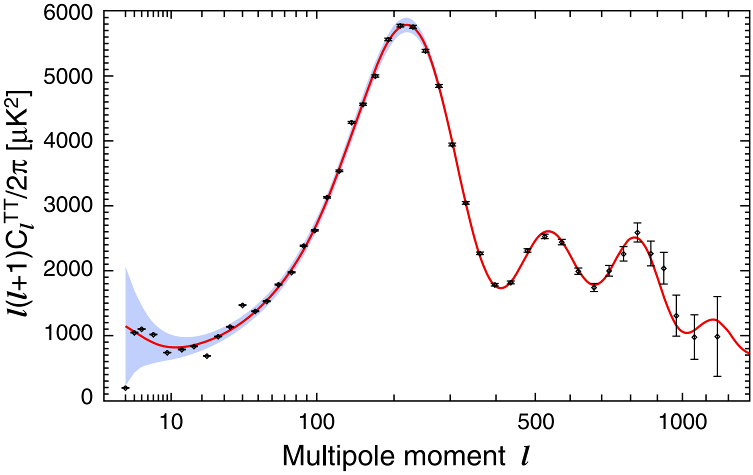


Figure 5.2: WMAP7 angular spectrum of the CMB inhomogeneities, showing the oscillations due to BAO. Figure from Larson et al. (2011).

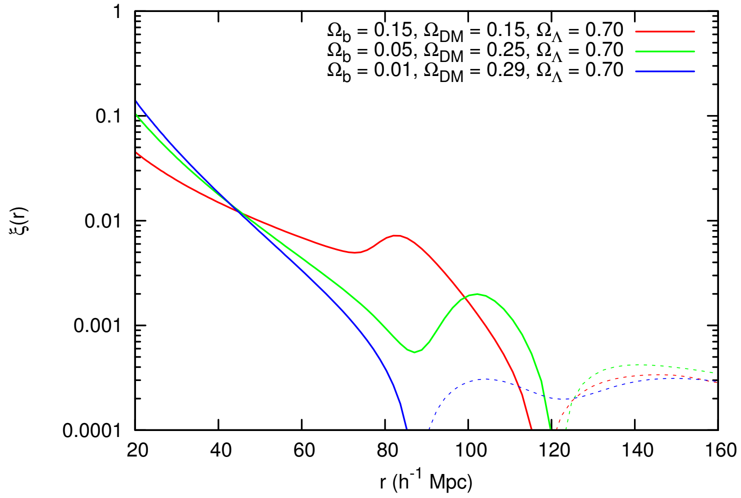
We can consider the early universe as a superposition of these point-like initial perturbations. Therefore, the present-day large-scale distribution of radiation and matter follows (in the linear approximation) from the convolution of this initial overdensity field with the individual solutions shown above. The effect in the radiation field is seen when studying the spectrum of the temperature anisotropies in the CMB (Fig. 5.2), as a series of wiggles, which are indeed acoustic features.

In the case of the matter distribution, the effect of the process described above is seen clearly in the two-point correlation function $\xi(r)$ as faint peak at the scale of r_s (Fig. 5.3a). The extra power at these scales correspond to the correlations between the central bump left at the position of the initial overdensity, and the acoustic shell surrounding it. The width of this peak depends on several factors, such as Silk damping, or the fact that the decoupling of baryons from photons at recombination does not happen instantly. At scales larger than this peak, $\xi(r)$ becomes negative. This is a consequence of the constraint

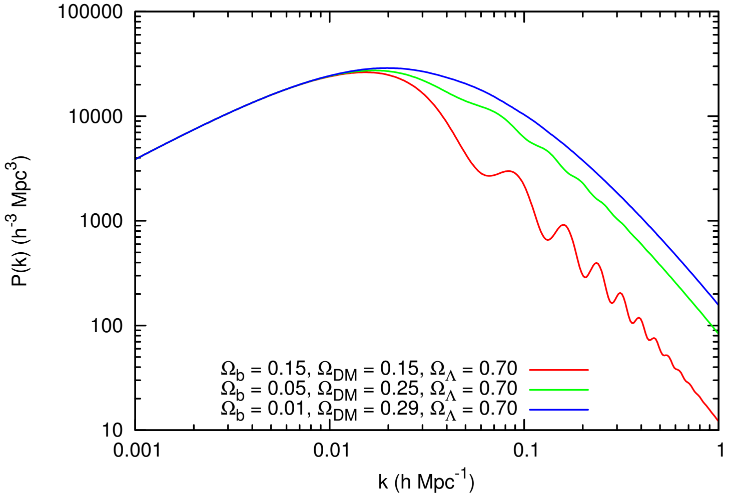
$$\int_0^\infty r^2 \xi(r) \mathrm{d}r = 0\,,$$

which requires $\xi(r)$ to become negative at some point. This condition is mathematically equivalent to the requirement $P(k=0)=0$ (Gabrielli et al., 2002), which is fullfilled in the case of the ΛCDM power spectrum (see Section 1.1.2).

The peak in the correlation function is converted into a series of oscillations or wiggles in the power spectrum $P(k)$ (Fig. 5.3b). The amplitude of the oscillations here is much smaller than those found in the CMB spectrum. This is due mainly



(a)



(b)

Figure 5.3: Correlation function (a) and power spectrum (b) of the matter distribution at $z = 0$ in the linear-theory ΛCDM model. The functions are shown for three different baryon fractions, with the parameters shown. In (a), the dashed lines represent negative values of $\xi(r)$. They were calculated using the Camb software (Lewis et al., 2000).

to the fact that only a small fraction of the total matter is part of the acoustic wave, while all the radiation is involved in the acoustic phenomenon.

5.1.2 Baryon acoustic oscillations as standard ruler

From the BAO model described above, we see that this feature provides us with a standard ruler which can be used to trace the expansion history of the Universe (Eisenstein et al., 1998b; Blake and Glazebrook, 2003; Seo and Eisenstein, 2005). We know that the scale of the BAO, $r_{\rm BAO}$, is given by the sound horizon r_s, and this scale remains fixed (if we neglect non-linear effects) in co-moving coordinates. Therefore, if we are able to measure the physical BAO scale from the galaxy distribution at a given redshift, this gives us information on the expansion rate $H(z)$, which depends on the Hubble constant H_0, the matter and dark energy densities, and the equation of state for dark energy (see Section 1.1.1). This provides a "geometrical" method to constraint combinations of these parameters. As shown above, the $r_{\rm BAO}$ scale can be measured from the position of the peak in $\xi(r)$, or the wavelength of the oscillations in $P(k)$.

The use of BAO in the low-redshift galaxy distribution as a way to measure cosmological parameters was first proposed by Eisenstein et al. (1998b). One of its main advantages is that the constraints obtained from BAO are complementary to those obtained from other observables, such as CMB anisotropies, type Ia supernovae, or the overall shape of the galaxy power spectrum. This means that the degeneracies between parameters are different for different observables, and hence combining them one can get very stringent constraints. When analysing BAO in the galaxy distribution, one normally takes the *a priori* value for r_s from the measurements of the CMB anisotropies, which determine it accurately. The results from WMAP5 (Komatsu et al., 2009) give a value of $r_s = 153.3 \pm 2.0\,{\rm Mpc} = 111.0 \pm 1.4\,h^{-1}\,{\rm Mpc}$ (where we used their best-fit value $h = 0.724$).

In the ideal case, one would like to measure independently the BAO scale in the line-of-sight and transverse directions, as these constrain different parameter combinations (Alcock and Paczynski, 1979). We can understand this difference if we assume we are measuring a spherical object at a fixed redshift z with a known co-moving size r_s. Following Hogg (1999), the line-of-sight co-moving distance between both sides of the object is then given by

$$r_s = d_c(z_2) - d_c(z_1) = \int_{z_1}^{z_2} \frac{c\mathrm{d}z'}{H(z')} \simeq \frac{c\Delta z}{H(z)}\,, \tag{5.1}$$

where $\Delta z = z_2 - z_1$ is the redshift difference we would actually measure between both ends of the object, and we assumed that $H(z)$ does not change significantly between z_1 and z_2. On the other side, the transverse co-moving size is given by

$$r_s = \Delta\theta(1+z)d_A(z)\ , \tag{5.2}$$

where $\Delta\theta$ is the angle we would actually measure, and $d_A(z)$ is the angular diameter distance. For a flat Universe, it is given by

$$d_A(z) = \frac{1}{1+z}\int_0^z \frac{c\mathrm{d}z'}{H(z')}\,.$$

Therefore, from the radial measurement Δz we could constrain directly $H(z)$ at the given redshift. From the transverse measurement $\Delta\theta$, the constraint could be on an integral over this function from now to redshift z, giving complementary information.

In practice, however, when measuring the BAO scale from a galaxy survey, one is not measuring a single Δz and $\Delta\theta$, but extracting the BAO scale using a statistical tool such as the correlation function. In order to do that, we need to transform the observed angles and redshifts into distances by using a set of fiducial values for the cosmological parameters. Assuming that the true parameters are close to the fiducial ones, one can just model this difference in parameters as a shift between the measured BAO scales $r^{\parallel}_{\mathrm{BAO}}$ and $r^{\perp}_{\mathrm{BAO}}$, and the known acoustic scale r_s. In this way, the parameters are obtained as

$$H(z) = \frac{r^{\parallel}_{\mathrm{BAO}}}{r_s} H^{\mathrm{fid}}(z)\ ,\ d_A(z) = \frac{r_s}{r^{\perp}_{\mathrm{BAO}}} d_A^{\mathrm{fid}}(z)\,,$$

where $H^{\mathrm{fid}}(z)$ and $d_A^{\mathrm{fid}}(z)$ are the values of these functions when using the fiducial parameters. In the most common case, moreover, one is not able to measure independently the BAO scale in the transverse and longitudinal directions, but only a spherically averaged scale r^{av}_{BAO}. In this case, we can proceed in the same way, but now we will constrain a combination of $H(z)$ and $d_A(z)$ given by the distance measure

$$d_V(z) \equiv \left[(1+z)^2 d_A^2(z)\frac{cz}{H(z)}\right]^{1/3}\,.$$

Using the scale shift method as above, we get

$$d_V(z) = \frac{r_s}{r^{av}_{\mathrm{BAO}}} d_V^{\mathrm{fid}}(z)\,.$$

However, when taking into account non-linear effects, the measured BAO scale does not correspond directly to the sound horizon at decoupling r_s as explained above. As these effects are well understood, the above equations can be modified to take into account this correction.

5.1.3 Previous detections of BAO in the galaxy distribution

In the early 1990s, several authors detected an excess clustering at scales close to $\sim 100\,h^{-1}\,\mathrm{Mpc}$ in the power spectrum of pencil beam surveys (Broadhurst et al., 1990; Szalay et al., 1993), and also of the slice-like Las Campanas survey (Landy et al., 1996). BAO were considered as a possible explanation for this excess. However, Eisenstein et al. (1998a) and Meiksin et al. (1999) showed that BAO could not account for the actual signal detected, for plausible values of the baryon and matter densities. Clearer hints for the presence of BAO in the power spectrum were obtained from the Abell cluster catalogue (Miller et al., 2002), and from early data from the 2dFGRS (Percival et al., 2001).

The first unambiguous detections of BAO were reported by Eisenstein et al. (2005) in the analysis of the correlation function of the Luminous Red Galaxies (LRGs, Eisenstein et al., 2001) catalogue from the SDSS Data Release 3 (DR3), and by Cole et al. (2005) in the power spectrum of the final 2dFGRS catalogue. Later, these detections were confirmed in later releases of the SDSS-LRG sample (Hütsi, 2006; Cabré and Gaztañaga, 2009), and of a combination of the SDSS Main and 2dFGRS samples (Percival et al., 2007a). The claimed significance of the BAO detection in these cases (although sometimes using divergent criteria) is at the $3 - 3.5\sigma$ level. Moreover, tentative detections have been claimed in the SDSS-LRG photometric catalogue (Padmanabhan et al., 2007), and in a galaxy cluster sample derived from SDSS data (Hütsi, 2010). In all these cases, the BAO signal was found in the angle-averaged, or “monopole”, two-point statistics (correlation function or power spectrum). Gaztañaga et al. (2009a) also found evidence for the presence of BAO in the SDSS-LRG sample using the three-point correlation function.

These results were used to put stringent constraints on the values of the cosmological parameters. Percival et al. (2007b, 2010) used the geometrical method described schematically above to constrain the distance measures to the redshifts of different galaxy samples. Other works (Tegmark et al., 2006; Sánchez et al., 2009; Reid et al., 2010) combined this method with a more thorough analysis of the full shape of the galaxy power spectrum, and different cosmological observables.

Okumura et al. (2008) were the first to attempt a separation between line-of-sight and transverse direction in the observed BAO signal. They found a ‘baryon ridge’ in the two-dimensional correlation function $\xi(r_\perp, r_\parallel)$[2] of the SDSS-LRG DR3 sample, corresponding to the BAO peak in the monopole. Later Gaztañaga et al. (2009b), using SDSS-LRG DR6, claimed a detection of the BAO peak in the correlation function along the radial direction only, and Gaztañaga et al. (2009c) explored the consequences of this signal on the values of the cosmological parameters. However, this detection is a controversial issue. Miralda-Escudé (2009) claimed, using statistics of pair counts, that the radial BAO detection could not be statistically significant. Kazin et al. (2010a) reproduced the results of Gaztañaga et al., but interpreted them in a different way, concluding that the radial BAO detection was not significant. Part of the divergences between these

---

[2] See Sections 1.2 and 2.1.

works come from different interpretations of the effect that lensing magnification would have on the BAO signal (Hui et al., 2007; Yoo and Miralda-Escudé, 2009). Recently, Tian et al. (2011) found an indication of the presence of the BAO peak in the line-of-sight direction from an analysis of the SDSS Main sample.

5.1.4 Aim of the work in this chapter

The aim of this chapter was to confirm the detection of the BAO peak in different galaxy samples drawn from the latest data from both SDSS-LRG and 2dFGRS galaxy surveys. These surveys cover different but slightly overlapping redshift ranges. Moreover, galaxies are selected in different ways in SDSS and 2dFGRS, so that the samples considered have intrinsically different properties, resulting in different clustering behaviour (Sánchez and Cole, 2008). Therefore, finding a compatible BAO signal in different galaxy samples would be a good indicator that this is a reliable feature of the matter density field, and not an artifact of a particular sample.

A further purpose of this work was to study the BAO feature, for the first time, in the last data release (DR7) of the SDSS-LRG catalogue, which is the largest galaxy survey to date, and to compare it to previous analysis using early versions of this catalogue. We follow here a qualitative approach when comparing the peak obtained in the correlation function of different samples, and do not perform a detailed calculation of the statistical significance of the detection.

In Chapter 6, we will further study the reliability of the detection of BAO in the galaxy distribution, but this time using a complementary approach. In that case we develop a new method, based on the use of a purpose-designed wavelet function directly on the density field traced by the galaxies. Apart from confirming the reliability of the BAO detection, this approach allows us to localise in configuration space the regions giving the largest contribution to the BAO signal.

5.2 SDSS and 2dFGRS data samples used

We studied the correlation function at the BAO scales for three samples drawn from the two largest redshift surveys to date, SDSS and 2dFGRS. A summary of the characteristics of the samples used, compared to the sample used by Eisenstein et al. (2005), is shown in Table 5.1. In the case of SDSS, we limited our study to the LRG sample. LRGs were selected by the SDSS team using several colour and magnitude cuts (Eisenstein et al., 2001) to obtain a highly biased sample reaching a higher redshift than the main, magnitude-limited part of the survey. They are therefore specially suited for correlation studies at large scales, and they constitute the galaxy sample covering the largest volume to date.

Sample	N	Absolute magnitude	z	Ω (sr)	V (h^{-3}Gpc^3)	$\bar{n}$ (h^3 Mpc^{-3})
DR3-LRG	46,748	$-23.2 < M_g^{0.3} < -21.2$	$[0.16, 0.47]$	1.16	0.75	6.3×10^{-5}
DR7-LRG	92,219	$-23.2 < M_g^{0.3} < -21.2$	$[0.16, 0.47]$	2.02	1.30	7.1×10^{-5}
DR7-LRG-VL	41,195	$-23.2 < M_g^{0.3} < -21.6$	$[0.16, 0.40]$	2.02	0.817	5.0×10^{-5}
2dFVL	32,388	$-21 < M_{b_J} < -20$	$[0.03, 0.19]$	0.45	0.024	1.4×10^{-3}

Table 5.1: Characteristics of the samples used and quoted in this chapter. DR3-LRG corresponds to the sample used in Eisenstein et al. (2005). For each sample, we give the number of galaxies N, the absolute magnitude range, the redshift (z) range, the effective solid angle covered Ω, the total volume V, and the mean galaxy number density $\bar{n} = N/V$. In the calculation of Ω and V we took into account the completeness mask for each survey. Absolute magnitudes M are normalised to $H_0 = 100$ km s^{-1} Mpc^{-1}.

We selected our two SDSS-LRG samples from the Data Release 7 (DR7, Abazajian et al., 2009). This was the last data release from the SDSS-II project, and corresponded to the completion of the original survey planned for the SDSS project.[3] The work presented here was the first study of galaxy clustering using this final SDSS catalogue. DR7 contains spectra for 206,797 LRGs within a total solid angle of 9380 deg^2. In order to minimise the influence of border corrections in our results, we restricted our work to a sub-sample of these data, corresponding to the main compact body of the survey footprint in the North galactic hemisphere. This area covers $\simeq 6800\,\mathrm{deg}^2$. From this catalogue, we selected two different samples, based on redshift and absolute magnitude cuts. The absolute magnitude used $M_g^{0.3}$ is the absolute magnitude in the g^* band at $z = 0.3$, which was K-corrected, and corrected for passive evolution, following Eisenstein et al. (2001) and Eyal Kazin (priv. comm.). Through this chapter, we always quote absolute magnitudes normalised to $h = 1$.

We selected a sample of LRG reproducing the selection used by Eisenstein et al. (2005) for the first BAO detection. This sample (which we label DR7-LRG) selects almost all the LRGs from the catalogue in the redshift range $z \in [0.16, 0.47]$, as it only applies a mild cut in absolute magnitude. In Table 5.1 we list the characteristics of this sample, and also, for comparison, the sample used by Eisenstein et al. (2005), labelled as DR3-LRG. We see that the DR7-LRG sample is about twice as large as the original DR3-LRG sample, which should imply a more reliable determination of the correlation function. Moreover, the sky footprint of the DR7 is much more compact than the DR3 one, so that the influence of border corrections in the calculations will also be smaller. DR7-LRG is not a volume-limited sample, but its number density changes with redshift, as shown in Fig. 5.4. The 'bump' in density seen at $z \simeq 0.35$ is due to the transition

---

[3] The SDSS is now being extended by the SDSS-III project, which involves several complementary surveys, see Eisenstein et al. (2011) for details.

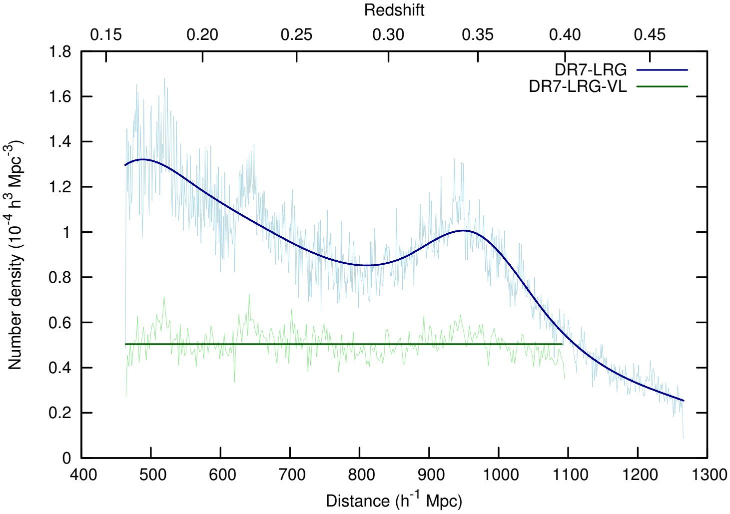


Figure 5.4: Number density of the SDSS samples used in this chapter as function of distance (in co-moving units) or, equivalently, redshift. For each of the samples, we show the measured density (thin line), and the smooth approximation we use for our calculations (thick line). In the case of DR7-LRG, this is a spline fit, while for the DR7-LRG-VL sample, we assume a constant density.

between the two different colour cuts used to select the LRGs (Eisenstein et al., 2001). We model the dependency of density on distance using a smooth spline fit, also shown in the figure.

We selected an additional sub-sample from the SDSS-LRG catalogue, which we label DR7-LRG-VL. In this case, we make a more stringent cut in absolute magnitude, $M_g^{0.3} < -21.6$, in order to obtain a real volume-limited sample in the range $z \in [0.16, 0.40]$. This sample contains galaxies more luminous on average than DR7-LRG, so we expect its bias to be larger. On the other side, the number density and size of the sample is smaller. In Fig. 5.4 we also show its number density as function of distance. We see that it is a reasonable approximation to assume that it is constant within the redshift range considered. We note that the SDSS samples used here are slightly different from those originally used in Martínez et al. (2009), but the results are nearly indistinguishable.

Finally, we also used a volume-limited sample drawn from the 2dFGRS. The 2dFGRS obtained $221,414$ galaxy redshifts distributed in two independent slices, one in the Northern and the other in the Southern galactic hemisphere, plus a set of random fields, which we do not use here. The survey covered a total of

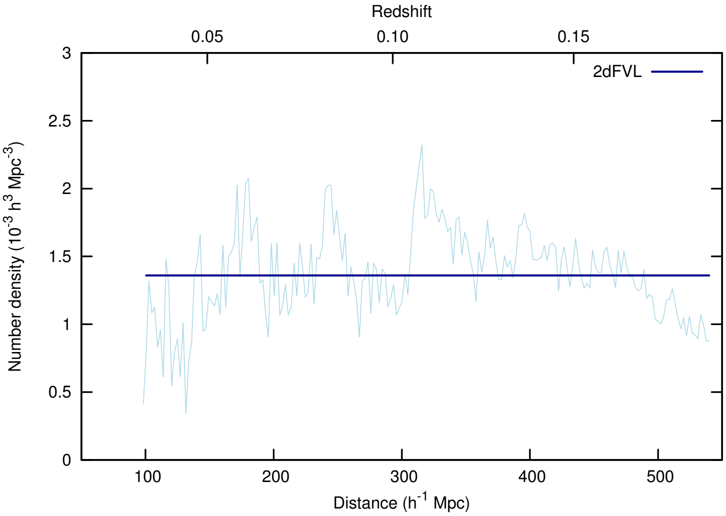


Figure 5.5: Same as Fig. 5.4, for the 2dFGRS sample used in this chapter.

$\simeq 1500\,\mathrm{deg}^2$ (Colless et al., 2003). The sample 2dFVL we use here was prepared by the 2dF team (Croton et al., 2004). It contains luminous galaxies, selected in absolute b_J magnitude by the cut $-21 < M_{b_J} < -20$, up to a redshift of $z < 0.19$. We show the number density as function of the distance for this sample in Fig. 5.5, where we see that it is approximately constant. The number density of this 2dFVL sample is more than an order of magnitude larger than that of the DR7-LRG sample.

Although the volume covered by the 2dFVL sample is smaller than that of the DR7-LRG, it is still useful to measure correlation on scales $\sim 100\ h^{-1}$ Mpc. As shown below, cosmic variance is not dominant, even if its effect is an order of magnitude larger for 2dFVL than for DR7-LRG. Moreover, the increase in density compensates the decrease in volume, so that the number of pairs of galaxies in each distance bin at these scales is similar in both samples. Hence, discreteness errors should be similar in both cases.

These three samples are important for the detection of the acoustic peak. Although the SDSS-LRG samples are larger and cover a redshift range less affected by nonlinear effects, the lower redshifts mapped by the 2dFVL sample are also important, providing the yardstick that can be compared with the characteristic BAO scales at larger redshifts. Moreover, by using galaxy samples with different selection functions and bias properties, we can assess the robustness of this detection, even if qualitatively.

In Fig. 5.6, we show two slices drawn from the DR7-LRG and 2dFVL samples.

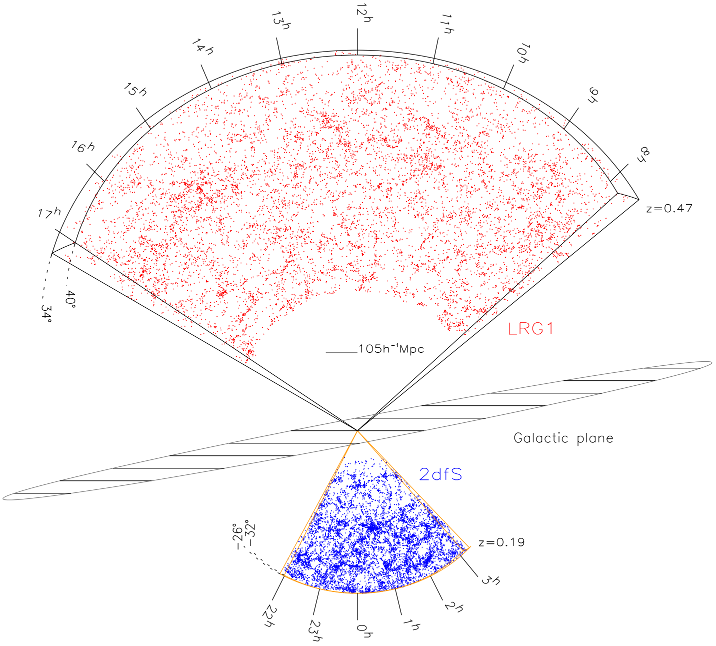


Figure 5.6: Configuration (redshift) space plot of two slices drawn from the DR7-LRG and 2dFVL samples. The large slice is drawn from the SDSS-LRG (DR7) survey. It is 6° wide in declination and the galaxy distribution is shown within the redshift range $0.16 \le z \le 0.47$. There are 10,136 LRGs within this slice depicted as red dots. The smaller slice with blue dots shows the galaxy distribution of 9,744 objects from the Southern Galactic hemisphere of the 2dFVL sample, reaching a depth of $z = 0.19$. To illustrate the scale of the acoustic peak a segment of length $105\,h^{-1}$Mpc is shown to scale.

This figure illustrates the differences in redshift range, volume and number density of these two samples. We also compare the sizes of the samples with the acoustic scale.

In all cases, we used a flat fiducial cosmological model with $\Omega_M = 0.27$, $\Omega_\Lambda = 0.73$ for the conversion of redshifts into co-moving distances.

5.3 Estimation of the correlation function

We estimated the spherically-averaged redshift-space correlation functions by using the Landy-Szalay border-corrected estimator (Landy and Szalay, 1993) that has good large-scale properties (see Section 1.2). We generated a random distribution of points following the selection function of each catalogue considered, and estimated the correlation function $\xi(s)$.

$$\widehat{\xi}_{\mathrm{LS}}(s) = 1 + \frac{DD(s)}{RR(s)} - 2\frac{DR(s)}{RR(s)} \tag{5.3}$$

where $DD(s)$, $RR(s)$ and $DR(s)$ are the probability densities of galaxy-galaxy, random-random and galaxy-random pairs, respectively, for a pair distance s. There are several recipes for the choice of the size N_{rd} of the random point set; as we are interested in large-distance correlations, where the numbers of pairs per bin (kernel width) are large, we used $N_{\mathrm{rd}} \simeq 5N$ (N is the number of galaxies in the sample). Increasing N_{rd} up to $20N$ led to point-wise differences less than a percent.

We estimate the probability densities by the kernel method, using the box spline of the third degree $B_3(x)$ (Saar et al., 2007). It is defined as

$$B_3(x) = \frac{1}{12}\left(|x-2|^3 - 4\,|x-1|^3 + 6\,|x|^3 - 4\,|x+1|^3 + |x+2|^3\right)\,, \tag{5.4}$$

and it has the important property of having compact support in $x \in [-2, 2]$. We sum the box spline $B_3(s/\Delta s)$ centred at each pair distance, taking in all cases a value of $\Delta s = 5\,h^{-1}$ Mpc. This results in an effective kernel width of $2\Delta s = 10\,h^{-1}$ Mpc. We sample the distributions at intervals of $1\,h^{-1}$ Mpc, smaller than the kernel width. In this way, we obtain a smooth estimation for the probability densities, and hence to $\xi(s)$. In any case, the information contained in this estimation is very similar to the one we would obtain using the usual bin estimation.

We generated the auxiliary random catalogues following the angular and radial selection function of the different samples. Regarding the radial selection, we assumed a constant density for the two volume-limited samples, DR7-LRG-VL and 2dFVL, and used our spline fit to the radial density function in the case of DR7-LRG (see Figs. 5.4 and 5.5).

To reproduce the angular selection of the SDSS catalogue (for our samples DR7-LRG and DR7-LRG-VL), we used the angular completeness masks provided by the Mangle software (Hamilton and Tegmark, 2004; Swanson et al., 2008). These masks are based on the DR7 New York University Value Added Catalogue (NYU-VAGC) prepared by Blanton et al. (2005). In the case of the 2dFVL sample, we used the 2dFGRS mask software provided by the 2dFGRS Team (Colless et al., 2003) to characterise the angular completeness of the survey.

The accepted statistical paradigm for the galaxy distribution is to model it as a Cox process – a Poisson point process where the local intensity is determined by a realisation of a random field. The two-stage nature of this process leads to two sources of errors of sample statistics (correlation functions, in our case). The first source is the sample (or cosmic) variance: how well the correlation function for the particular realisation of the random field in the volume considered reproduces the one defining the field. The second source is due to the discreteness of the point process. As cosmic variance and discreteness variance are independent, they add for the total variance.

The cosmic variance can be estimated, assuming that the random field is Gaussian, and that its expected correlation function and power spectrum are more or less known. The Gaussian assumption is valid at the large scales considered here. In this case, the covariance of the correlation function estimate can be found as a convolution of the correlation function itself, or as an integral over the power spectrum squared. A rough estimate is (Cohn, 2006)

$$\begin{aligned}\mathrm{Var}\left[\widehat{\xi}(r)\right] &\simeq \frac{1}{2\pi^2}\frac{1}{Vr^2}\int P^2(k)\mathrm{d}k \simeq \\ &\simeq 7\times 10^{-8}(V/h^{-3}\,\mathrm{Gpc}^3)^{-1}(r/100\,h^{-1}\,\mathrm{Mpc})^{-2}\,,\end{aligned}$$

where the numerical value was obtained by using the Eisenstein and Hu (1998) approximation for the power spectrum. The sample volume V and pair distance r in this formula are typical for the DR7-LRG and the baryonic peak. At these scales, the rms error is about 2×10^{-4} for DR7-LRG, and about 1.7×10^{-3} for the 2dFVL sample. As we shall see below, this is much smaller than the discreteness error in both cases, and we shall neglect it for the rest of this work.

Regarding the discreteness errors, we estimate them using a bootstrap method (Barrow et al., 1984) on the pair probability distributions $DD(s)$ and $DR(s)$ appearing in equation (5.3). These are the correct functions to bootstrap as, contrary to the correlation function itself, they are smooth functions of sample means (see, e.g. Efron and Tibshirani, 1993; Lahiri, 2003). As the sample points themselves are correlated, we apply the bootstrap method to (overlapping) blocks of points. The radius R of these blocks is chosen to approximately match the correlation length of each galaxy sample. We used $R = 12.0\,h^{-1}$ Mpc for the two SDSS samples, and $R = 6.5\,h^{-1}$ Mpc for our 2dFVL sample. For details on this error estimation method, and a test using a point process with known correlation function, see Martínez et al. (2009).

5.4 Results

Figure 5.7 shows the redshift-space correlation function we obtained for our DR7-LRG sample. We see that the function $\xi(s)$ remains positive in most of the range shown (up to $s \sim 180\,h^{-1}$ Mpc). It also shows confidently the BAO peak, with a maximum at $s = 102\ \ h^{-1}$ Mpc.

The inset shows the quantity $s^2\xi(s)$, in order to highlight the peak behaviour of the correlation function. We compare our estimate to the results for the original BAO detection from Eisenstein et al. (2005). We also plot here a model $\xi(r)$

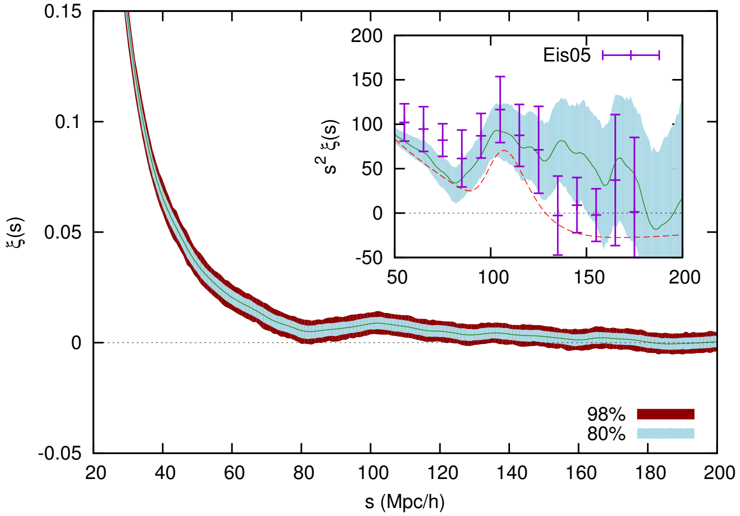


Figure 5.7: Redshift-space correlation function of the DR7-LRG sample. The main panel shows the 98% and 80% symmetric confidence regions in dark red and pale blue, together with the estimated function as a green line. In the inset, we show the quantity $s^2\xi(s)$ to highlight the large-scale behaviour, and compare it to the original discovery data from Eisenstein et al. (2005) (labelled as 'Eis05'), and to the ΛCDM model prediction (red dashed line, see the main text for details).

corresponding to the WMAP7 best-fit cosmological parameters (Komatsu et al., 2011). We calculated it from the power spectrum $P(k)$ obtained with the software Camb (Lewis et al., 2000), by a Fourier transformation. We only change the amplitude to match that of our observed $\xi(s)$ at $s = 40\,h^{-1}$ Mpc. This model therefore lacks some effects, such as those from redshift-space distortions. Moreover, we do not fit this model, but only show it for a fixed set of parameters, for illustration purposes.

We see that the DR7 data shows the BAO peak with a higher confidence than the original DR3 data, as seen by the consistently smaller error-bars obtained for $s \lesssim 120\,h^{-1}$ Mpc. The agreement with the model is also good at these scales, specially taking into account the caveats described above. However, there is a slight disagreement for scales $s \sim 120 - 150\,h^{-1}$ Mpc, as the peak we obtain is wider than that seen in DR3-LRG, and also than the theoretical expectations.

We show the correlation functions obtained for our two volume-limited samples in Fig. 5.8. Both our DR7-LRG-VL and 2dFVL samples exhibit the signature of a BAO peak, at approximately the same scale than our full DR7-LRG sample.

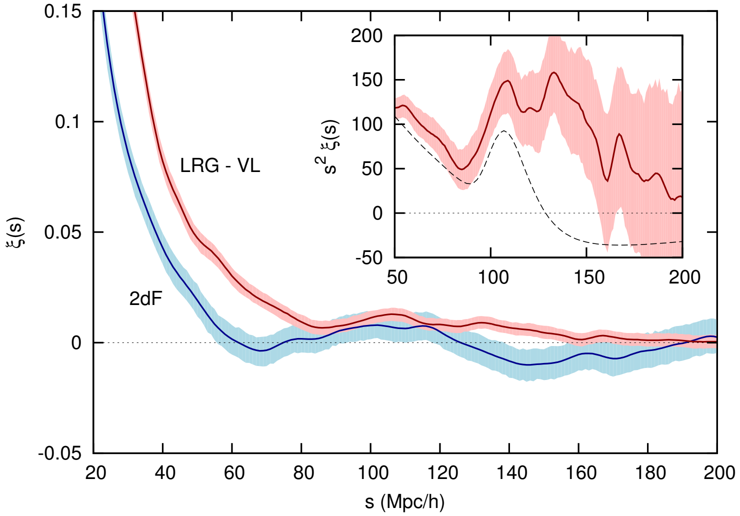


Figure 5.8: Redshift-space correlation function for our two volume-limited samples. The main panel shows the $\xi(s)$ estimate and 80% symmetric confidence band for the DR7-LRG-VL (red) and the 2dFVL (blue). The inset shows $s^2\xi(s)$ for the DR7-LRG-VL sample only, compared to the expected value in a simple ΛCDM model (dashed line), similar to that shown in Fig. 5.7.

However, the signal is weaker in these cases, specially the 2dFVL sample, as expected given the smaller volume covered here. At smaller scales, $s \lesssim 60\,h^{-1}\,\mathrm{Mpc}$, we see clearly the effect of the different bias values for the samples. The LRG-VL sample contains very luminous red galaxies, while the 2dFVL sample contains much fainter galaxies selected in the blue. This results in an approximate bias ratio of $b_{LRG-VL}/b_{2dFVL} \simeq 1.4$ at $s \simeq 40\,h^{-1}\,\mathrm{Mpc}$.

In the inset of Fig. 5.8 we show in more detail the large-scale behaviour for the DR7-LRG-VL, and compare it to a model $\xi(r)$ calculated as before for DR7-LRG. As expected, this sample exhibits a behaviour very similar to DR7-LRG at these scales, although with a larger bias ($b_{\mathrm{LRG-VL}}/b_{\mathrm{LRG}} \sim 1.2$ at these scales). It also shows that the detected BAO peak is wider than expected from the theoretical model.

In the case of the 2dFVL sample, the correlation function crosses zero at $s \simeq 60\,h^{-1}\,\mathrm{Mpc}$, reaching a local minimum with $\xi(s) < 0$ at $s \simeq 65\,h^{-1}\,\mathrm{Mpc}$. It has been argued that this zero crossing implies a deviation from the standard ΛCDM model (Sylos Labini et al., 2009; Sylos Labini, 2010), as this model predicts the crossing at larger scales, and the discrepancy can not be attributed to differences in galaxy bias. However, Kazin et al. (2010b) used a set of N-body simulations to show that one would expect this behaviour in $\sim 7\%$ of cases due to sample variance, even for LRG-sized samples. At $s > 120\,h^{-1}$ Mpc, $\xi(s)$ takes negative values again. In this case, this result agrees with the theoretical model, as it requires $\xi(r)$ to become negative at large scales (see Fig. 5.3a).

5.5 Discussion

Overall, we have shown that the baryon acoustic peak is a reliable feature of the galaxy correlation function, as it is found in several different samples. In fact, this was the first reported detection of the peak in the correlation function of a galaxy sample drawn from 2dFGRS. The fact of observing the peak in both the SDSS and the 2dFGRS is specially interesting as those two surveys select galaxies with different properties, and thus this indicates that the peak is not an artifact of the galaxy – dark matter bias, but a feature in the large-scale matter distribution.

Later works have reinforced this conclusion, by also detecting the BAO signature in different LRGs samples from SDSS-DR7 (Sánchez et al., 2009; Kazin et al., 2010b; Reid et al., 2010; Percival et al., 2010), and also in other two low redshift samples, the SDSS Main sample (Tian et al., 2011) and the 6dF Galaxy Survey (Beutler et al., 2011). Recently, Blake et al. (2011) also detected the BAO signature in a spectroscopic sample in the redshift range $z \in [0.3, 0.9]$ drawn from the WiggleZ survey (Drinkwater et al., 2010).

However, the actual statistical significance of the different BAO detections is still a controversial topic. Although most papers studying the BAO signal in SDSS-LRG claim detections in the range $3 - 3.5\sigma$, Cabré and Gaztañaga (2011) are not finding such level of detection. In that work, they clearly separate model selection from the determination of the parameters. Using similar data as we used in this chapter, they claim only a $\sim 2.2\sigma$ preference for the full ΛCDM model compared to a no-BAO model.[4] They do not claim this result to be in contradiction with the standard ΛCDM model, but to be a consequence of insufficient data. In spite of their low-level detection, they do find that, assuming the ΛCDM model to be correct, an accurate determination of the BAO scale, and its use as standard ruler (see Section 5.1.2) is still possible. This contradicts the view of Bassett and Afshordi (2010), who argue that low-level detections are not sufficient to robustly estimate the cosmological parameters.

From the results in this chapter, we also see that the observed BAO peak in SDSS-LRG is much wider than expected, and also than that found in the original detection data by Eisenstein et al. (2005). This same effect has been reported in other works dealing with SDSS data (both in DR6 and DR7), see e.g. Cabré and

---

[4] Cabré and Gaztañaga (2011) actually only found a $\sim 1\sigma$ significance when using the data from the original detection by Eisenstein et al. (2005), while in the original work a 3.4σ significance was claimed, although computed in a different way.

Gaztañaga (2009); Kazin et al. (2010b). Although this may just be a statistical fluctuation, it may also be an indication of a deviation from ΛCDM at very large scales. In fact Labatie et al. (2011), using a simple comparison to a fixed ΛCDM model, similar to that shown in the inset of Fig. 5.7, found a $\sim 2.7\sigma$ deviation at these scales. This behaviour may be related to the extra power observed at large scales at some photometric samples at higher redshift (Sawangwit et al., 2009; Thomas et al., 2011).

6

Wavelet analysis of baryon acoustic structures

As we have seen in the previous chapter, baryon acoustic oscillations (BAO) are an interesting feature of the galaxy distribution at large scales. They produce a faint signal in the form of a peak at a scale of $\simeq 105\,h^{-1}\,\mathrm{Mpc}$ in the correlation function. This signal has been detected in the correlation function and power spectrum of several of the largest galaxy samples to date (see Section 5.1.3, and our results in Section 5.4). However, the actual significance of this detection is a controversial topic (Bassett and Afshordi, 2010; Cabré and Gaztañaga, 2011).

It is important, therefore, to find evidence of BAO in the galaxy distribution based on complementary methods. A step further is to search for real structures in the galaxy distribution that are responsible for the BAO feature in these second order statistics. The detection of these structures would be a confirmation of the existence of the baryon acoustic phenomenon. Moreover, if we are able to localise these structures in configuration space, this would allow us to study in more detail the properties of BAO.

In this chapter, we introduce a new method for the detection of BAO, which is closely tied to the underlying physics of the process, explained in Section 5.1.1. This method (described in Section 6.1) is based on analysing directly the 3D galaxy distribution using a very specific wavelet function (which we called 'BAOlet'), which is specially designed to search for BAO features. The method makes use of two different tracers, one to map the overall density field (including the BAO shells), and the other to locate the position of the largest overdensities, which should correspond to centres of the shells. Moreover, as we study directly the galaxy distribution in configuration space, this method gives valuable information about the localisation of the BAO signal in the volume studied. This allows us to identify regions of space where the BAO signal is stronger or fainter.

We study the expected signal using this method in the ΛCDM model in Section 6.2, using both analytical prediction and a N-body simulation catalogue. We then apply this method to a galaxy catalogue drawn from the Sloan Digital Sky Survey (SDSS), consisting on a sample of Luminous Red Galaxies (LRGs), and a sample drawn from the 'Main' catalogue. We describe these samples in Section 6.3, and show the results obtained in this case in Section 6.4. We make a simple test to assess the significance of these results, and explore the implications of this analysis regarding the localisation of BAO structures. We finally discuss our results and possibilities for future work in Section 6.5. The work in this chapter has been submitted for publication as Arnalte-Mur et al. (2011).

6.1 The wavelet detection method

The basis of the new BAO detection method we present here is to focus on the positions of massive dark matter haloes, which correspond to the location of large initial perturbations, and to look for the presence of structures resembling the acoustic shells around these. Once we locate the positions of the large overdensities, we need to study the density field to identify the structures corresponding to the acoustic shells around these centres.

An appropriate method for identification of structures in continuous fields is wavelet analysis (Farge, 1992; Chui, 1992; Starck and Murtagh, 2006; Jones, 2009). Wavelet transforms are a kind of integral transforms designed to localise information in both scale and in space. In this sense, they mean an improvement with respect to Fourier transforms, which are suited to localise the signal in scale, but lose all spatial information. Therefore, wavelets can be used for the analysis of data at different scales, and identification of characteristic patterns or structures.

Wavelet transforms are widely used in many areas, especially in image analysis (Mallat, 2008; Starck et al., 2010). They have been used in Cosmology for the analysis of the large-scale structure, and of the CMB anisotropies (Martínez et al., 1993; Rauzy et al., 1993; Vielva et al., 2004; Starck et al., 2006; Saar, 2009).

In the case of the detection of BAO, we first design a family of wavelets adapted for the detection of shell-like structures (our 'BAOlet', Section 6.1.1). We then use this wavelet transform to map the presence of these shells in the sampled volume, as function of radius and width, through the wavelet coefficients $W_{R,s}$ defined below. We illustrate this idea using a simple toy model in two dimensions in Section 6.1.2. However, as explained in Section 5.1.1, the density field we analyse is a convolution of the BAO profile (Fig. 5.1) with the initial overdensity field. As this field has a mean of zero, BAO could not be detected using the usual summary statistics on $W_{R,s}$, which are linear in density. Our approach, then, is to look for the exceptional cases of massive haloes, corresponding to large initial perturbations, where we expect the BAO signal to be correspondingly high on average. This is the basis for our statistic $B(R, s)$ presented in Section 6.1.3 which, once the positions of these haloes are independently identified, has a positive expected value, and can therefore be used for the detection of BAO.

6.1.1 Design of the BAOlet function

Standard wavelet functions, like the Mexican hat, are typically designed for the detection of simple overdensities, or 'bumps', of different scales. They are therefore not suitable for the detection of shells. Instead, we need a family of wavelets whose shape matches the type of structures we want to find in our data. Therefore, for this work we use a specially designed wavelet (the 'BAOlet'), well adapted to the search of BAO features – shell-like structures around our selected centres.

We design this new family of wavelet functions as a transformation of the wide-used B-spline wavelets (Saar, 2009). These $\psi_{R,s}(\mathbf{x})$ functions are spherically symmetric, and their radial profiles are defined as

$$\psi^r_{R,s}(r) = \frac{\alpha_{R,s}}{4\pi r^2}\left[2B_3\left(2\frac{r-R}{s}\right) - B_3\left(\frac{r-R}{s}\right)\right] , \tag{6.1}$$

where R and s are the two parameters that define the scale (radius) and width of the BAOlet function, $\alpha_{R,s}$ is the normalisation constant defined so that

$$||\psi_{R,s}||^2 \equiv \int |\psi_{R,s}(\mathbf{x})|^2 \mathrm{d}^3\mathbf{x} = 1 \,, \tag{6.2}$$

and $B_3(x)$ is the box spline of the third degree, defined by

$$B_3(x) = \frac{1}{12}\left(|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3\right) .$$

The BAOlet function is shown in Fig. 6.1. It can be thought of as a spherical shell of radius R and width s, with zero amplitude at its centre and therefore adapted to the detection of spherical shells of a given radius. This specific choice is motivated by the fact that the integrated profile is the widely-used one-dimensional 'B-spline' wavelet function that has a null mean and compact support $[-2, 2]$. These properties directly translate onto the BAOlet that has also a null mean –a requirement for any wavelet function– if $R > 2s$, and takes non-zero values only for $R - 2s \leq |\mathbf{x}| \leq R + 2s$.

Given a density map $\delta(\mathbf{x})$, properly normalised as in equation (6.2), we can construct, for each pair of values in the parameter space (R, s), a BAOlet coefficient map as the convolution of our density field with the corresponding wavelet:

$$W_{R,s}(\mathbf{x}) \equiv \int_{\Re^3} \psi_{R,s}(\mathbf{y})\delta(\mathbf{y} - \mathbf{x})\mathrm{d}^3\mathbf{y} \,. \tag{6.3}$$

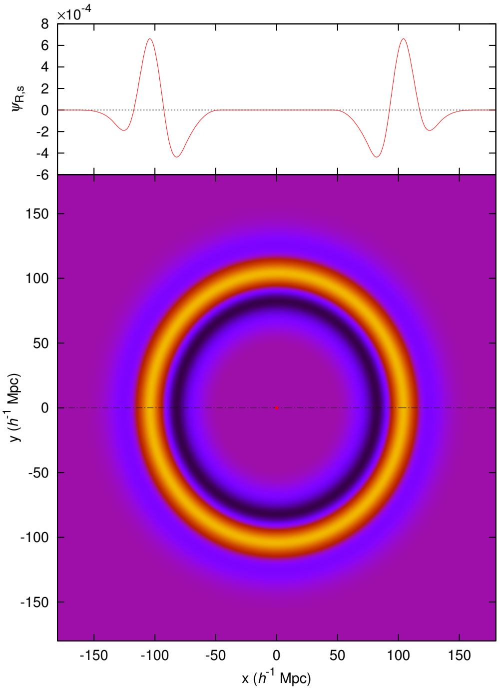


Figure 6.1: The BAOlet function used in our analysis, as defined in equation (6.1). Here we show a 2D plot (bottom) of the wavelet $\psi_{R,s}(\mathbf{x})$, together with a 1D slice (top) along the dashed-dotted axis. The wavelet is plotted here for $R = 105\,h^{-1}$ Mpc, $s = 30\,h^{-1}$ Mpc. The red dot marks the centre of the wavelet. This function has a null mean (provided that $R > 2s$), and compact support. It takes non-zero values only for $R - 2s \leq |\mathbf{x}| \leq R + 2s$.

Qualitatively, the value of $W_{R,s}$ at a given position $\mathbf{x}$ in space gives an indication of how the density distribution in the region surrounding $\mathbf{x}$ resembles the wavelet function for these values of the parameters (R, s). In this way, the BAOlet acts as a matched filter, which is sensitive to data containing shells of different radius and different widths. Its property of zero mean is also of high importance since it makes the statistics derived from the BAOlet coefficients independent of the background level. Indeed, it is obvious that any constant added to the input data would not change the BAOlet coefficients. In comparison, the estimation of such a baseline level is a very delicate aspect of the BAO detection in the two-point correlation function.

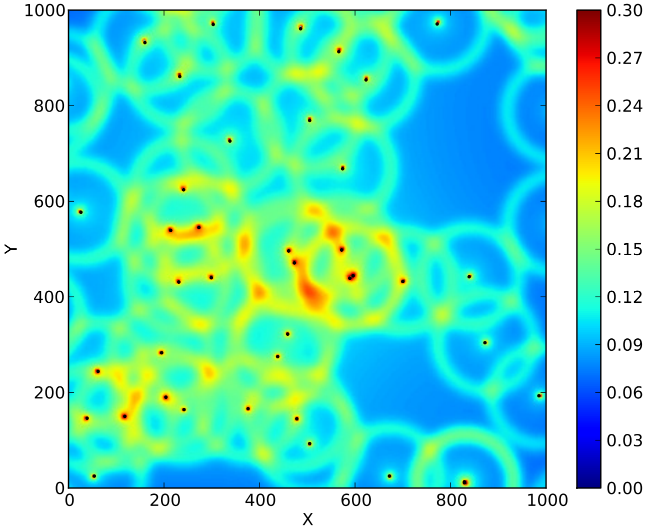


Figure 6.2: Density field generated for our 2D toy model. The centres of the shells generated are marked by the black points. Units are arbitrary.

6.1.2 Illustration of the method using a 2D toy model

In order to illustrate the meaning of the coefficient map $W_{R,s}(\mathbf{x})$, and of our method, we use simple un-physical model for the density distribution in two dimensions. We consider a 2D window of 1000×1000 (all units are arbitrary in this model). We generate randomly a set of 40 centres inside this region. We then generate an isotropic density field around each of those centres following a radial profile given by the addition of a power law plus a small Gaussian peak located at a radius of $r_0 = 100$, with width $\sigma = 20$. In this way, the profile is qualitatively similar to that expected from a large initial perturbation in the ΛCDM model (see e.g. Fig. 5.1). The total density field, shown in Fig. 6.2, is given by the superposition of the individual fields surrounding each of the centres.

We calculated the $W_{R,s}(\mathbf{x})$ coefficient map for this density field following equation (6.3)[1] for different values of the parameters (R, s). We show the results in Fig. 6.3, for two sets of values of (R, s), with the original positions of the shell centres superimposed. In the top panel, the values of R, s used for the calculation of the coefficients is close to the characteristics of the simulated shells. One can see by eye that in this case the original shell centres lie generally close to maxima of the $W_{R,s}$ map, as expected. In the bottom panel, however, the value of R used

---

[1] Note that in two dimensions the correct normalisation for the BAOlet function defined in equation (6.1) is slightly different, as we substitute the $4\pi r^2$ term in the denominator by $2\pi r$.

is much smaller than the radius of the shells. As there are no shells of this radius in the density field (except if they happen by chance), the $W_{R,s}$ map in this case should be due mostly to noise, and it is un-correlated with the positions of the centres.

Using these results, we can explain qualitatively the idea of our BAO detection method. We assume we know the positions $\{\mathbf{x}_c\}$ of the massive haloes which should be the centres of BAO shells. Using a different mass tracer, we build a map of the overall density field, and construct the BAOlet coefficient maps $W_{R,s}(\mathbf{x})$. If BAO are present in the sample, we should find a correlation between the centre positions $\{\mathbf{x}_c\}$, and the coefficient maps $W_{R,s}(\mathbf{x})$, for the values of (R, s) characteristics of the BAO shells. In the absence of BAO, however, we should not find any significant relation between the massive haloes and $W_{R,s}(\mathbf{x})$ for any value of (R, s). We measure this correlation between the possible centres and the shells by means of the $B(R, s)$ statistic defined below.

As a next step, once BAO are detected, we can use the map $W_{R,s}(\mathbf{x})$ for the characteristic parameters of the shell, as a map giving the intensity of the BAO shell signal for each position in space. Hence, we can use it to identify the actual structures in configuration space which are responsible for BAO.

6.1.3 The detection statistic B(R,s)

Due to the properties of the wavelet, the coefficient maps $W_{R,s}(\mathbf{x})$ should have a null mean when averaged over all points in the volume considered. Equivalently, if we sampled these maps at N random points uniformly distributed in the volume ($\mathbf{x}_r^{(i)}$), the expected value of the average of the coefficients is zero,

$$E\left\{\left\langle W_{R,s}(\mathbf{x}_r^{(i)})\right\rangle_N\right\} = 0\,. \tag{6.4}$$

This condition holds even in the presence of shell-like structures in the density field, and for any value of (R, s). However, if we are able to identify the positions of N massive haloes in the same volume ($\mathbf{x}_c^{(i)}$), we can define a new statistic $B(R, s)$ as the mean value of the coefficients $W_{R,s}(\mathbf{x})$ at these positions:

$$B(R,s) = \left\langle W_{R,s}(\mathbf{x}_c^{(i)})\right\rangle_N\,. \tag{6.5}$$

If there are indeed shell-like structures around the selected haloes $\mathbf{x}_c^{(i)}$, these positions should correspond, on average, to positive values of the coefficient map $W_{R,s}(\mathbf{x})$, as seen in Fig. 6.3. In this case, we should find positive values of $B(R, s)$ with the maximum of B at the (R, s) values characterising these shells.

We can obtain further information from the wavelet coefficients $W_{R,s}(\mathbf{x})$, as we have information on the actual dependence of the signal picked up by the BAOlet function on the position. In particular, fixing a set of parameters of

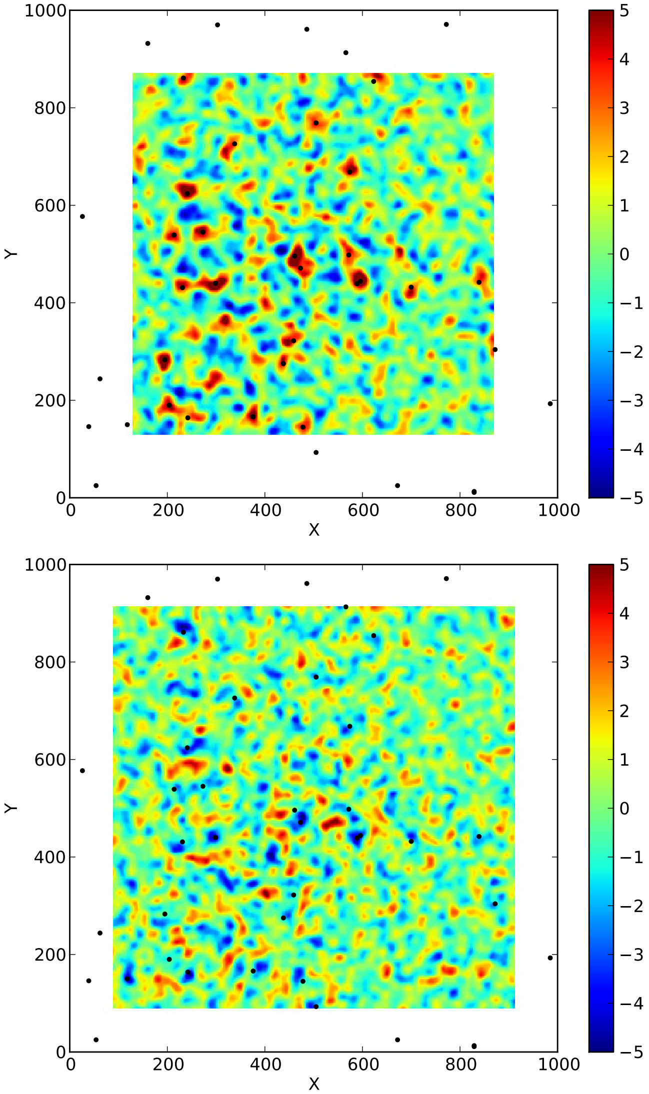


Figure 6.3: Coefficient map $W_{R,s}(\mathbf{x})$ for the 2D toy model density field shown in Fig. 6.2. We show the map for two sets of parameter values: $R = 105, s = 26.25$ (top), and $R = 65, s = 24.375$ (bottom). The black points represent the original centres used to generate the density field. We do not calculate $W_{R,s}(\mathbf{x})$ in the regions close to the boundaries (the white bands in the figures), in order to avoid possible border effects.

interest (R_i, s_i), we could use the coefficients $W_{R_i,s_i}(\mathbf{x}_c)$ to identify what is the signal that comes from each of the selected massive haloes for these characteristics of the shells. In the context of BAO, the parameters R_i, s_i can be chosen *a priori* using a theoretical model, or *a posteriori* using the parameters for which the function $B(R, s)$ attains its maximum. This spatial information can be used, for example, to localise in configuration space the structures responsible for the largest BAO signal in a given sample.

For our calculation of $B(R, s)$, we sample the (R, s) parameter space on a grid. For each point (R, s), we calculate the coefficient map $W_{R,s}(\mathbf{x})$ as the convolution of the BAOlet with the density field (equation 6.3). We perform the convolution in Fourier space using a Fast Fourier Transform (FFT) technique. To avoid problems with the FFT, we zero-pad a large region around our density cube. To obtain $B(R, s)$, we sample $W_{R,s}(\mathbf{x})$ at the position of the N selected centres, and calculate the average value (equation 6.5).

Therefore, to apply this method, we need a way to map the overall density field $\delta(\mathbf{x})$, but also to locate the position of massive matter haloes $\mathbf{x}_c^{(i)}$. We have to use two different populations of mass tracers, so that they play the appropriate role in the analysis. The idea of using two different tracer sets, one for the small perturbations and another for the high peaks, in a cross-correlation analysis was anticipated by Eisenstein et al. (2007). We implement here a similar idea, but using a wavelet tool directly on the density field. As detailed below, we use galaxies from the 'Main' and LRG samples of SDSS in this case. However, this choice would depend on the kind of data available in each case.

6.2 Prediction from ΛCDM

In order to better understand our method, we show here which results we expect according to the ΛCDM model, and the effect of BAO in our new statistic $B(R, s)$. We use for this aim both the analytical approximation to the transfer function of Eisenstein and Hu (1998), and the results from the MareNostrum Institut de Ciències de l'Espai (MICE) simulation (Fosalba et al., 2008).

In the first place, we use the ΛCDM transfer function, which allows us to study directly the effect of the BAO. However, in this case, we must do a series of approximations in order to make a prediction for $B(R, s)$. We want to predict which is the typical result for the wavelet coefficient $W_{R,s}$ at the position of the massive matter haloes $\mathbf{x}_c$, as a function of R, s. From equation (6.3), we see that this is equivalent to studying the typical density profile around such haloes, $\delta(\mathbf{y}-\mathbf{x}_c)$. The ΛCDM transfer function allows us to calculate this profile, provided we know which is the initial perturbation corresponding to the selected haloes. We follow Eisenstein et al. (2007), and make the simple approximation of considering that these initial perturbations are point-like and spherically symmetric, and can thus be simply described by a Dirac delta function in configuration space. This corresponds to a constant value in Fourier space. As the transfer function $T(k)$ describes the relative evolution of the different Fourier modes (see Section 1.1.2), the present day radial density profile corresponding to such initial perturbation will we given simply by

$$\rho(r) = C\widetilde{T}(r) \tag{6.6}$$

where $\widetilde{T}(r)$ is the Fourier transform of the transfer function $T(k)$, and C is a normalisation constant that depends on the details of the initial perturbation, and on the cosmic growth function $D_1(z)$. This is the same profile as shown in Fig. 5.1, although we consider here the transfer function for all the matter (dark matter plus baryons). From equations (6.3) and (6.5), we see that the effect of C will be just to change the overall normalisation of our statistic $B(R, s)$.

We used the fitting formulae to the linear-theory transfer function $T(k)$ from Eisenstein and Hu (1998), and obtained the expected $W_{R,s}$ at the position of a large overdensity using equations (6.6) and (6.3). In order to highlight the particular signature of BAO, we also calculated $W_{R,s}$ using the 'no wiggle' transfer function formula, in which the BAO signal has been edited out. We used here the values $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$, $\Omega_b = 0.044$, and $h = 0.7$ for the cosmological parameters, to allow for a direct comparison with the MICE simulation. Following Eisenstein and Hu (1998), the sound horizon scale in this case is $r_s = 109.3\,h^{-1}$ Mpc.

The results for both cases are shown in Fig. 6.4. In the plot, we mask the region $R < 2s$, as for these values of the parameters our BAOlet is not compensated (its mean is different from 0). Comparing both panels of the Figure, we see clearly which is the effect of the presence of BAO in our statistic. In the case without BAO $W_{R,s}$ is always negative, and it presents a smooth gradient across the (R, s) plane. This gradient is due to the overall shape of the radial profile (equation (6.6). However, in the presence of BAO, $W_{R,s}$ shows a prominent peak with positive values. This clearly shows the idea behind the $B(R, s)$ statistic. The BAOlet $\psi_{R,s}$ acts as matched filter with a shape adapted to detect spherical shells. Therefore the positive values in the coefficients $W_{R,s}$ correspond to the cases in which the radial profile is matched by the BAOlet shape. The values at which $W_{R,s}$ attains its absolute maximum, $R_{\max} = 110\,h^{-1}$ Mpc and $s_{\max} = 22\,h^{-1}$ Mpc, correspond thus to the characteristics of the shell that best matches the observed profile around the selected centres.

In order to test the reliability of the method, and of this ΛCDM prediction, we calculated the $B(R, s)$ for a halo catalogue drawn from the MICE simulations. We used the publicly available halo catalogue from the 'MICE3072' run (Crocce et al., 2010). This particular run contains 2048^3 particles in a box of side $3072\,h^{-1}$ Mpc, therefore covering a volume of $29\,h^{-3}\,\mathrm{Gpc}^3$. The simulation was run with the GADGET-2 code (Springel, 2005), assuming a ΛCDM model with the parameters mentioned above. Haloes in the simulation were selected using a friends-of-friends (FoF) algorithm.

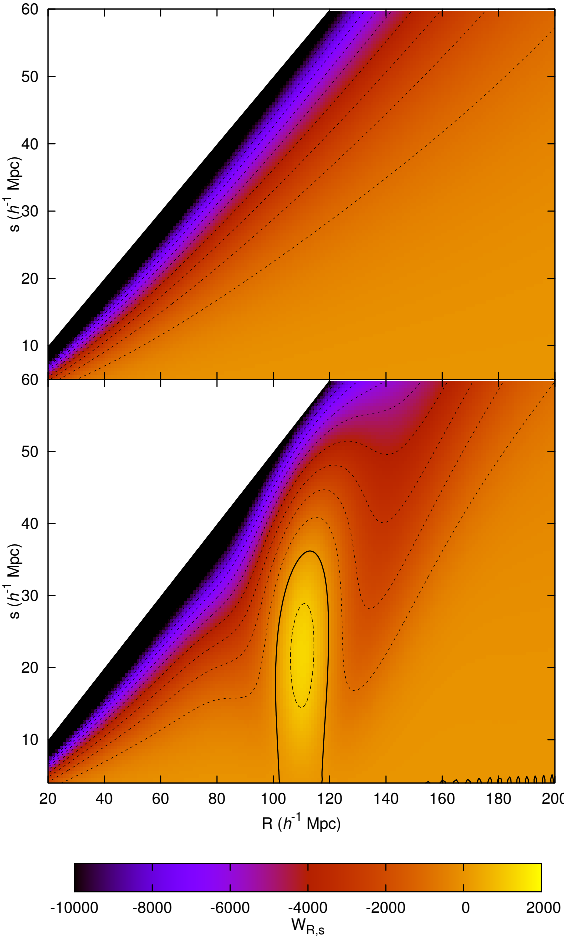


Figure 6.4: Values of the BAOlet coefficients $W_{R,s}$ expected at the positions of large initial point-like perturbations, as a function of the BAOlet parameters (R, s). The bottom panel shows the result using a standard linear-theory ΛCDM transfer function, while the top panel shows the result using a transfer function with the BAO wiggles edited out (Eisenstein and Hu, 1998). The normalisation is arbitrary. The contours are drawn at steps of 1000 for $W_{R,s} < 0$ (dotted), $W_{R,s} = 0$ (solid), and $W_{R,s} > 0$ (dashed). The map attains its maximum at $R = 110\,h^{-1}$ Mpc, $s = 22\,h^{-1}$ Mpc.

We used the halo catalogue at $z = 0$, which contains a total of 2819031 haloes formed by 143 or more particles. This corresponds to haloes with masses $\geq 3.35 \times 10^{13}\, h^{-1}\, \mathrm{M}_{\odot}$. The halo number density is thus $9.72 \times 10^{-5}\, h^3 \mathrm{Mpc}^{-3}$. We used the full halo catalogue as a tracer of the overall density field. We then selected as centres for the calculation of $B(R, s)$ in equation (6.5) only the haloes with a mass $\geq 1.76 \times 10^{14}\, h^{-1}\, \mathrm{M}_{\odot}$. We chose this mass threshold in order to select approximately the 10% most massive haloes in the simulation box. This choice is somewhat arbitrary, but serves for the purpose of testing the BAOlet method and illustrating the expected result. We considered only the real-space positions of the haloes, so the possible effects of redshift-space distortions are not included in this calculation.

Fig. 6.5 shows the BAOlet result $B(R, s)$ for these MICE samples, compared to the theoretical results obtained above from the Eisenstein and Hu (1998) transfer functions. We obtain a result very similar to that of Fig. 6.4, as $B(R, s)$ shows a clear peak, and attains its absolute maximum for $R_{\mathrm{max}} = 108\, h^{-1}\, \mathrm{Mpc}$, $s_{\mathrm{max}} = 28\, h^{-1}\, \mathrm{Mpc}$. This indicates that our BAOlet method can be applied to two sets of mass tracers, although the details of the tracers used here are very different from the ones we use later on the SDSS samples. This also confirms the expected effect of the presence of BAO in the $B(R, s)$ function: the presence of a large peak with positive values of B, located approximately at the values of R and s corresponding to the radius and width of the acoustic shells. The fact that we obtain here slightly different values for R_{max} and s_{max} than those predicted above may be due to non-linear evolution effects, which slightly reduce the radius and increase the width of the shells. As mentioned in Section 5.1.1, a similar effect is present in the correlation function (see e.g. Crocce and Scoccimarro, 2008).

6.3 SDSS samples used

We used data from two different samples of the latest data release (DR7) of the spectroscopic SDSS. On one side, we used the ‘Main’ galaxy sample (Strauss et al., 2002) as mass tracers for reconstructing the overall density field $\delta(\mathbf{x})$. On the other, we used the LRGs as tracers of the central over-densities, and therefore used them as the selected centres $\mathbf{x}_c^{(i)}$ to compute $B(R, s)$.

Luminous Red Galaxies were selected by the SDSS team using several colour and magnitude cuts to obtain a highly biased sample reaching high redshift (Eisenstein et al., 2001). The galaxies selected in this way are known to reside near the centres of massive dark matter haloes (Zheng et al., 2009) and are thus adequate tracers for the centres of baryon acoustic structures. We applied an extra cut in the K-corrected, evolved, g-band absolute magnitude: $-23.2 < M_g < -21.2$, as in the previous BAO analysis by Eisenstein et al. (2005), and as in our DR7-LRG sample in Chapter 5. This results in an approximately volume-limited sample in the redshift range $0.15 < z < 0.30$.

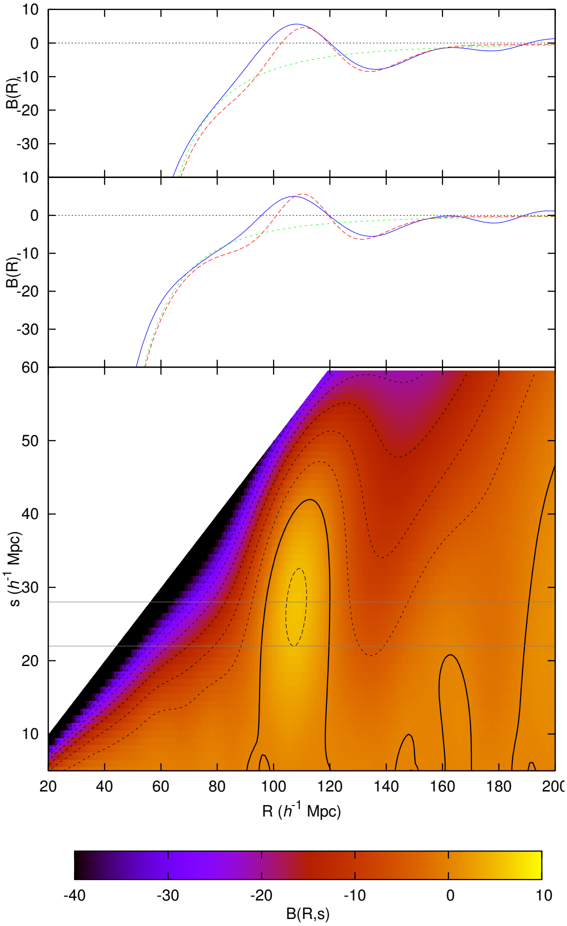


Figure 6.5: The BAOlet statistic B calculated for the MICE simulation sample described in the text as a function of the parameters (R, s) (bottom panel). The contours are drawn at steps of 5 for $B < 0$ (dotted), $B = 0$ (solid), and $B > 0$ (dashed). This function attains its maximum for $R = 108\,h^{-1}$ Mpc, $s = 28\,h^{-1}$ Mpc. The top two panels show cuts at the values $s = 28\,h^{-1}$ Mpc (top) and $s = 22\,h^{-1}$ Mpc (middle), marked with grey horizontal lines in the 2D panel. In each case, the solid blue line corresponds to the value obtained from MICE, the dashed red line corresponds to the theoretical expectation from the Eisenstein and Hu (1998) transfer function (bottom panel of Fig. 6.4), and the dotted green line to the theoretical expectation using the 'no wiggle' transfer function (top panel of Fig. 6.4). These theoretical prediction have been re-normalised to get the same value at the maximum in $B(R, s)$.

‘Main’ galaxies in the SDSS constitute a much denser sample, and are therefore more suitable to map small density changes such as BAO shells. We used the ‘Main’ sample from the Value-Added Galaxy Catalogue (Blanton et al., 2005), which constitutes a magnitude limited sample in the r band, with $r < 17.6$. We applied an extra simple cut, $M_r < -20$, in order to slightly reduce the changes in the radial selection function of this sample. However, this cut only affects a small volume at $z \lesssim 0.11$, eliminating the faintest galaxies in this region, and does not affect significantly our results.

For the conversion of angles and redshifts into co-moving distances, we used a fiducial cosmology with the parameters $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$. We converted the distribution of ‘Main’ galaxies into a density field $\delta(\mathbf{x})$ binning it into a grid with cubic pixels of $3\,h^{-1}$ Mpc side. We corrected for the selection effects by weighting each galaxy by the inverse of the average density at its redshift. Although this weighting may not be optimal, it should not affect significantly our results, given that the wavelet method does not depend on the local background level. We used the density field constructed in this way for the calculation of the BAOlet coefficients following equation (6.3).

In our calculations, we could only use the region in which these two samples overlap, which corresponds to the redshift limits $0.15 < z < 0.26$. To minimise border effects in the $W_{R,s}(\mathbf{x})$ and $B(R,s)$ calculations, we defined a buffer region of $r_{\mathrm{buff}} = 175\,h^{-1}$ Mpc from any of the borders of the ‘Main’ sample volume. We used as centres only the LRGs in the inner volume. This allows us to use the density field, as traced by the ‘Main’ sample galaxies, from $z > 0.09$. In order to minimise angular selection effects and border effects, we use a compact area of the sky where the angular completeness is nearly uniform. This area covers $5511\,\mathrm{deg}^2$ and is defined, in the SDSS survey coordinates (Stoughton et al., 2002), by the limits $-31.25° < \eta < 28.75°$, $-54.8° < \lambda < 51.8°$. This results in finally using the density field in a volume of $2.2 \times 10^8\,h^{-3}\,\mathrm{Mpc}^3$, as traced by $N_{\mathrm{Main}} = 198342$ galaxies. The number of LRGs used as centres (avoiding the buffer region) is $N_{\mathrm{LRG}} = 1599$.

In Fig. 6.6 we show a slice of the SDSS survey showing both the ‘Main’ galaxies and the LRGs. To illustrate the idea of the method, we show a zoom around a given LRG galaxy, for which the coefficient map $W_{R,s}(\mathbf{x})$ has a large value (for $R_{\max}, s_{\max}$). Even for this single centre, a slight over-density of ‘Main’ galaxies can be seen at a radius of $\simeq 105\,h^{-1}$ Mpc.

6.4 Results for the SDSS samples

We performed the calculation of $B(R,s)$ for the SDSS in an analogous way to the case of the MICE simulation, using the samples defined in Section 6.3. Our results are shown in Fig. 6.7. As above, we mask the region $R < 2s$. As we are not introducing any border correction when calculating the $B(R,s)$ statistic, we

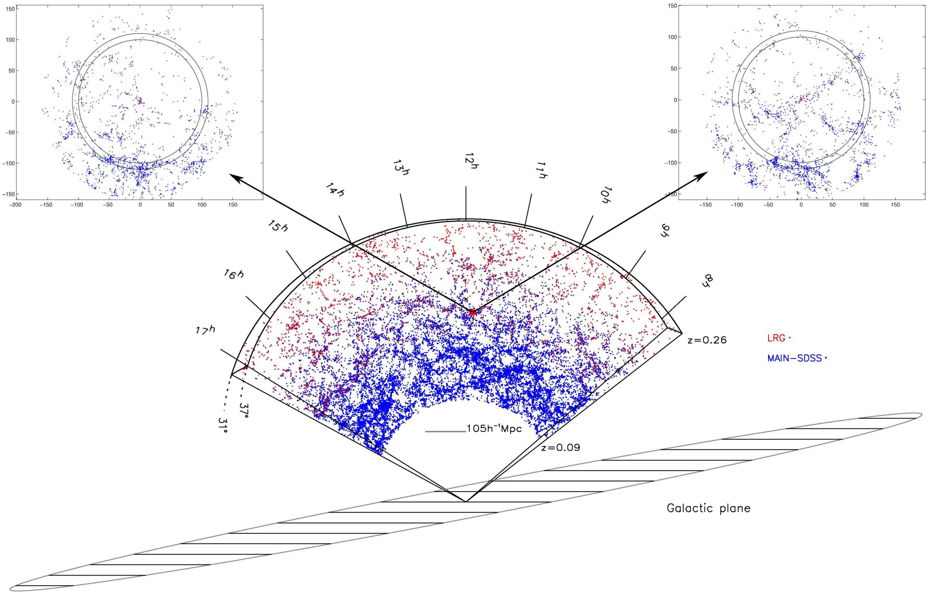


Figure 6.6: A slice drawn from the SDSS catalogues we used, showing the distribution of both the LRGs (red), and the 'Main galaxies (blue). Although we show here only a 6°-thick slice, we used the samples covering a total area of 5511 $\deg^2$ in our analysis. The radius of a typical BAO shell is shown as a segment. At the top insets we show two orthogonal slices of width $20\,h^{-1}$ Mpc, centred on a particular LRG. The two circles have the radii of 100 and 110 h^{-1} Mpc, and a slight over-density can be appreciated at that scale.

also mask the region corresponding to the values $R > r_{\rm buf} - s$. Values obtained at those large values of R could contain some spurious signal, as the calculation of $W_{R,s}$ would rely on the density field in regions outside of the survey boundaries.

The resulting $B(R, s)$ map is qualitatively very similar to that expected, either using an analytical ΛCDM model (Fig. 6.4), or the MICE simulation (Fig. 6.5). This is an indication that the observed pattern does not originate from spurious features in the SDSS but is closely related to the large scale structure and more specifically the BAO. $B(R, s)$ attains a maximum at $R_{\rm max} = 116\,h^{-1}$ Mpc, $s_{\rm max} = 36\,h^{-1}$ Mpc. This maximum is clearly related to the characteristics of the BAO structures present in our samples. However, the position of the maximum obtained slightly deviates from that obtained for the MICE simulation (Fig. 6.5), with larger values for both R and s. This shift may be due to different observational

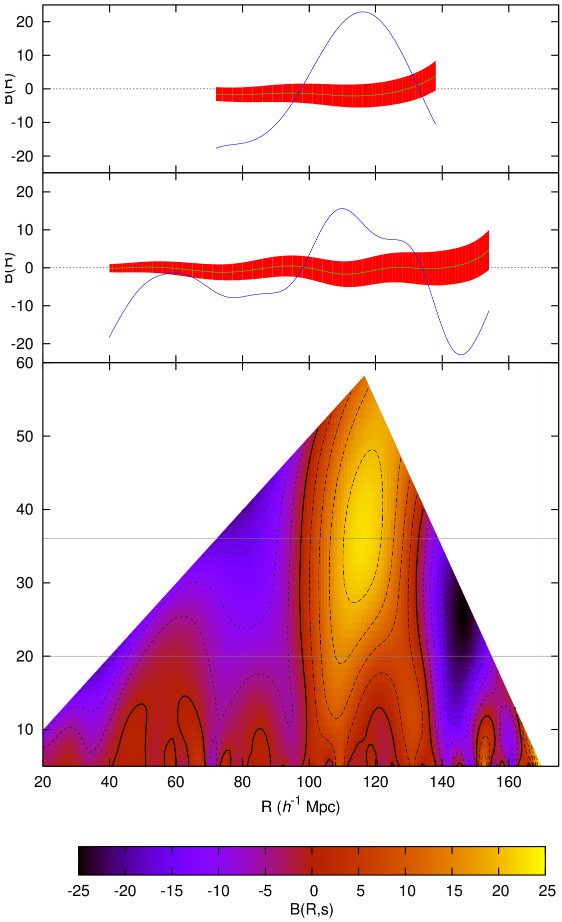


Figure 6.7: The BAOlet statistic B calculated for SDSS data as a function of the parameters (R, s). The bottom panel shows the results in the full parameter space considered, where we sampled both R and s at intervals of $1\,h^{-1}$ Mpc. We mask two areas, at the upper right and left corners, where our results are not reliable (see details in the text). The contours are drawn at steps of 5 for $B < 0$ (dotted), $B = 0$ (solid), and $B > 0$ (dashed). The top two panels show cuts at the arbitrarily chosen values $s = 36\,h^{-1}$ Mpc (top) and $s = 20\,h^{-1}$ Mpc (middle), marked with grey horizontal lines in the 2D panel. In these panels, the blue line is $B(R, s)$, while the green line and the red band show the mean $(\overline{B}^{MC})$ and 1-σ^{MC} interval for the Monte Carlo realisations of random centres. We obtain a clear significant peak at different values of s, with a maximum for $R = 116\,h^{-1}$ Mpc, $s = 36\,h^{-1}$ Mpc.

issues in the SDSS samples, such as redshift-space distortions or selection effects, which we did not model in the simulated catalogues. It may also be due to small differences between the simulated model and the actual large-scale matter distribution. A similar effect is also found in the detection of BAO using the two-point correlation function, where the observed peak is slightly wider than expected in the ΛCDM model (see Chapter 5, and Cabré and Gaztañaga, 2009; Kazin et al., 2010b; Labatie et al., 2011). In sum, we can not use the scale and the width of the observed maximum of $B(R,s)$ as direct estimates of the radius or width of the shells.

We performed a simplified significance test in order to assess the reliability of the BAO detection with this method. To do so, we focused on the value of $B(R,s)$ obtained at the maximum, $B_{\max} = B(R_{\max}, s_{\max}) = 22.9 \pm 3.7$[2]. We assess the probability of finding such a maximum in the case in which there are not baryon acoustic structures present in our sample. We model this null hypothesis by using randomly distributed centres for the calculation of $B(R,s)$ in equation (6.5), instead of the LRGs, while keeping the $W_{R,s}(\mathbf{x})$ coefficients from the observed density field (traced by SDSS 'Main' galaxies). The expected value of $B(R,s)$ in this case is 0 (see equation 6.4), and we expect to obtain a significantly higher signal in the data. In this way, we are testing the null hypothesis that, either there are not shell-like structures in the density field traced by the 'Main' sample, or these shell-like structures are not found preferentially around LRG centres. In either case, that would mean that there are not BAO-like structures present in our sample.

To perform the significance test, we generated 10^5 random realisations of a Poisson process, with the mean number of points $N_{\rm LRG}$, in the same volume as the LRGs considered in the calculation. We calculated $B(R,s)$ for each of such realisations. We obtain then the mean value $\overline{B}^{MC}(R,s)$, and the standard deviation $\sigma^{MC}(R,s)$ of the Monte-Carlo realisations of the centres. We show $\overline{B}^{MC}(R,s)$ and a band of $1\sigma^{MC}(R,s)$ around it in the top panels of Fig. 6.7.

We now calculate our signal-to-noise ratio at the maximum as $SNR_{\max} = B_{\max}/\left[\sigma^{MC}(R_{\max}, s_{\max})\right] = 6.60$, and assess the probability of finding such a large value of $SNR_{\max}$ anywhere in the parameter space for the Monte Carlo realisations. We computed the maximum value of SNR for each realisation j in the full (R,s) range, $SNR_{\max}^{MC(j)}$. The distribution of the values of $SNR_{\max}^{MC(j)}$ is shown in Fig. 6.8, where it is compared to the value of $SNR_{\max}$ obtained in the real data. We found that only one of the realisations gave a value of $SNR_{\max}^{MC(j)}$ larger than $SNR_{\max}$. Thus, the probability of obtaining a maximum with such a large SNR in the case of randomly distributed centres (our null hypothesis) is $p \simeq 10^{-5}$, equivalent to a $\sim 4.4\sigma$ detection in the Gaussian case.

We should stress here that the significance found here can not be compared

---

[2] This error in $B_{\max}$ is obtained from the variance of the coefficients at the $N_{\rm LRG}$ different LRGs. However, our significance test is independent of this error value.

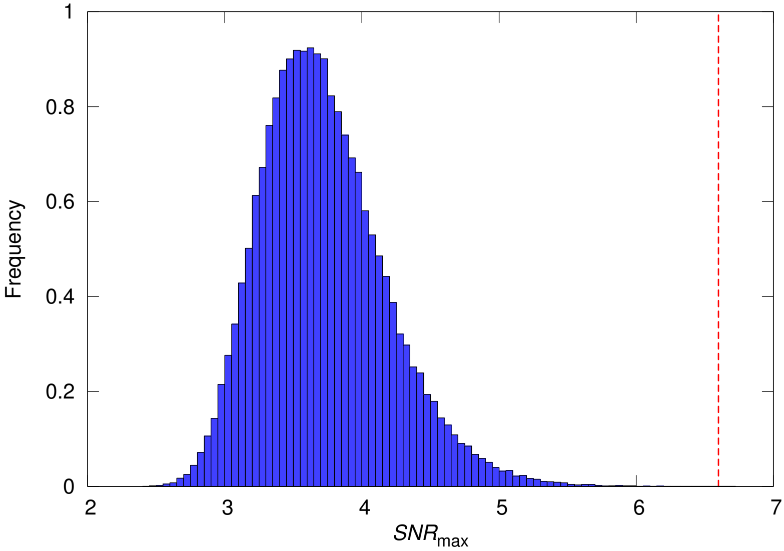


Figure 6.8: Histogram showing the distribution of the maximum SNR values obtained, in the full (R, s) space, for the 10^5 Monte Carlo realisations of Poisson-distributed centres ($SNR_{\rm max}^{MC(j)}$). This histogram has a mean of 3.72 and a standard deviation of 0.46. We show as a dashed vertical line the value obtained from the data (using the LRGs as centres), $SNR_{\rm max} = 6.60$. Only one of the Monte Carlo realisations give a maximum value larger than $SNR_{\rm max}$.

directly to other BAO detection levels found in the literature. We are not comparing our results with an analytical no-BAO model of $B(R, s)$ (such as that shown in the top panel of Fig. 6.4), since to do so would require the detailed modelling of all the selection effects affecting the two samples used. Even when dealing only with the value of $B_{\rm max}$, the null hypothesis we use above is missing some possible important effects (e.g. we do not account for the clustering of centres, as expected for LRGs), so that the 4.4σ detection level is probably over-estimated.

As explained in Section 6.1, we can extract more information about the BAO phenomenon in our samples making further use of the BAOlet coefficient maps $W_{R,s}(\mathbf{x})$. Once we have detected the BAO in our samples through the use of the $B(R, s)$ statistic, we can restrict to the map for the parameters $R_{\rm max}$, $s_{\rm max}$, which correspond to the characteristics of the BAO shells present in our samples. In this way, the values $W_{\rm max} \equiv W_{R_{\rm max}, s_{\rm max}}$ are a measure of how strong is the signal coming from a BAO shell around a given point, and in particular, a given LRG.

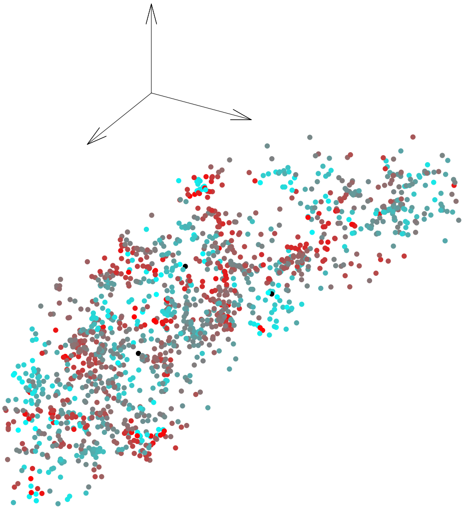


Figure 6.9: The distribution in redshift space of the LRGs used in the analysis (i.e. inside our buffer region). Each LRG is coloured according to the value of $W_{\rm max}$ at its position. Blue points correspond to low values of $W_{\rm max}$, and red points to high values.

Therefore, using $W_{\rm max}$ we can localise in configuration space the regions of the volume covered by our samples where the BAO signal is mostly coming from.

We illustrate this idea in Fig. 6.9, where we plot the distribution of the $N_{\rm LRG} = 1599$ LRGs used as centres in our analysis, showing also the value of $W_{\rm max}$ for each of them, following a colour gradient. The highest values of $W_{\rm max}$ correspond to the red points in the plot. We do not perform here a detailed analysis of the spatial information we get from $W_{R,s}$. However, from this figure, it appears that the LRGs giving larger signal (larger values of $W_{\rm max}$, in red) tend to be clustered together in high density regions. On the contrary, LRGs with low values of $W_{\rm max}$ (in blue) seem to be spread over the whole volume covered.

It is worth to emphasise that this approach would be impossible with any statistical BAO detection method used this far, since the spatial localisation of the shells is completely lost in the correlation function or in the power spectrum, while the local nature of the wavelet approach allows us to identify the positions of the most representative structures in our sample.

6.5 Discussion

In summary, we have designed a new method for the detection of baryon acoustic oscillations in the galaxy distribution and for the localisation, in configuration space, of the structures responsible for them. This method is based on the use of a specially designed wavelet (the 'BAOlet') applied directly on the density field. Our approach also relies on the use of two different tracers: one for the overall density distribution, and the other for the central overdensities of the baryon acoustic structures.

On a first step, we calculated the expected signal for our detection statistic $B(R, s)$, and tested the method with simulations. We then applied this method to the detection of baryon acoustic structures in a sample drawn from the SDSS. In this case, we used galaxies from the 'Main' catalogue to trace the overall density field, and galaxies from the LRG catalogue to trace the location of massive dark matter haloes. We clearly detect BAO in the sample providing a confirmation of the detection obtained previously using general two point statistics such as the power spectrum and correlation function (see Chapter 5). In fact, our approach provides an independent method for the detection.

Recent works have proposed alternative methods to study the BAO based on wavelets (Xu et al., 2010; Tian et al., 2011). In particular, Tian et al. use a Mexican hat wavelet function with two parameters, conceptually similar to ours. They use it to search for a peak in the two point correlation function of the 'Main' SDSS sample, obtaining a detection with a p-value $p = 0.002$ (equivalent to 3.1σ in the Gaussian case). As in our case, this shows the utility of using the 'Main' sample to reduce the shot noise in the calculation and to obtain significant detections. However, these works apply the wavelet to the measured two point correlation function, instead of directly to the density field. In this way, they use the wavelets to characterise accurately the BAO signal (in terms of radius and width), but they are not able to get any information about the localisation of these structures in space.

In our case, we are calculating the full coefficient maps $W_{R,s}$ directly from the density field, so we keep all the spatial information about the BAO signal. This kind of information is completely lost when using the standard two point statistics. We showed how this method allows us, through the use of $W_{\mathrm{max}}(\mathbf{x})$, to localise in configuration space the actual structures responsible for the BAO signal obtained.

Although we do not explore them here, this spatial information has several potential applications for a detailed analysis of the phenomenon of BAO. For example, it allows to study the relation of the BAO signal at a given LRG to its properties or the environment. It could also be used to make a selection of LRG centres with high signal, and use them to refine the measurements of the BAO characteristics.

We expect that this new method for studying BAO will be of much use for ongoing or planned surveys, such as the WiggleZ Survey (Drinkwater et al., 2010), the Baryon Oscillation Spectroscopic Survey (BOSS, Eisenstein et al., 2011), or the Physics of the Accelerating Universe (PAU) Survey (Benítez et al., 2009a), which will cover a much larger volume than studied here, and will explore higher redshifts.

In particular, the PAU Survey will cover 8000 deg^2, obtaining redshifts for $\sim$ 12 million LRGs in the range $0.1 < z < 0.9$, and hundreds of millions of other galaxies. This means a potential increase of a factor ~ 40 in volume, and of an order of magnitude in number density, with respect to the data used here. This implies that the detection and characterisation of the BAO using our statistic $B(R, s)$ would be much more reliable in this case. Moreover, the possibilities of our method to localise the BAO signal in configuration space could be used to study the evolution of the BAO characteristics over the large redshift range covered. However, the BAOlet method will need to be adapted to the characteristics of the PAU data, specially given the expected uncertainty in the photometric redshift determination, $\sigma_z \sim 0.003(1 + z)$.

A different application of our method was suggested by Beutler et al. (2011). They proposed to use the BAOlet method presented in this chapter to study BAO at very low redshifts, using the galaxies from the 6dF Galaxy Survey (6dFGS, Jones et al., 2004, 2009) as tracers of the central over-densities, and data from the future Wide field ASKAP L-band Legacy All-sky Blind surveY (WALLABY)[3] to map the overall density field. WALLABY is a H I survey planned for the Australian Square Kilometre Array Pathfinder telescope (ASKAP), currently under construction, which will contain half a million galaxies at a mean redshift ~ 0.04, having a large overlap with 6dFGS. Although the volume will not be very large in this case, the study of BAO at those small redshifts gives interesting cosmological information, as shown by Beutler et al. using the correlation function of 6dFGS alone.

---

[3] http://www.atnf.csiro.au/research/WALLABY

7

Conclusions

The work presented in this thesis has dealt with the analysis of data from recent galaxy redshift surveys, and with the development of new statistical tools to tackle specific problems. We divided the thesis in two parts, according to the ranges of scales considered in each case, and therefore to the scientific problems explored in each case. Conclusions have been drawn at the end of each chapter. Here we just summarise them for the whole thesis, indicating some further work to be done in the future. Most of the work presented in this thesis has been published, or submitted for publication, in refereed astronomical journals. Appendix A presents a list of these articles.

In Part I of the thesis, we studied the clustering of galaxies at small to intermediate scales ($r \lesssim 20\,h^{-1}\,\mathrm{Mpc}$), focusing on the evolution of this clustering, and on its dependence on galaxy properties, or *segregation*. Chapters 2 and 3 dealt directly with the problem of studying clustering at high redshift making use of the possibilities of the ALHAMBRA survey, while in Chapter 4 we made a more general review of the mark statistics methods that can be applied to the analysis of segregation.

In Chapter 2 we developed a method for the recovery of the real-space clustering from photometric surveys with the characteristics of ALHAMBRA. This method is based on the use of the projected correlation function, and is adapted to data with typical photometric redshift errors $\Delta z \lesssim 0.015(1+z)$. We showed, using data from an N-body simulation, that this method recovers the real-space correlation function to within 5% for the scales of interest. We also tested that the method is reliable in the presence of a small fraction of catastrophic redshifts, or outliers. A limitation of this method is the fact that it imposes a minimum redshift range to be considered for the calculation, which depends on Δz.

Although this method was designed in the first place with the aim of analysing ALHAMBRA data, it has a more general application, as it can be used when analysing data from any survey with similar redshift accuracy. This is the case of other photometric surveys using a large number of medium- or narrow-band filters, such as that of Ilbert et al. (2009) in the COSMOS field, or the planned J-PAS survey (Benítez et al., 2009a), which will cover $\sim 8000\,\mathrm{deg}^2$ in ~ 50 optical filters, thus obtaining a much better redshift accuracy.

We applied this method to the calculation of the correlation function for several samples drawn from the ALHAMBRA survey in Chapter 3. The characteristics of the ALHAMBRA survey allow us to explore a large range in redshift, $z \in [0.3, 1.5]$, so this is an ideal tool to study the evolution of galaxy clustering through cosmic time. We divided our sample in three redshift bins, and selected several galaxy samples in each of them based on B-band luminosity. Given the photometric depth, we considered in some cases samples of relatively faint galaxies, going down to $L = 0.1L^*$ for $z \simeq 0.5$, or $L = 0.5L^\star$ for $z \simeq 1.2$. This is not possible with similar pencil-beam spectroscopic surveys, as they need to target brighter sources for spectroscopy.

Although the ALHAMBRA data analysed is still preliminary, we already obtained some interesting results. In the range of scales studied, $r \in [0.2, 15]\,h^{-1}\,\mathrm{Mpc}$, the correlation function for all samples was well fitted by a power law, which makes the comparison between samples easier. We observed clearly the effects of evolution, as our high redshift ($z \simeq 1.2$) samples are much less clustered than samples at lower redshift. Regarding luminosity segregation, we saw clearly its effects for all the redshift ranges analysed. The fact of observing luminosity segregation for these faint samples with $L \lesssim L^*$ implies a difference with respect to local ($z \simeq 0$) studies, where this kind of segregation is significant only for $L \geq L^*$.

These results showed the possibilities of ALHAMBRA for the study of large-scale structure. The work done here can be extended naturally by the use a more detailed model of galaxy clustering to interpret the results obtained for the correlation function. Moreover, although we only studied here the dependence of clustering properties on luminosity, the ALHAMBRA data set will allow for a study of galaxy segregation in more detail, given the amount of information the survey obtains for each galaxy (colour, spectral type, morphology).

The phenomenon of galaxy segregation can be analysed in the framework of the statistics of marked point processes. We reviewed the basic tools available in this case in Chapter 4, and illustrated their application using a galaxy sample drawn from the 2dFGRS, characterised by a spectral classification parameter. We showed how these mark statistics tools can be used to study segregation as a function of scale both for continuous marks (such as luminosity, colour, etc.) and for discrete categories. In the latter case, we introduced the mark connection function, showing that it gives valuable information when analysing different galaxy populations defined by some set of galaxy characteristics. Our results for 2dFGRS confirmed previous observations, and showed in detail how passive galaxies tend to cluster more strongly than active ones at small scales.

The segregation analysis methods presented here are general tools for the analysis of data from any survey obtaining both the spatial distribution and characterisation of galaxies. In particular, they will be used for the detailed analysis of galaxy segregation using ALHAMBRA data. In this case, they will need to be adapted for their use with photometric redshift data, in a way similar as we did in Chapter 2 with the correlation function.

In Part II of the thesis, we turned to correlations at larger scales, focusing on the phenomenon of baryon acoustic oscillations (BAO). BAO are a feature imprinted in the cosmological density field in the early epochs of the universe, and provide a useful standard ruler to constrain the cosmological parameters. They were only recently detected (Eisenstein et al., 2005) in the galaxy distribution.

We reviewed the physics of BAO, and the status of their detection using the standard two-point statistics in Chapter 5. We measured the two-point correlation function for several samples drawn from the largest surveys to date, 2dFGRS and SDSS. We obtained a peak corresponding to BAO at the expected scale in all cases, which shows the reliability of the detection of this feature. Other authors have reinforced these conclusions using other independent samples. However, we saw a small deviation from the ΛCDM model prediction, as the observed peak was wider than expected. Although this deviation may not be significant, it deserves further investigation.

Finally, in Chapter 6 we developed a new method for the analysis of the BAO phenomenon. This method makes use of the possibilities of wavelets methods to look for the actual structures in configuration space which are responsible for the BAO. It is also based on the use of two complementary mass tracers, and we illustrated it using a catalogue formed by 'Main' and LRG galaxy samples from SDSS. In this way, we showed how we were able not only to detect BAO in the samples, but also to localise regions giving lower or higher BAO signal. This kind of information is completely lost when using the traditional two-point statistics methods.

This approach opens many new possibilities for the analysis of BAO, which we did not explore in detail in this thesis. Some of its possible applications could be to study the relation of these baryon acoustic structures with the properties of the galaxies that form them, or the selection of regions with high BAO signal to refine the characterisation of the BAO. Furthermore, this spatial information could be used to study the dependence of BAO characteristics (and in particular the BAO scale) on redshift in a continuous way. These possible applications are however limited by the survey volume available, specially due to the need for using two different tracers. In this sense, we expect that this method will be of much use for those ongoing and future surveys covering volumes much larger than that of SDSS.

Part III

Appendices

A

List of publications

Below, we list the articles published, or submitted for publication, in refereed journals, and containing some of the results presented in this thesis:

- P. Arnalte-Mur, A. Fernández-Soto, V. J. Martínez, E. Saar, P. Heinämäki, I. Suhhonenko (2009): Recovering the real-space correlation function from photometric redshift surveys.
  MNRAS, **394**:1631. `[arXiv:0812.4226]`
  (Contains the work presented in Chapter 2)

- V. J. Martínez, P. Arnalte-Mur, D. Stoyan (2010): Measuring galaxy segregation using the mark connection function.
  A&A, **513**:A22. `[arXiv:1001.1294]`
  (Contains the work presented in Chapter 4)

- V. J. Martínez, P. Arnalte-Mur, E. Saar, P. de la Cruz, M. J. Pons-Bordería, S. Paredes, A. Fernández-Soto, E. Tempel (2009): Reliability of the Detection of the Baryon Acoustic Peak.
  ApJL, **696**:L93–L97. Erratum: **703**:L184. `[arXiv:0812.2154]`
  (Contains the work presented in Sections 5.2–5.5)

- P. Arnalte-Mur, A. Labatie, N. Clerc, V. J. Martínez, J.-L. Starck, M. Lachièze-Rey, E. Saar, S. Paredes (2011): Wavelet analysis of baryon acoustic structures in the galaxy distribution.
  Submitted to *A&A*. Accepted, waiting for revision. `[arXiv:1101.1911]`
  (Contains the work presented in Chapter 6)

B

Agrupament de galàxies: evolució, segregació i oscil·lacions acústiques bariòniques

Aquest apèndix conté un resum del contingut de la tesi en català. En lloc d'incloure ací figures o taules noves, farem referència a les que apareixen al cos de la tesi en els capítols anteriors.

B.1 Introducció

La cosmologia és la ciència que té per objectiu comprendre l'origen, evolució i estructura de l'univers en el seu conjunt. Ha estat una part integral del desenvolupament de la ciència en la història, des del seu començament. En l'últim segle, diversos avanços teòrics i, principalment, observacionals l'han convertida en una ciència madura en el camp de la física, apropant-nos a comprendre les propietats bàsiques del nostre univers.

Una de les eines observacionals principals en aquest procés ha estat la construcció de grans cartografiats de galàxies cobrint fraccions significatives del cel. Mitjançant l'anàlisi de la distribució de les galàxies en l'espai, permeten l'estudi de l'estructura a gran escala (LSS, per les sigles en anglès) de l'univers. Aquests cartografiats també es centren en l'estudi dels processos de formació i evolució de les galàxies. En les tres últimes dècades, juntament amb altres eines cosmològiques, els cartografiats de galàxies han experimentat una gran millora en termes de cobertura, profunditat i precisió, gràcies a importants desenvolupaments tècnics. Al mateix temps, també s'han anat desenvolupant les eines teòriques i estadístiques necessàries per analitzar aquestes dades.

B.1.1 El model cosmològic estàndard

El model cosmològic estàndard, conegut com "Λ – Matèria fosca freda" (ΛCDM, per les sigles en anglès), s'ha anat desenvolupant al llarg del segle XX. La idea principal és la d'una "Gran explosió calenta" (o *Hot Big Bang*), basada en tres hipòtesis: la homogeneïtat i isotropia de l'univers a gran escala, la teoria de la relativitat general per descriure la dinàmica gravitatòria, i el fet que l'univers s'està expandint a partir d'un estat inicial molt dens i calent. Les principals evidències observacionals en favor d'aquest model són la llei de Hubble que relaciona la distància a les galàxies amb la seua velocitat de recessió, les observacions del fons còsmic de microones (CMB, per les seues sigles en anglès), i la mesura de les abundàncies d'elements lleugers.

La geometria de l'univers està determinada per les hipòtesis anteriors, i es descriu per la mètrica de Robertson-Walker (RW), que incorpora la homogeneïtat i isotropia, i descriu l'expansió de l'univers a partir del 'factor d'escala' $a(t)$. En aquest context, el 'desplaçament al roig' o *redshift* cosmològic z experimentat per un fotó que viatja des d'una galàxia llunyana fins nosaltres és una mesura del temps cosmològic t, o del factor d'escala a, corresponent a l'instant d'emissió del fotó. Seguint la relativitat general, podem relacionar les propietats geomètriques de la mètrica amb el contingut de matèria i energia de l'univers, emprant les anomenades equacions de Friedmann-Lemaître. La combinació de diferents resultats observacionals ens dóna una imatge detallada de l'univers. Aquest és espacialment pla, un $\sim 73\%$ del seu contingut està format per 'energia fosca', un $\sim 22\%$ per 'matèria fosca', i un $\sim 5\%$ per la matèria bariònica que coneguem. La matèria fosca seria un tipus de matèria que es comporta com a tal pel que fa a la gravitació, però no interacciona de cap altra manera. L'energia fosca, o constant cosmològica, és un terme que cal introduir per explicar el fet que l'expansió de l'univers està accelerant, i que podria ser assimilat a un terme d'energia del buit. L'explicació de la matèria i energia fosca és un dels temes de recerca actuals més importants en la física.

Estudiant la distribució de les galàxies en l'espai, s'observa que la matèria no es troba distribuïda uniformement, sinó que s'agrupa formant diverses estructures. Necessitem per tant estudiar el camp de pertorbacions en densitat $\delta(\mathbf{x}, t)$ i les seues propietats estadístiques. En particular, s'estudia la funció de correlació a dos punts de la distribució,

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle \ ,$$

i la seua transformada de Fourier, l'espectre de potències $P(k)$. En el context del model ΛCDM, les propietats del camp $\delta(\mathbf{x}, t)$ vénen determinades principalment per l'efecte de la gravetat sobre el camp inicial de pertorbacions. Aquestes pertorbacions, amb una amplitud inicial molt petita, creixen, principalment per acreció gravitatòria, per formar les estructures observades en l'actualitat. Per tant, tenint a més en compte altres efectes com per exemple la pressió deguda a la radiació, es pot fer una predicció acurada per $\xi(r)$ i $P(k)$ en aquest model (Fig. 1.6), i comparar-la amb l'estudi observacional de l'estructura a gran escala.

B.1.2 Anàlisi estadística de l'agrupament de galàxies

La manera més directa d'estudiar l'estructura a gran escala de l'univers és mitjançant l'anàlisi de la distribució de galàxies en l'espai. Per aquesta anàlisi, considerem la distribució de galàxies com una realització d'un procés puntual aleatori, i hi emprem els mètodes estadístics associats. En aquest cas, assumint homogeneïtat i isotropia, definim la funció de correlació a dos punts $\xi(r)$ de la següent manera. Considerem dues esferes de volums infinitesimals $\mathrm{d}V_1$ i $\mathrm{d}V_2$, separades per una distància r. Aleshores, la probabilitat què en ambdues esferes hi trobem un punt del procés puntual és

$$\mathrm{d}P_{12} = \lambda^2 \left[1 + \xi(r)\right] \mathrm{d}V_1 \mathrm{d}V_2 \,,$$

on λ és la densitat numèrica del procés.

Dins del model ΛCDM, el camp de densitat de galàxies està relacionat amb el camp de densitat de matèria, en l'aproximació més simple, a través del paràmetre de biaix, o *bias*, b, com

$$\delta_g(\mathbf{x}) = b\delta(\mathbf{x}) \,,$$

i per tant podem relacionar les funcions de correlació corresponents. Aquest paràmetre b conté informació sobre els processos de formació i evolució de les galàxies i varia en funció del tipus de galàxia que considerem. Açò implica que les propietats d'agrupament de poblacions diferents de galàxies són diferents, un fenomen conegut com 'segregació'.

Existeixen diversos mètodes per l'estimació de $\xi(r)$ a partir de les dades obtingudes per un cartografiat, és a dir, a partir d'una realització del procés puntual en un volum (o finestra) determinat. Tots aquests mètodes es basen en el comptatge de parells de punts i en una certa correcció pels efectes provocats per les vores del volum. Els mètodes més emprats en cosmologia es basen en la utilització d'un catàleg auxiliar artificial sense agrupament (és a dir, un procés de Poisson). Aquest és el cas de l'estimador de Landy i Szalay (1993) que emprem en aquesta tesi.

B.1.3 Cartografiats de redshifts de galàxies

Els cartografiats de galàxies són una de les eines principals de la cosmologia observacional. L'objectiu bàsic d'aquests és acumular informació observacional sobre totes les galàxies, complint un determinat criteri de selecció, en una regió del cel. D'aquesta manera, es poden estudiar estadísticament les propietats de les diferents poblacions de galàxies, i la seua distribució en l'espai. Inicialment, aquests cartografiats contenien només la posició angular de les galàxies. Tanmateix, a partir dels anys 1980, començant amb el cartografiat del *Center for Astrophysics* (CfA, Huchra et al., 1983; de Lapparent et al., 1986), apareixen cartografiats que mesuren també l'espectre de cada galàxia, de manera que es pot calcular el seu *redshift* i per tant la seua distància.

En parts d'aquesta tesi emprem dades provinents dels dos majors cartografiats completats fins ara: el *Two-degree Field Galaxy Redshift Survey* (2dFGRS, Colless et al., 2001) i el *Sloan Digital Sky Survey* (SDSS, York et al., 2000). El 2dFGRS va obtenir espectres per 221414 galàxies cobrint una àrea de $\sim 1500\,\mathrm{deg}^2$. La mostra espectroscòpica es va seleccionar en magnitud aparent com $b_J \leq 19.45$, i el *redshift* medià de les galàxies seleccionades és $z_m = 0.11$. Pel que fa al SDSS, aquest cobreix una àrea total de $\sim 8000\,\mathrm{deg}^2$, i considera dues mostres diferents de galàxies. La mostra principal, o '*Main*', es selecciona per un tall $r \leq 17.77$. Conté ~ 560000 galàxies, amb un *redshift* medià de $z_m \simeq 0.1$. Una segona mostra selecciona galàxies roges lluminoses (LRGs, per les sigles en anglès), mitjançant una selecció més complexa en color i magnitud. Aquesta mostra és més profunda, arribant fins $z \sim 0.5$, i conté ~ 200000 galàxies.

Una alternativa a la mesura espectroscòpica dels *redshifts* és l'ús de *redshifts* fotomètrics a partir de dades fotomètriques en diverses filtres (Baum, 1962; Koo, 1985). En aquest cas, la precisió en la mesura del *redshift* és molt menor, però l'ús de la fotometria permet obtenir mostres més profundes, o cobrint regions del cel majors, emprant menors temps d'observació. El cartografiat ALHAMBRA (*Advanced Large Homogeneous Area Medium-Band Redshift Astronomical survey*, Moles et al., 2008) es troba a meitat camí entre els cartografiats fotomètrics clàssics i els espectroscòpics, ja que obté fotometria emprant un total de 23 filtres, reduint així l'error comés en la mesura dels *redshifts* fotomètrics. Presentem amb detall les característiques d'aquest cartografiat al Capítol 3 [B.2.2].

B.1.4 Objectius d'aquesta tesi

En aquesta tesi, estudiem diferents temes relacionats amb l'anàlisi de l'estructura a gran escala de l'univers basada en l'agrupament de galàxies. Ens centrem en l'anàlisi de dades provinents de recents cartografiats de *redshifts* de galàxies. Al mateix temps, desenvolupem una sèrie de mètodes estadístics nous dirigits a abordar alguns dels problemes oberts en l'estudi de l'agrupament de galàxies.

La tesi està dividida en dues parts. En la Part I [B.2] ens centrem en l'agrupament de galàxies a escales petites i intermèdies. En aquestes escales, la distribució de galàxies depèn tant de la distribució global de la matèria a gran escala, com dels processos que governen la formació i evolució de galàxies, a través del *bias*. Per tant, és essencial entendre com l'agrupament de galàxies es troba afectat per les propietats intrínseques de les galàxies en les mostres considerades (segregació). També és important estudiar l'evolució amb el temps de les propietats de l'agrupament, intentant separar l'evolució de la LSS de l'evolució de les propietats de les galàxies. Ací ens centrem en els possibles estudis que es poden dur a terme emprant el cartografiat ALHAMBRA.

En primer lloc, al Capítol 2 [B.2.1], desenvolupem un mètode per recuperar la funció de correlació en espai real a partir de *redshifts* fotomètrics, centrant-nos en la precisió típica esperada per ALHAMBRA. Posem a prova aquest mètode emprant catàlegs extrets de la simulació d'halos tipus *pencil beam* de Heinämäki et al. (2005). A continuació apliquem aquest mètode a dades preliminars del cartografiat ALHAMBRA al Capítol 3 [B.2.2]. Hi incloem una descripció de les característiques del cartografiat. Calculem la funció de correlació per diverses mostres de galàxies a diferents *redshifts* (fins $z = 1.5$) i lluminositats. Açò ens permet estudiar la segregació per lluminositat de les galàxies, i la seua evolució, durant un període que s'estén durant el $\sim 70\%$ de la vida de l'univers. Finalment, al Capítol 4 [B.2.3], examinem una sèrie de mètodes estadístics que permeten una anàlisi més detallada de la segregació de galàxies. Aquests mètodes es basen en el formalisme de l'estadística de marques, utilitzat de manera habitual en altres camps. Ací, il·lustrem els diferents mètodes emprant dades del 2dFGRS, i introduïm per primera vegada la funció de connexió de marques per al seu ús en cosmologia.

En la Part II [B.3], estudiem escales majors, i ens centrem en particular en les oscil·lacions acústiques bariòniques (BAO, per les seues sigles en anglès) en la distribució de galàxies. Les BAO són una dèbil característica gravada en la distribució de galàxies per les ones acústiques presents en el plasma de barions i fotons abans de la recombinació. Per tant, ens proporcionen una connexió entre la LSS que observem avui i la física de les primeres etapes de l'història de l'univers. També ens proporcionen una regla estàndard que es pot emprar per estudiar en detall la història de l'expansió de l'univers.

Les BAO només s'han pogut detectar en la distribució de galàxies recentment (Eisenstein et al., 2005; Cole et al., 2005), emprant els majors cartografiats de galàxies disponibles. Al Capítol 5 [B.3.1] repassem la física de les BAO, i a continuació ens centrem en estudiar la fiabilitat de la seua detecció amb les dades actuals. Amb aquest fi, calculem la funció de correlació a grans escales per diverses mostres de galàxies extretes de SDSS i 2dFGRS. Finalment, al Capítol 6 [B.3.2] desenvolupem un nou mètode, complementari als anteriors, per la detecció de les BAO. Aquest mètode es basa en l'ús d'ondetes (*wavelets*), i té com objectiu detectar les estructures responsables del fenomen de les BAO. A més a més, gràcies a l'enfocament de les ondetes, aquest mètode proporciona més informació que els estadístics a dos punts habituals, ja que conserva la informació referent a la localització del senyal de les BAO en l'espai de configuració. En aquell capítol presentem el mètode, i l'apliquem a un catàleg de galàxies extret de SDSS.

B.2 Evolució de l'agrupament i segregació de galàxies a escales petites

B.2.1 Recuperació de la funció de correlació en espai real a partir de cartografiats de redshifts fotomètrics

Als últims anys, s'han proposat i dut a terme una sèrie de cartografiats cosmològics basats en la utilització de fotometria multibanda per a la determinació dels *redshifts* de les galàxies. El principal avantatge d'aquests cartografiats front a aquells basats en l'espectroscòpia és la seua major profunditat i densitat d'objectes que poden assolir. A canvi, la incertesa en la mesura dels *redshifts* és molt major, i aquesta es trasllada a una incertesa en la mesura de les posicions tridimensionals de les galàxies.

L'objectiu d'aquest capítol és estudiar un mètode que ens permeta recuperar el valor de la funció de correlació a dos punts, $\xi(r)$, a partir de les dades provinents de cartografiats fotomètrics, tenint per tant en compte la presència d'errors grans en els *redshifts*. Aquest mètode empra la funció de correlació projectada, introduïda per Davis and Peebles (1983) per al cas dels cartografiats espectroscòpics. Donat que la nostra intenció és la utilització d'aquest mètode per estudiar l'agrupament de galàxies emprant dades d'ALHAMBRA, estudiem el cas d'un cartografiat fotomètric estret i profund (de tipus *pencil beam*). Pel que fa al valor de la incertesa en *redshift*, aquest depèn principalment del nombre de filtres emprats. Ací estudiem diferents casos corresponents a les propietats de diferents cartografiats realitzats o en projecte.

El treball presentat en aquest capítol va ser publicat com a Arnalte-Mur et al. (2009).

Mètode per la recuperació de $\xi(r)$

El mètode que emprem ací per recuperar la funció de correlació en l'espai real es basa en la descomposició de les distàncies entre parells de galàxies en una component paral·lela ($r_\parallel$) i una altra perpendicular ($r_\perp$) a la direcció de la visual (Fig. 2.1). Una vegada mesurem la funció de correlació ξ en termes d'aquestes coordenades, podem calcular la funció de correlació projectada com

$$w(r_\perp) \equiv 2 \int_0^\infty \xi(r_\perp, r_\parallel) \mathrm{d}r_\parallel \,. \tag{B.1}$$

Donat que w depèn únicament de la separació transversal, no es veu afectada (a primer ordre) pels errors en *redshift*. Així, es pot emprar per recuperar la funció de correlació en espai real com

$$\xi_r(r) = -\frac{1}{\pi} \int_r^\infty \frac{\mathrm{d}w(r_\perp)}{\mathrm{d}r_\perp} \frac{\mathrm{d}r_\perp}{\left(r_\perp^2 - r^2\right)^{1/2}} \,. \tag{B.2}$$

Catàlegs fotomètrics simulats

Per tal de comprovar la validesa d'aquest mètode, emprem una sèrie de catàlegs fotomètrics simulats basats en la simulació de con de llum de Heinämäki et al. (2005). Aquesta simulació correspon a una àrea en el cel de $2^\circ \times 0\overset{\circ}{.}5$, i en aquest treball emprem el catàleg d'halos de matèria fosca en el rang en *redshift* $z \in [2, 3]$.

A partir de les posicions reals dels halos en la simulació, generem els catàlegs fotomètrics artificials, pertorbant aleatòriament les coordenades radials seguint una distribució gaussiana, amb desviació estàndard corresponent a la incertesa en *redshift* considerada en cada cas. D'aquesta manera generem tres catàlegs diferents, amb els valors següents de l'error en *redshift*: $\Delta z/(1 + z) = 0.005, 0.015, 0.05$. Aquests valors es corresponen, aproximadament, als casos de cartografiats fotomètrics amb ~ 5 filtres, amb ~ 23 filtres (com és el cas d'ALHAMBRA), o amb $\sim 30 - 50$ filtres (com és el cas dels cartografiats en el camp COSMOS, Ilbert et al., 2009, o el cartografiat PAU, Benítez et al., 2009a).

Proves del mètode emprant els catàlegs simulats

A l'hora de calcular la funció de correlació en espai real per als nostres catàlegs simulats, comencem per calcular la funció de correlació en dues dimensions, $\xi(r_\perp, r_\parallel)$ mitjançant l'estimador de Landy i Szalay (1993). A continuació, calculem la funció de correlació projectada $w(r_\perp)$ integrant numèricament l'equació (B.1). És necessari en aquest cas fixar un límit d'integració superior finit, $r_{\parallel,\mathrm{max}}$, que influeix directament sobre el resultat del nostre càlcul. Vam realitzar una sèrie de proves emprant els nostres catàlegs simulats, arribant a la conclusió que un valor raonable per aquesta límit és $r_{\parallel,\mathrm{max}} = 4r(\Delta z)$, on $r(\Delta z)$ és la distància comòbil corresponent al valor típic de l'error en *redshift*, Δz, en cada cas. Finalment, obtenim la funció de correlació deprojectada $\xi_{\mathrm{dep}}(r)$ emprant l'equació (B.2).

D'aquesta manera, apliquem el mètode per la recuperació de la funció de correlació en espai real als tres catàlegs fotomètrics simulats, i els comparem amb la funció de correlació obtinguda directament sense introduir-hi els errors en *redshift*, pel rang d'escales $0.5 < r < 30\,h^{-1}$ Mpc (Fig. 2.7). Aquests resultats mostren que, per als dos catàlegs simulats amb millor qualitat de la mesura dels *redshifts*, $\Delta z/(1 + z) = 0.005, 0.015$, som capaços de recuperar la funció de correlació amb una precisió d'un $\sim 5\%$ per $r < 10\,h^{-1}$ Mpc, i entre un 10% i un 20% per les escales majors. Tanmateix, per al cas amb errors en *redshift* majors, $\Delta z = 0.05(1+z)$, el mètode només ens permet recuperar la funció de correlació buscada per les escales més petites, $r < 2\,h^{-1}$ Mpc, però falla a escales majors. Açò és degut, en part, al fet que per valors tan grans de Δz, el valor de $r_{\parallel,\mathrm{max}}$ que necessitem emprar per la integració de l'equació (B.1) és comparable a la dimensió radial del volum cobert pel catàleg.

Finalment, realitzem dues proves addicionals del mètode. Per un costat, estudiem la robustesa del mètode de deprojecció en presència d'un 5% d'objectes amb *redshifts* 'catastròfics', és a dir, amb un error molt major que l'assumit per al conjunt de la mostra. Aquest tipus d'errors poden aparèixer en la determinació de *redshifts* fotomètrics, per exemple degut a confusions entre diferents característiques espectrals. En aquest cas, el nostre mètode també permet recuperar la funció de correlació en espai real (Fig. 2.9), tot i que ara la precisió és lleugerament menor. D'altra banda, repetim el procés emprant un catàleg de punts generat mitjançant un model de segments de Cox (Martínez i Saar, 2002), ja que en aquest cas la funció de correlació té una forma analítica. El fet que en aquest cas també siguem capaços, almenys per als catàlegs amb errors en *redshift* menors, de recuperar la funció de correlació (Fig. 2.10) ens indica que els resultats obtinguts no depenen en gran mesura dels detalls del procés puntual estudiat.

Concloem, per tant, que el mètode de la funció de correlació projectada presentat en aquest capítol és un mètode fiable per estudiar les correlacions en espai real emprant catàlegs de *redshifts* fotomètrics. En el següent capítol, emprem aquest mètode per estudiar la funció de correlació en espai real emprant mostres de galàxies provinents del cartografiat ALHAMBRA.

B.2.2 Evolució de l'agrupament de galàxies des de $z = 1.5$ en el cartografiat ALHAMBRA

L'estudi de la distribució de galàxies a alt *redshift*, comparant-la amb els resultats obtinguts per mostres locals, és una eina útil per l'estudi de l'estructura a gran escala i la seua evolució, així com del procés de formació i evolució de les galàxies. En concret, la funció de correlació $\xi(r)$ és una eina estadística útil per aquesta anàlisi. Per mostres locals, $\xi(r)$ es comporta aproximadament com una llei de potències per un gran rang d'escales (Totsuji and Kihara, 1969; Peebles, 1974; Martínez, 1999). Tanmateix, recentment s'han observat desviacions significatives respecte a aquest comportament, que són compatibles amb el model dels halos (Zehavi et al., 2004).

Una qüestió important en aquest estudi és entendre com evoluciona la dependència del *bias* amb la lluminositat, és a dir, el fet que les galàxies més lluminoses mostren un agrupament més fort que aquelles més dèbils. Aquest efecte ha estat estudiat amb detall en mostres locals, amb $z \simeq 0$. Alguns cartografiats espectroscòpics profunds han estès aquest estudi a *redshifts* majors, fins $z \sim 1$.

En aquest capítol estudiem l'agrupament de galàxies, i la dependència d'aquest amb la lluminositat, emprant dades provisionals del cartografiat fotomètric ALHAMBRA. Aquest cartografiat ens permet realitzar aquest estudi per *redshifts* fins $z = 1.5$. A més a més, donada la profunditat del cartografiat, podem estudiar mostres de galàxies més dèbils que allò possible emprant espectroscòpia. Per aquest estudi, emprem la funció de correlació $\xi(r)$, calculada mitjançant el mètode presentat en el capítol anterior, per tenir en compte els errors presents en la mesura dels *redshifts*.

El cartografiat ALHAMBRA

El cartografiat ALHAMBRA (Moles et al., 2008) és un cartografiat fotomètric que cobreix un àrea total de $4\,\mathrm{deg}^2$, emprant un total de 23 filtres en l'òptic i l'infraroig proper (NIR). S'ha dut a terme emprant el telescopi de 3.5 metres del *Centro Astronómico Hispano-Alemán* (CAHA) a Calar Alto, Almería. El seu objectiu principal és produir un catàleg cobrint un gran rang en *redshift*, per tal d'estudiar l'evolució còsmica.

El sistema de filtres va ser dissenyat per tal d'optimitzar la mesura de *redshifts* fotomètrics (Benítez et al., 2009b). L'espectre òptic (entre 3500 i 9700 Å) es cobreix amb 20 filtres de banda mitjana (Fig. 3.1), mentre que al NIR s'empren els tres filtres J, H, K_s estàndard. L'àrea total del cartografiat es divideix en 8 camps separats en el cel. Cadascun d'aquests està format per dues tires de $1^\circ \times 15'$, separades per $\sim 15'$, degut a la geometria de la càmera òptica emprada.

El catàleg emprat en aquest treball és el catàleg IDR3, que conté dades provisionals per una àrea nominal de $2.44\,\mathrm{deg}^2$. El nombre total d'objectes amb *redshift* mesurat en aquest catàleg és de 501868. Els objectes d'aquest catàleg es van detectar emprant una imatge profunda construïda en cada camp com a combinació de les exposicions de millor qualitat, en diferents filtres. La profunditat assolida és de $AB \simeq 24.5$ per als filtres òptics, i $AB \simeq 22$ per als filtres en el NIR. Els *redshifts* fotomètrics es van mesurar emprant el mètode implementat en *Bayesian Photometric Redshift* (BPZ, Benítez, 2000), basat en l'ajust a 6 models d'espectres de galàxies fent ús del formalisme bayesià.

Selecció d'un catàleg per l'estudi de l'estructura a gran escala

Per tal de poder estudiar amb precisió l'estructura a gran escala utilitzant aquest catàleg, però, cal una selecció adequada dels objectes. D'una banda, per evitar introduir en els càlculs objectes espuris (deguts a soroll, defectes de la imatge, etc.), estels, o galàxies amb *redshifts* mesurats amb baixa qualitat. De l'altra banda, necessitem caracteritzar amb precisió la funció de selecció del cartografiat, per tal de poder estudiar estadístics com la funció de correlació.

En primer lloc, estudiem la funció de selecció angular del cartografiat. La descrivim emprant una màscara angular que defineix quines regions del cel s'han observat amb fiabilitat, però no entrem a estudiar variacions en la completesa dins d'aquesta màscara. Ens basem en les imatges profundes en cada camp emprades per la detecció d'objectes. D'una banda, eliminem les regions de la imatge amb temps d'exposició total baix, corresponents als límits de la imatge, i que contenen una gran quantitat de soroll. A més a més, també eliminem una sèrie de regions al voltant d'objectes extensos o saturats, o amb píxels negatius, que podrien afectar la correcta detecció d'objectes propers. Les figures 3.4 i 3.5 il·lustren el procés de creació d'aquestes màscares. L'àrea efectiva seleccionada finalment és de $1.877\,\mathrm{deg}^2$ (un 77% de l'àrea nominal total per IDR3), i tenim 376175 objectes dins de les màscares.

A continuació, emprem un tall en el diagrama de color (A457M - J) – (J - K_s) per tal d'eliminar un $\sim 23\%$ d'estels presents en la mostra (Fig. 3.7). Aquest mètode és similar al mètode BzK (Daddi et al., 2004) emprat habitualment.

Finalment, fem una selecció en termes de la qualitat en la determinació dels *redshifts* dels objectes. Per aquest fi, emprem el paràmetre $p_{\rm odds}$, proporcionat per BPZ, que ens dóna una mesura de com de concentrada és la distribució de probabilitat de z al voltant del valor més probable. Per tal de minimitzar la presència d'objectes amb errors catastròfics en el *redshift*, i restringir-nos a aquells amb els menors errors, seleccionem objectes amb valors $p_{\rm odds} > 0.85$.

El catàleg que utilitzem finalment conté un total de 106713 galàxies. La distribució d'objectes en *redshift* (Fig. 3.8) es concentra en la regió $0.3 \lesssim z \lesssim 1.5$, amb una mediana de $z_m = 0.810$, encara que hi ha objectes també a *redshifts* més alts, arribant fins $z_{\rm max} = 5.38$. L'error relatiu mitjà en la determinació dels *redshifts* del catàleg és $\overline{\left(\frac{\sigma_z}{1+z}\right)} = 0.012$.

Per cadascun d'aquests objectes, calculem la distància comòbil corresponent al seu *redshift* (emprant els paràmetres cosmològics determinats per WMAP7), i la seua magnitud absoluta en banda B, M_B. En aquest cas, apliquem la correcció K (Hogg et al., 2002) corresponent, utilitzant els mateixos models d'espectres emprats en la determinació dels *redshifts* fotomètrics.

Funcions de correlació per als catàlegs d'ALHAMBRA

Per tal d'estudiar la funció de correlació per poblacions de galàxies a diferents *redshifts* i lluminositats, dividim el catàleg seleccionat en un total de 12 mostres. Pel que fa al *redshift*, emprem tres intervals, corresponents a $z \in [0.3, 0.6]$, $z \in [0.6, 1.0]$ i $z \in [1.0, 1.5]$. Respecte a la lluminositat, creem diferents mostres per cada interval en *redshift* imposant un límit inferior variable sobre la lluminositat en banda B. D'aquesta manera seleccionem 5 mostres diferents per l'interval més proper en *redshift*, 4 mostres per l'intermedi, i 3 pel cas més llunyà. Estudiant la lluminositat mediana per cada mostra i comparant-la amb la lluminositat característica de la funció de lluminositat L^*, es verifica en tots els casos que $L^{\rm med} \lesssim L^*$. Els detalls de les mostres emprades es troben en la Taula 3.3, i la selecció realitzada s'il·lustra en la Fig. 3.14.

Per cadascuna d'aquestes mostres, calculem la funció de correlació projectada $w(r_\perp)$, i la funció de correlació en espai real deprojectada $\xi_{\rm dep}(r)$, fent ús del mètode presentat al Capítol 2 [B.2.1]. Per tal de fer el càlcul amb l'estimador de Landy i Szalay (1993), necessitem modelar amb detall la funció de selecció de cada mostra. Pel que fa a la funció de selecció angular, emprem la màscara definida a l'apartat anterior. La funció de selecció radial per cadascuna de les mostres la calculem directament a partir d'un suavitzat del perfil de densitat radial observat. Estimem els errors i la matriu de covariàncies pels nostres resultats de $w(r_\perp)$ i

$\xi_{\rm dep}(r)$ en cada cas emprant el mètode de *jackknife* (e.g. Norberg et al., 2009).

Seguint el procés descrit, calculem $w(r_\perp)$ per les nostres mostres en el rang aproximat $r_\perp \in [0.02, 30]\,h^{-1}\,\mathrm{Mpc}$ (Fig. 3.16). Qualitativament, s'observa que $w(r_\perp)$ es comporta com una llei de potències, encara que amb algunes desviacions. A més a més, comparant els resultats per mostres amb diferents lluminositats, s'observa clarament l'efecte de segregació, ja que les mostres de major lluminositat mostren valors majors de $w(r_\perp)$. Els resultats que obtenim per $\xi_{\rm dep}(r)$ són qualitativament semblants. Tanmateix, els errors són majors en aquest cas, i és per això que realitzarem la comparació amb models directament emprant els resultats per $w(r_\perp)$.

Comparem els nostres resultats per $w(r_\perp)$ amb dos models relativament simples. D'una banda, considerem un model de llei de potències per $\xi(r)$,

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

que es transforma en una llei de potències per $w(r_\perp)$, amb la següent forma:

$$w(r_\perp) = r_\perp \left(\frac{r_0}{r_\perp}\right)^{\gamma} \frac{\Gamma(1/2)\Gamma\left[(\gamma-1)/2\right]}{\Gamma(\gamma/2)} ,$$

on $\Gamma(\cdot)$ és la funció gamma. De l'altra banda, considerem un model basat en el la teoria estàndard de ΛCDM, emprant el programari Camb (Lewis et al., 2000). D'aquesta manera, calculem la funció de correlació projectada esperada per la distribució de matèria, $w_m(r_\perp)$, una vegada fixats els paràmetres cosmològics, i només variem el paràmetre de *bias* de la mostra de galàxies,

$$w(r_\perp) = b^2 w_m(r_\perp) .$$

Fem l'ajust a ambdós models, emprant el mètode estàndard de minimització de χ^2, per cadascuna de les mostres considerades. En tots els casos, restringim l'ajust al rang d'escales $r_\perp \in [0.2, 15]\,h^{-1}\,\mathrm{Mpc}$. Els paràmetres obtinguts per l'ajust en cada cas es troben a la Taula 3.4. No són evidents diferències significatives entre ambdós models pel que fa a la bondat de l'ajust.

Estudiant els valors de r_0 obtinguts en l'ajust al model de potències (amb γ fixat) per les diferents mostres (Fig. 3.22), obtenim una imatge clara de les propietats d'agrupament per les diferents poblacions de galàxies. D'una banda, veiem una clara dependència de r_0 amb la magnitud absoluta M_B de les mostres, un senyal de la segregació per lluminositat. De l'altra, observem com l'agrupament és més fort per les mostres a *redshift* més baix, com s'espera pel procés de formació d'estructures.

Finalment, emprant els ajustos al model basat en ΛCDM, podem quantificar la dependència del *bias* b en la lluminositat L de les mostres (Fig. 3.23). Obtenim una clara variació de b per tot el rang de lluminositats emprat. Aquest comportament és diferent a aquell observat en cartografiats locals, on la segregació per lluminositat només és significativa per les mostres més lluminoses, amb $L > L^*$ (Hamilton, 1988; Norberg et al., 2001; Tegmark et al., 2004). Aquesta diferència per mostres a alt *redshift* ja havia estat observada per Coil et al. (2006) emprant el cartografiat DEEP2, encara que per un rang en lluminositats més limitat que l'estudiat ací.

En definitiva, en aquest capítol mostrem que el mètode de la funció de correlació projectada és útil per a l'estudi de l'estructura a gran escala emprant *redshifts* fotomètrics en el cas d'un cartografiat com ALHAMBRA. A més a més, tot i que els resultats són encara preliminars, demostrem les capacitats d'ALHAMBRA per estudiar l'evolució de l'agrupament de les galàxies, i la segregació d'aquestes. El catàleg final d'ALHAMBRA ens permetrà realitzar una anàlisi més detallada, fent ús de models d'agrupament més complets, i estenent l'estudi de la segregació a altres propietats de les galàxies.

B.2.3 Mesures de la segregació de galàxies amb estadística de marques

És un fet conegut que diferents tipus de galàxies presenten diferents propietats pel que fa a l'agrupament. Aquest fenomen s'anomena segregació. Les propietats d'agrupament depenen de diverses característiques intrínseques de les galàxies, com la lluminositat, el color, el tipus espectral o la morfologia. A grans trets, podem identificar dos tipus de galàxies. Les galàxies de tipus el·líptic, amb color més roig i ritme de formació estel·lar baix, es troben preferentment en entorns amb densitat alta, mentre que les de tipus espiral, amb colors blaus i ritme de formació estel·lar alt, són les dominants en el camp (Davis and Geller, 1976; Dressler, 1980). Aquestes diferències es corresponen amb les diferents amplituds i pendents observades en la funció de correlació per aquestes dues poblacions (Loveday et al., 1995; Hermit et al., 1996; Guzzo et al., 1997). A més del tipus de galàxia, com hem explicat al capítol anterior, l'agrupament també depèn de la lluminositat de les galàxies (e.g. Hamilton, 1988).

En aquest capítol, investiguem un enfocament del fenomen de la segregació de galàxies basat en considerar la distribució de galàxies com una realització d'un procés puntual marcat (Stoyan i Stoyan, 1994; Illian et al., 2008). En aquest cas, cada punt del procés té associada una característica o 'marca', i existeixen mètodes estadístics dissenyats per estudiar conjuntament les propietats de la distribució dels punts en l'espai i de la distribució de les marques. L'objectiu del capítol és revisar alguns d'aquests mètodes, i mostrar la seua utilitat per l'anàlisi de la distribució de galàxies. Per tal d'il·lustrar l'aplicació dels diferents mètodes, emprem una mostra de galàxies limitada en volum extreta del cartografiat 2dFGRS.

El treball presentat en aquest capítol va ser publicat com a Martínez et al. (2010).

Mètodes d'anàlisi de l'agrupament per processos puntuals marcats

Considerem per tant la distribució de galàxies com una realització d'un procés puntual marcat X^M. Aquesta és una seqüència de punts marcats aleatoris $X = (\mathbf{x}_i, m_i)$, on $\mathbf{x}_i$ és la posició d'un punt, i m_i és la marca associada a aquest. La marca pot ser o bé una variable contínua (per exemple alguna propietat com la lluminositat o el color), o bé una variable discreta. En aquest darrer cas, tenim típicament una descripció qualitativa de diferents classes d'objectes.

En aquest treball tractem una sèrie d'estadístics de segon ordre que són una extensió de la funció de correlació a dos punts per al cas dels processos marcats. En el cas en què tractem les marques com una variable contínua, definim la funció de correlació de marques normalitzada $k_{mm}(r)$, i el variograma de marques $\gamma(r)$. En ambdós casos, podem considerar aquests estadístics com una mena de probabilitat condicionada a trobar parells de punts del procés separats per una distància r.

Així, $k_{mm}(r)$ mesura el valor esperat del producte normalitzat de les marques per aquests parells. Si la distribució de marques és independent de l'agrupament espacial dels punts tindrem $k_{mm}(r) = 1$. D'altra banda, $\gamma(r)$ és una mesura de la semblança entre les marques per parells separats per una distància r. En el cas en què les marques són independents de la distribució espacial, tenim $\gamma(r) = \sigma_m^2$, on σ_m^2 és la variància de la distribució de les marques.

Per al cas en què dividim la mostra en poblacions discretes, considerem ací les funcions de correlació parcials $\xi_{ij}(r)$, i les funcions de connexió de marques $p_{ij}(r)$. Les primeres es defineixen de manera anàloga a la funció de correlació usual $\xi(r)$, a partir de la probabilitat de trobar en un cert element de volum $\mathrm{d}V_1$ un punt de la població i, i en l'element de volum $\mathrm{d}V_2$ un punt de la població j, sent r la distància que separa els dos elements de volum.

Per la seua banda, les funcions $p_{ij}(r)$ són semblants als estadístics $k_{mm}(r)$ i $\gamma(r)$ pel seu sentit condicional. $p_{ij}(r)$ representa la probabilitat de trobar dos punts dels tipus i i j separats per una distància r sota la condició que en eixes posicions hi ha punts del procés global.

Mostra de dades emprada

Per il·lustrar l'aplicació d'aquests estadístics a la distribució de galàxies, hem seleccionat una mostra limitada en volum extreta del cartografiat 2dFGRS (Croton et al., 2004), i seleccionada en magnitud absoluta com $-20 < M_{b_J} < -19$. Per tal de simplificar els càlculs, ens limitem a una regió de l'espai amb forma de paral·lelepípede, cobrint un volum de $V \simeq 10^6\, h^{-3}\,\mathrm{Mpc}^3$, i contenint un total de $N = 7741$ galàxies.

Assignem marques a les galàxies emprant el paràmetre η definit per Madgwick et al. (2002). Aquest classifica els espectres de les galàxies, de manera que valors baixos de η corresponen a galàxies amb poca formació estel·lar (passives), mentre que valors alts corresponen amb galàxies amb alta formació estel·lar (actives). Donat que la distribució del paràmetre η en les dades és clarament bimodal (Fig. 4.4), per a l'estudi amb marques discretes dividim la mostra en dues poblacions amb un nombre semblant de galàxies, emprant un llindar adequat en el valor de η.

Resultats: segregació en el 2dFGRS

Calculem els diferents estadístics presentats abans per aquesta mostra de galàxies. En cada cas, realitzem el càlcul basant-nos en l'estimador de Stoyan i Stoyan (1994) per la funció de correlació, que inclou explícitament la correcció pels efectes de les vores del volum. Comparem els nostres resultats amb la hipòtesi nul·la què la distribució de les marques és independent de la posició espacial de les galàxies (el 'model de marques aleatòries').

Calculem en primer lloc els estadístics corresponents a marques contínues, $k_{mm}(r)$ (Fig. 4.6) i $\gamma(r)$ (Fig. 4.7). Per escales curtes, $r \lesssim 15\,h^{-1}\,\mathrm{Mpc}$, obtenim signes clars de la segregació de galàxies, ja que en aquest cas $k_{mm}(r) < 1$ i $\gamma(r) < \sigma_m^2$. Açò ens indica que, a escales curtes, els parells de punts que es troben a la mostra estan formats preferentment per dues galàxies amb valors baixos de η, com esperàvem.

Pel que fa a la nostra divisió en dues classes de galàxies, calculem en primer lloc les funcions de correlació parcials $\xi_{ij}(r)$ (Fig. 4.8). Aquestes mostren l'alt grau d'agrupament present en ambdues poblacions. Tanmateix, és en estudiar la funció de connexió de marques $p_{ij}(r)$ (Fig. 4.9) quan es veu clarament l'efecte de la segregació, i és possible estudiar a quines escales hi apareix. Per escales $r \gtrsim 20\,h^{-1}\,\mathrm{Mpc}$ no trobem signes de segregació. Per escales més curtes observem com el nombre relatiu de parells format per dues galàxies passives augmenta, mentre que aquells formats per dues galàxies actives disminueix. Pel que fa als parells creuats, aquest efecte és present només a les escales més curtes, $r \lesssim 10\,h^{-1}\,\mathrm{Mpc}$.

En definitiva, mostrem com els estadístics associats als processos puntuals marcats ens poden proporcionar informació útil sobre la segregació de galàxies. Açò ocorre tant per al cas de les marques contínues, com per al cas d'una classificació discreta de galàxies. En aquest darrer cas, hem introduït la funció de connexió de marques $p_{ij}(r)$ per primera vegada en cosmologia, mostrant la seua utilitat.

B.3 Agrupament a grans escales i oscil·lacions acústiques bariòniques

B.3.1 Fiabilitat de la detecció del pic acústic bariònic en la funció de correlació

En aquest capítol, fem en primer lloc una introducció a la física de les oscil·lacions acústiques bariòniques (BAO, per les seues sigles en anglès), i la seua importància per la determinació de paràmetres. A continuació, estudiem la fiabilitat de la seua detecció en la funció de correlació, emprant mostres extretes dels cartografiats 2dFGRS i SDSS. Part del treball presentat en aquest capítol va ser publicat com Martínez et al. (2009).

Oscil·lacions acústiques bariòniques

Les BAO són una característica de l'estructura a gran escala deguda a les ones acústiques presents en el plasma de fotons i barions abans de la recombinació. Donada una sobredensitat adiabàtica inicial, la pressió de radiació fa que es forme una ona esfèrica viatjant cap a l'exterior, que està formada tant pels fotons com pels barions, que en eixe moment es troben acoblats (Fig. 5.1). Al mateix temps, la sobredensitat inicial creix degut a l'acreció gravitatòria de la matèria fosca que no es veu afectada per la pressió de radiació. En el moment de la recombinació ($z \sim 1100$), els barions es desacoblen de la radiació, de manera que formen una closca esfèrica al voltant de la pertorbació inicial, amb un radi corresponent a l'horitzó acústic en el moment de la recombinació, r_s. A partir d'ací, tant la sobredensitat central com la closca esfèrica augmenten en amplitud degut a l'acreció de matèria fosca i de barions. Pel que fa al radi de la closca esfèrica, aquest ja només augmenta seguint el flux de Hubble, de manera que es manté constant en coordenades comòbils.

La distribució de matèria a baix *redshift* manté signes d'aquestes característiques gravades en les èpoques més primerenques. En el cas de la funció de correlació $\xi(r)$, hi apareix un pic a l'escala corresponent al radi de les closques esfèriques, $r_s \simeq 105\,h^{-1}$ Mpc en coordenades comòbils. Aquest pic es transforma en una sèrie d'oscil·lacions atenuades presents en l'espectre de potències $P(k)$ (Fig. 5.3).

El fet que l'escala típica de les BAO, r_s, es mantinga fixa (en coordenades comòbils) des del moment de la recombinació fins ara fa que aquestes es puguen emprar com una mena de 'regla estàndard'. El valor de r_s està determinat acuradament a partir de mesures del CMB. Així, en detectar BAO en la distribució de galàxies a un cert *redshift*, comparant el valor de $r_{\rm BAO}$ mesurat amb el valor de r_s conegut a partir del CMB, obtenim informació sobre el ritme d'expansió $H(z)$, i per tant sobre els paràmetres cosmològics.

Les BAO van ser detectades per primera vegada en la mostra de LRGs del

SDSS per Eisenstein et al. (2005), i en el 2dFGRS per Cole et al. (2005). Treballs posteriors han confirmat aquesta detecció, i han emprat les BAO per la determinació de paràmetres cosmològics (e.g. Tegmark et al., 2006; Percival et al., 2010). En aquest capítol, estudiem la fiabilitat de la detecció del pic bariònic en la funció de correlació, seleccionant diferents mostres del catàleg de 2dFGRS i del catàleg final (DR7) de SDSS.

Mostres de dades del SDSS i 2dFGRS emprades

Realitzem els càlculs per tres mostres diferents, extretes dels cartografiats SDSS i 2dFGRS. En primer lloc, emprem el catàleg de LRGs de SDSS. D'una banda, definim una mostra equivalent a l'emprada per Eisenstein et al. (2005), seleccionant pràcticament totes les LRGs en l'interval $z \in [0.16, 0.47]$, amb un volum total de $V = 1.30\,h^{-3}\,\mathrm{Gpc}^3$. De l'altra banda, fem un tall més exigent en magnitud absoluta, per obtenir una mostra de les galàxies més lluminoses, que està limitada en volum en l'interval $z \in [0.16, 0.40]$. Finalment, emprem també una mostra limitada en volum extreta del 2dFGRS (Croton et al., 2004), cobrint un interval en *redshift* menor, $z \in [0.03, 0.19]$ i per tant amb un volum molt menor. Les característiques de les diferents mostres es troben en la Taula 5.1.

Funcions de correlació per les mostres de SDSS i 2dFGRS

Calculem la funció de correlació a escales grans per cadascuna de les mostres emprant l'estimador de Landy i Szalay (1993), i tenint en compte les diferents funcions de selecció angulars i radials de les mostres. Pel que fa a l'error en la funció de correlació, emprem un nou mètode d'estimació basat en el *bootstrap* (Barrow et al., 1984) aplicat al comptatge de parells de punts.

Els nostres resultats per les tres mostres estudiades (Figs. 5.7 i 5.8) confirmen la presència del pic bariònic en la funció de correlació, i són bàsicament compatibles amb la predicció del model ΛCDM. Tanmateix, s'observa que l'amplada del pic és major que la esperada, i que l'observada en l'estudi original de Eisenstein et al. (2005). Contràriament a la predicció del model, la funció de correlació per la mostra de 2dFGRS té un mínim amb $\xi(r) < 0$ per escales $r \simeq 60\,h^{-1}\,\mathrm{Mpc}$. Aquesta darrera característica pot ser deguda a la variància mostral, donat el petit volum cobert en aquest cas.

En definitiva, comprovem que el pic bariònic és una característica ben determinada de la funció de correlació a grans escales, ja que s'observa en mostres amb propietats molt diferents. En particular, el detectem per primera vegada en el catàleg final (DR7) de SDSS, i en una mostra del 2dFGRS. Estudis posteriors han confirmat aquests resultats, incloent la discrepància en l'amplada del pic, tant en el SDSS com en dades provinents d'altres cartografiats.

B.3.2 Anàlisi en ondetes de les estructures acústiques bariòniques

En aquest capítol, presentem un mètode alternatiu per la detecció i l'anàlisi de les BAO en la distribució de galàxies, cercant les estructures en l'espai de configuració que es corresponen amb aquest fenomen. El mètode està basat en analitzar directament la distribució de galàxies emprant una transformada d'ondeta, o *wavelet*[1], dissenyada especialment per estudiar les característiques de les BAO. Per aquest fi, emprem dos traçadors diferents, un pel camp global de densitat (incloses les closques esfèriques degudes a les BAO), i l'altre per localitzar grans sobredensitats, que han de correspondre als centres d'aquestes closques esfèriques. A més de detectar les BAO emprant un mètode complementari als estadístics a dos punts, aquest mètode ens permet obtenir informació addicional sobre la distribució del senyal de les BAO en diferents regions de l'espai. El treball presentat en aquest capítol ha estat acceptat per la seua publicació (pendent de revisió) com Arnalte-Mur et al. (2011).

El mètode de detecció mitjançant ondetes

La base del nostre mètode és centrar-nos en la posició dels halos més massius, i estudiar el camp de densitat al seu voltant, cercant estructures semblants a les closques esfèriques de les BAO. L'anàlisi en ondetes és un mètode adequat per a la identificació d'estructures en camps continus, i ha estat emprat en diferents àrees, incloent-hi l'anàlisi de l'estructura a gran escala i el CMB en cosmologia.

En el nostre cas, el tipus particular d'estructura que busquem (closques esfèriques) fa necessari dissenyar una família d'ondetes específica. Aquestes funcions $\psi_{R,s}(\mathbf{x})$, que anomenem 'BAOlets', són funcions amb la forma aproximada d'una closca esfèrica de radi R i amplada característica s (Fig. 6.1). A més a més, tenen suport compacte i mitjana nul·la, propietats útils per al tipus d'anàlisi que realitzem. Una vegada definida la família d'ondetes, podem calcular la transformada corresponent per al camp de sobredensitats $\delta(\mathbf{x})$,

$$W_{R,s}(\mathbf{x}) = \int_{\Re^3} \psi_{R,s}(\mathbf{y})\delta(\mathbf{y}-\mathbf{x})\mathrm{d}^3\mathbf{y}\,.$$

El valor del coeficient $W_{R,s}(\mathbf{x})$ en un cert punt ens dóna una mesura de com s'assembla el camp de densitat $\delta(\mathbf{x})$ al voltant d'eixe punt a la forma de la funció d'ondeta, per a eixos valors dels paràmetres R, s.

Donades les propietats de les ondetes, el valor mitjà dels coeficients $W_{R,s}(\mathbf{x})$ sobre tot el volum ha de ser nul. Tanmateix, si identifiquem les posicions $\mathbf{x}_c^{(i)}$ de

---

[1] Hi ha diferents traduccions possibles al català d'aquest terme, com 'oneta', 'ondícula' o 'ondeleta' (a partir de l'original francès '*ondelette*'). Ací emprarem el terme 'ondeta', recomanat pel centre de terminologia Termcat.

N halos massius, podem definir l'estadístic

$$B(R,s) = \left\langle W_{R,s}(\mathbf{x}_c^{(i)}) \right\rangle_N ,$$

que tindrà valors positius si tenim estructures en forma de closca esfèrica preferentment centrades en les posicions $\mathbf{x}_c^{(i)}$. A més, trobarem el valor màxim de B pels valors de (R,s) característics d'aquestes closques esfèriques.

Predicció del model ΛCDM

En primer lloc, estudiem els resultats per $B(R,s)$ esperats en el model ΛCDM estàndard. Per aquest fi, emprem d'una banda l'aproximació analítica a la funció de transferència d'Eisenstein i Hu (1998), i de l'altra el catàleg d'halos obtingut de la simulació de N-cossos MareNostrum Institut de Ciències de l'Espai (MICE, Fosalba et al., 2008). Els resultats per $B(R,s)$ obtinguts (Figs. 6.4 i 6.5) mostren com l'efecte esperat de les BAO és un pic en el pla (R,s). La posició del màxim d'aquest pic caracteritza l'escala i amplada dels BAO pels valors dels paràmetres cosmològics emprats ($R = 108\,h^{-1}\,\mathrm{Mpc}$, $s = 28\,h^{-1}\,\mathrm{Mpc}$ en el cas de MICE, que incorpora els efectes de la teoria no lineal).

Resultats per les mostres del SDSS

Finalment, emprem el mètode per estudiar les BAO en el catàleg de SDSS. Fem ús del catàleg principal ('*Main*'), com a traçador del camp de densitat global, i les LRGs com a traçadores dels centres, ja que aquestes galàxies tendeixen a trobar-se prop dels centres dels halos més massius (Zheng et al., 2009). Podem estudiar aleshores només el rang en *redshift* en què les dues mostres coincideixen, la qual cosa redueix significativament el volum disponible. D'aquesta manera, estudiem el camp de densitat en un volum de $2.2 \times 10^8\,h^{-3}\,\mathrm{Mpc}^3$, on trobem $N_{LRG} = 1599$ LRGs com a centres.

Calculant $B(R,s)$ en aquest cas (Fig. 6.7), obtenim un resultat qualitativament semblant a allò esperat en el model ΛCDM, obtenint un màxim per $R_{\max} = 116\,h^{-1}\,\mathrm{Mpc}$, $s_{\max} = 36\,h^{-1}\,\mathrm{Mpc}$. Açò és una indicació que estem detectant les BAO presents en la mostra, tot i que els valors obtinguts per $R_{\max}$ i $s_{\max}$ es desvien lleugerament dels esperats. Aquest fet pot ser degut a possibles efectes observacionals que no hem modelat en l'apartat anterior. Estudiant el valor del màxim en B respecte a allò obtingut emprant centres distribuïts aleatòriament, obtenim una significació de la detecció de 4.4σ, tot i que aquest valor podria ser menor si empràrem un model menys simplificat.

Un dels avantatges del mètode de les ondetes és que podem emprar els mapes complets dels coeficients $W_{R,s}(\mathbf{x})$ per obtenir informació addicional sobre les BAO, i en particular sobre la distribució del senyal obtingut en l'espai de configuració.

Il·lustrem aquesta idea en la Fig. 6.9, on es mostra la distribució de les LRGs en l'espai, i al mateix temps el valor corresponent de $W_{\rm max} \equiv W_{R_{\rm max}, s_{\rm max}}$.

En definitiva, hem introduït un nou mètode que ens proporciona una detecció independent del fenomen de les BAO en la distribució de galàxies. D'aquesta manera, confirmem les deteccions obtingudes prèviament emprant els estadístics més generals (funció de correlació i espectre de potències). A més a més, aquest mètode proporciona informació addicional sobre la localització de les estructures responsables pel senyal de BAO. Aquesta informació pot ser útil per un estudi amb més detall de les BAO.

B.4 Conclusions

El treball presentat en aquesta tesi ha tractat l'anàlisi de dades de cartografiats de galàxies recents, i el desenvolupament de noves tècniques estadístiques per estudiar alguns problemes específics. La tesi està dividida en dues parts, depenent del rang d'escales considerats en cada cas, i per tant dels problemes científics relacionats amb cada rang d'escales.

En la primera part de la tesi, hem estudiat l'agrupament de galàxies a escales petites i intermèdies ($r \lesssim 20\,h^{-1}\,$Mpc), centrant-nos en l'evolució d'aquest agrupament, i la seua dependència de les propietats de les galàxies, o "segregació". Els Capítols 2 [B.2.1] i 3 [B.2.2] han tractat directament el problema d'estudiar l'agrupament a alt *redshift* traient partit de les possibilitats que ofereix el cartografiat ALHAMBRA, mentre que al Capítol 4 [B.2.3] hem fet una revisió més general sobre els mètodes de l'estadística de marques que es poden aplicar a l'anàlisi de la segregació.

Al Capítol 2 [B.2.1] hem desenvolupat un mètode per la recuperació de la informació sobre l'agrupament en espai real a partir de cartografiats fotomètrics amb les característiques d'ALHAMBRA. Aquest mètode es basa en la utilització de la funció de correlació projectada, i està adaptat per dades amb un error típic en la mesura dels *redshifts* fotomètrics de $\Delta z \lesssim 0.015(1+z)$. Hem mostrat, fent ús de dades d'una simulació de N-cossos que aquest mètode recupera la funció de correlació en l'espai real dins d'un marge de 5% per les escales estudiades. També hem comprovat que aquest mètode és fiable davant la presència d'una fracció petita d'errors "catastròfics" en la mesura dels *redshifts*. Una limitació d'aquest mètode és el fet que imposa un mínim en el rang de *redshifts* a considerar pel càlcul, depenent de Δz.

Tot i que aquest mètode es va dissenyar amb l'objectiu d'analitzar dades provinents d'ALHAMBRA, té un rang d'aplicació major, ja que es pot emprar per l'anàlisi de dades provinents de qualsevol cartografiat amb precisió semblant en la mesura de *redshifts*. Aquest és el cas d'altres cartografiats fotomètrics que empren un nombre gran de filtres de banda estreta o mitjana, com el realitzat per Ilbert et al. (2009) en el camp COSMOS, o el cartografiat en projecte J-PAS

(Benítez et al., 2009a), que cobrirà un àrea de $\sim 8000\,\mathrm{deg}^2$ en ~ 50 filtres en l'òptic, obtenint així una precisió molt major en la mesura de *redshifts*.

Hem aplicat aquest mètode al càlcul de la funció de correlació per diverses mostres provinents del cartografiat ALHAMBRA al Capítol 3 [B.2.2]. Les característiques d'ALHAMBRA ens permeten explorar un rang gran en *redshift*, $z \in [0.3, 1.5]$, i per tant aquesta és una eina ideal per l'estudi de l'evolució de l'agrupament de galàxies amb el temps còsmic. Hem dividit la nostra mostra original en tres intervals en *redshift*, i seleccionat diverses mostres de galàxies en cadascun d'ells depenent de la lluminositat en banda B. Donada la profunditat fotomètrica del cartografiat, en alguns casos arribem a emprar mostres de galàxies relativament febles, arribant a $L = 0.1L^*$ per $z \simeq 0.5$, o $L = 0.5L^*$ per $z \simeq 1.2$. Açò seria impossible en el cas de cartografiats espectroscòpics, ja que en aquest cas és necessari emprar fonts més brillants.

Les dades d'ALHAMBRA analitzades són encara preliminars. Tot i això, ja hem obtingut alguns resultats interessants. En el rang d'escales estudiat, $r \in [0.2, 15]\,h^{-1}$ Mpc, la funció de correlació de totes les mostres s'ajusta bé a una llei de potències, el que fa que la comparació entre diferents mostres siga més fàcil. Hem observat clarament els efectes de l'evolució, ja que les nostres mostres a alt *redshift* ($z \simeq 1.2$) exhibeixen un agrupament molt més feble que les mostres a *redshift* més baix. Pel que fa a la segregació per lluminositat, observem clarament els seus efectes en cadascun dels intervals en *redshift* analitzats. El fet d'observar segregació per lluminositat per mostres febles amb $L \leq L^*$ implica una diferència amb els estudis amb mostres locals ($z \simeq 0$), on aquest tipus de segregació és significatiu només per $L \geq L^*$.

Aquests resultats mostren les possibilitats d'ALHAMBRA per l'estudi de l'estructura a gran escala. El treball presentat ací es pot estendre de manera natural emprant un model més detallat de l'agrupament de galàxies per interpretar els resultats obtinguts per la funció de correlació. A més a més, encara que ací només hem estudiat la dependència de les propietats d'agrupament en la lluminositat, les dades d'ALHAMBRA permetran realitzar un estudi en més detall de la segregació de les galàxies, donada la quantitat d'informació que el cartografiat obté per cada objecte (color, tipus espectral, morfologia).

El fenomen de la segregació de galàxies pot ser analitzat en el marc de l'estadística de processos puntuals marcats. Hem examinat les eines bàsiques disponibles en aquest cas al Capítol 4 [B.2.3], il·lustrant la seua aplicació emprant una mostra de galàxies extreta del cartografiat 2dFGRS, i caracteritzada per un paràmetre de classificació espectral. Hem mostrat com aquestes eines de l'estadística de marques es poden emprar per estudiar la segregació en funció de l'escala, per al cas tant de marques contínues (com lluminositat, color, etc.) com de categories discretes. En aquest últim cas, hem introduït la funció de connexió de marques, i hem mostrat com aquesta ens dóna informació important quan analitzem diferents poblacions de galàxies, definides per un cert conjunt de característiques d'aquestes. Els nostres resultats per 2dFGRS confirmen observacions prèvies, mostrant en detall com les galàxies passives tendeixen a agrupar-se entre elles més fortament que les actives a escales petites.

Els mètodes d'anàlisi de la segregació que hem presentat ací són eines generals per l'anàlisi de dades en qualsevol cartografiat que obtinga informació tant sobre la distribució espacial com sobre la caracterització de les galàxies. En particular, seran útils per una anàlisi detallada de la segregació de galàxies emprant les dades d'ALHAMBRA. En aquest cas, haurem d'adaptar aquests mètodes per la seua utilització amb dades contenint *redshifts* fotomètrics, de manera semblant a allò que hem fet al Capítol 2 [B.2.1] amb la funció de correlació.

En la segona part de la tesi, hem passat a l'estudi de les correlacions a escales majors, centrant-nos en el fenomen de les BAO. Les BAO són una característica gravada al camp cosmològic de densitat durant les èpoques inicials de l'univers, i ens proporcionen una regla estàndard útil per restringir els possibles valors dels paràmetres cosmològics. Les BAO només han pogut ser detectades recentment (Eisenstein et al., 2005) en la distribució de galàxies.

Hem resumit la física de les BAO, i l'estatus de la seua detecció emprant la funció de correlació a dos punts estàndard al Capítol 5 [B.3.1]. Hem mesurat la funció de correlació per diverses mostres provinents dels majors cartografiats realitzats fins ara, el 2dFGRS i el SDSS. En tots els casos, hem obtingut un pic corresponent a les BAO a l'escala esperada, mostrant així la fiabilitat de la detecció d'aquesta característica. Resultats d'altres autors reforcen aquesta conclusió mitjançant l'estudi d'altres mostres independents. Tanmateix, hi ha una petita desviació respecte a la predicció del model ΛCDM, ja que el pic observat és més ample que allò esperat. Tot i que aquesta desviació podria no ser significativa, mereix una investigació més a fons.

Finalment, al Capítol 6 [B.3.2] hem desenvolupat un nou mètode per l'anàlisi del fenomen de les BAO. Aquest mètode empra les possibilitats proporcionades pels mètodes d'ondetes per cercar les estructures en espai de configuració conseqüència de les BAO. També està basat en l'ús de dos traçadors de massa complementaris, i l'il·lustrem emprant un catàleg format per galàxies de les mostres *'Main'* i 'LRG' del SDSS. D'aquesta manera, hem mostrat com no només som capaços de detectar les BAO en les mostres, sinó també de localitzar les regions que donen major o menor senyal de BAO. Aquesta informació es perd completament quan s'utilitzen els mètodes estadístics a dos punts tradicionals.

Aquest enfocament obre moltes noves possibilitats per l'anàlisi de les BAO, tot i que no les hem explorat en detall en aquesta tesi. Algunes de les possibles aplicacions podrien ser l'estudi de la relació entre aquestes estructures acústiques bariòniques i les propietats de les galàxies que les formen, o la selecció de regions amb alt senyal de BAO per refinar la caracterització del fenomen de BAO. Així mateix, aquesta informació espacial pot ser emprada per estudiar de forma contínua la dependència de les característiques de les BAO –i en particular l'escala de BAO– en el *redshift*. Tanmateix, aquestes possibles aplicacions estan limitades pel volum cartografiat disponible, especialment degut a la necessitat d'emprar dos traçadors de massa diferents. En aquest sentit, esperem que aquest mètode podrà ser d'utilitat pels cartografiats en marxa i futurs que cobreixen volums molt majors que el SDSS.

Bibliography

Abazajian, K.N., et al. (2009): The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, **182**:543. [`arXiv:0812.0649`].

Abbas, U., et al. (2010): The VIMOS-VLT Deep Survey: evolution in the halo occupation number since z $\sim$ 1. *MNRAS*, **406**:1306. [`arXiv:1003.6129`].

Abell, G.O. (1958): The Distribution of Rich Clusters of Galaxies. *ApJS*, **3**:211.

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

Alnes, H., Amarzguioui, M., and Grøn, Ø. (2006): Inhomogeneous alternative to dark energy? *Phys. Rev. D*, **73**(8):083519. [`arXiv:astro-ph/0512006`].

Aparicio Villegas, T., et al. (2010): The ALHAMBRA Photometric System. *AJ*, **139**:1242. [`arXiv:1001.3383`].

Arnalte-Mur, P., Fernández-Soto, A., Martínez, V.J., Saar, E., Heinämäki, P., and Suhhonenko, I. (2009): Recovering the real-space correlation function from photometric redshift surveys. *MNRAS*, **394**:1631. [`arXiv:0812.4226`].

Arnalte-Mur, P., Labatie, A., Clerc, N., Martínez, V.J., Starck, J.L., Lachièze-Rey, M., Saar, E., and Paredes, S. (2011): Wavelet analysis of Baryon Acoustic Structures in the galaxy distribution. *A&A*, **submitted**. [`arXiv:1101.1911`].

Babcock, H.W. (1939): The rotation of the Andromeda Nebula. *Lick Observatory Bulletin*, **19**:41.

Bardeen, J.M., Bond, J.R., Kaiser, N., and Szalay, A.S. (1986): The statistics of peaks of Gaussian random fields. *ApJ*, **304**:15.

Barrow, J.D., Bhavsar, S.P., and Sonoda, D.H. (1984): A bootstrap resampling analysis of galaxy clustering. *MNRAS*, **210**:19P.

Bassett, B.A. and Afshordi, N. (2010): The Non-Gaussian Sting in Posteriors arising from Marginal Detections. *ArXiv e-prints*. [`arXiv:1005.1664`].

Bassett, B.A. and Hlozek, R. (2010): Baryon Acoustic Oscillations. In: P. Ruiz-Lapuente (ed.), *Dark Energy*, pp. 246–278 (Cambridge University Press, Cambridge). [`arXiv:0910.5224`].

Baugh, C.M., Cole, S., and Frenk, C.S. (1996): Evolution of the Hubble sequence in hierarchical models for galaxy formation. *MNRAS*, **283**:1361. [`arXiv:astro-ph/9602085`].

Baum, W.A. (1962): Photoelectric Magnitudes and Red-Shifts. In: G.C. McVittie (ed.), *Problems of Extra-Galactic Research*, *IAU Symposium*, volume 15, pp. 390–+.

Beisbart, C. and Kerscher, M. (2000): Luminosity- and Morphology-dependent Clustering of Galaxies. *ApJ*, **545**:6. [`arXiv:astro-ph/0003358`].

Benítez, N. (2000): Bayesian Photometric Redshift Estimation. *ApJ*, **536**:571. [`arXiv:astro-ph/9811189`].

Benítez, N., et al. (2009a): Measuring Baryon Acoustic Oscillations Along the Line of Sight with Photometric Redshifts: The PAU Survey. *ApJ*, **691**:241. [`arXiv:0807.0535`].

—— (2009b): Optimal Filter Systems for Photometric Redshift Estimation. *ApJ*, **692**:L5. [`arXiv:0812.3568`].

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

Bernardeau, F., Colombi, S., Gaztañaga, E., and Scoccimarro, R. (2002): Large-scale structure of the Universe and cosmological perturbation theory. *Phys. Rep.*, **367**:1. [`arXiv:astro-ph/0112551`].

Bertin, E. and Arnouts, S. (1996): SExtractor: Software for source extraction. *A&AS*, **117**:393.

Beutler, F., et al. (2011): The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *MNRAS*, pp. 1164–+. [`arXiv:1106.3366`].

Blake, C. and Bridle, S. (2005): Cosmology with photometric redshift surveys. *MNRAS*, **363**:1329. [`arXiv:astro-ph/0411713`].

Blake, C. and Glazebrook, K. (2003): Probing Dark Energy Using Baryonic Oscillations in the Galaxy Power Spectrum as a Cosmological Ruler. *ApJ*, **594**:665. [`arXiv:astro-ph/0301632`].

Blake, C., et al. (2011): The WiggleZ Dark Energy Survey: testing the cosmological model with baryon acoustic oscillations at $z = 0.6$. *MNRAS*, pp. 951–+. [`arXiv:1105.2862`].

Blanton, M.R., et al. (2005): New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *AJ*, **129**:2562. [`arXiv:astro-ph/0410166`].

Bolzonella, M., Miralles, J.M., and Pelló, R. (2000): Photometric redshifts based on standard SED fitting procedures. *A&A*, **363**:476. [`arXiv:astro-ph/0003380`].

Bond, J.R. and Efstathiou, G. (1984): Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. *ApJ*, **285**:L45.

Broadhurst, T.J., Ellis, R.S., Koo, D.C., and Szalay, A.S. (1990): Large-scale distribution of galaxies at the Galactic poles. *Nature*, **343**:726.

Butcher, H. and Oemler, Jr., A. (1978): The evolution of galaxies in clusters. I - ISIT photometry of C1 0024+1654 and 3C 295. *ApJ*, **219**:18.

—— (1984): The evolution of galaxies in clusters. V - A study of populations since $z \sim 0.5$. *ApJ*, **285**:426.

Cabré, A. and Gaztañaga, E. (2009): Clustering of luminous red galaxies - I. Large-scale redshift-space distortions. *MNRAS*, **393**:1183. [`arXiv:0807.2460`].

—— (2011): Have baryonic acoustic oscillations in the galaxy distribution really been measured? *MNRAS*, **412**:L98. [`arXiv:1011.2729`].

CARROLL, S.M. (2001): The Cosmological Constant. *Living Reviews in Relativity*, **4**:1. [`arXiv:astro-ph/0004075`].

CARROLL, S.M., PRESS, W.H., and TURNER, E.L. (1992): The cosmological constant. *ARA&A*, **30**:499.

CHUI, C.K. (1992): *An Introduction to Wavelets* (Academic Press, San Diego).

CLOWE, D., BRADAČ, M., GONZALEZ, A.H., MARKEVITCH, M., RANDALL, S.W., JONES, C., and ZARITSKY, D. (2006): A Direct Empirical Proof of the Existence of Dark Matter. *ApJ*, **648**:L109. [`arXiv:astro-ph/0608407`].

COE, D., BENÍTEZ, N., SÁNCHEZ, S.F., JEE, M., BOUWENS, R., and FORD, H. (2006): Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology. *AJ*, **132**:926. [`arXiv:astro-ph/0605262`].

COHN, J.D. (2006): Power spectrum and correlation function errors: Poisson vs. Gaussian shot noise. *New Astronomy*, **11**:226. [`arXiv:astro-ph/0503285`].

COIL, A.L., NEWMAN, J.A., COOPER, M.C., DAVIS, M., FABER, S.M., KOO, D.C., and WILLMER, C.N.A. (2006): The DEEP2 Galaxy Redshift Survey: Clustering of Galaxies as a Function of Luminosity at z = 1. *ApJ*, **644**:671. [`arXiv:astro-ph/0512233`].

COLE, S., ET AL. (2005): The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, **362**:505. [`arXiv:astro-ph/0501174`].

COLLESS, M., ET AL. (2001): The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, **328**:1039. [`arXiv:astro-ph/0106498`].

—— (2003): The 2dF Galaxy Redshift Survey: Final Data Release. *ArXiv Astrophysics e-prints*. [`arXiv:astro-ph/0306581`].

COLLISTER, A.A. and LAHAV, O. (2004): ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP*, **116**:345. [`arXiv:astro-ph/0311058`].

CONNOLLY, A.J., CSABAI, I., SZALAY, A.S., KOO, D.C., KRON, R.G., and MUNN, J.A. (1995): Slicing Through Multicolor Space: Galaxy Redshifts from Broadband Photometry. *AJ*, **110**:2655. [`arXiv:astro-ph/9508100`].

COORAY, A. and SHETH, R. (2002): Halo models of large scale structure. *Phys. Rep.*, **372**:1. [`arXiv:astro-ph/0206508`].

CROCCE, M. and SCOCCIMARRO, R. (2006): Renormalized cosmological perturbation theory. *Phys. Rev. D*, **73**(6):063519. [`arXiv:astro-ph/0509418`].

—— (2008): Nonlinear evolution of baryon acoustic oscillations. *Phys. Rev. D*, **77**(2):023533. [`arXiv:0704.2783`].

CROCCE, M., FOSALBA, P., CASTANDER, F.J., and GAZTAÑAGA, E. (2010): Simulating the Universe with MICE: the abundance of massive clusters. *MNRAS*, **403**:1353. [`arXiv:0907.0019`].

CROTON, D.J., ET AL. (2004): The 2dF Galaxy Redshift Survey: voids and hierarchical scaling models. *MNRAS*, **352**:828. [`arXiv:astro-ph/0401406`].

CYBURT, R.H., FIELDS, B.D., and OLIVE, K.A. (2008): An update on the big bang nucleosynthesis prediction for ^{7}Li: the problem worsens. *JCAP*, **11**:12. [`arXiv:0808.2818`].

DA COSTA, L.N., PELLEGRINI, P.S., DAVIS, M., MEIKSIN, A., SARGENT, W.L.W., and TONRY, J.L. (1991): Southern Sky Redshift Survey - The catalog. *ApJS*, **75**:935.

DA COSTA, L.N., ET AL. (1998): The Southern Sky Redshift Survey. *AJ*, **116**:1. [`arXiv:astro-ph/9804064`].

DADDI, E., CIMATTI, A., RENZINI, A., FONTANA, A., MIGNOLI, M., POZZETTI, L., TOZZI, P., and ZAMORANI, G. (2004): A New Photometric Technique for the Joint Selection of Star-forming and Passive Galaxies at $1.4 \lesssim z \lesssim 2.5$. *ApJ*, **617**:746. [`arXiv:astro-ph/0409041`].

DAVIS, M. and GELLER, M.J. (1976): Galaxy Correlations as a Function of Morphological Type. *ApJ*, **208**:13.

DAVIS, M. and PEEBLES, P.J.E. (1983): A survey of galaxy redshifts. V - The two-point position and velocity correlations. *ApJ*, **267**:465.

DAVIS, M., MEIKSIN, A., STRAUSS, M.A., DA COSTA, L.N., and YAHIL, A. (1988): On the universality of the two-point galaxy correlation function. *ApJ*, **333**:L9.

DAVIS, M., ET AL. (2003): Science Objectives and Early Results of the DEEP2 Redshift Survey. In: P. Guhathakurta (ed.), *Discoveries and Research Prospects from 6- to 10-Meter-Class Telescopes II*, *Proceedings of the SPIE*, volume 4834, pp. 161–172. [`arXiv:astro-ph/0209419`].

DE LAPPARENT, V., GELLER, M.J., and HUCHRA, J.P. (1986): A slice of the universe. *ApJ*, **302**:L1.

DE RAVEL, L., ET AL. (2011): The zCOSMOS redshift survey : Influence of luminosity, mass and environment on the galaxy merger rate. *ArXiv e-prints.* [`arXiv:1104.5470`].

DMITRIEV, V.F., FLAMBAUM, V.V., and WEBB, J.K. (2004): Cosmological variation of the deuteron binding energy, strong interaction, and quark masses from big bang nucleosynthesis. *Phys. Rev. D*, **69**(6):063506. [`arXiv:astro-ph/0310892`].

DODELSON, S. (2003): *Modern cosmology* (Academic Press, Amsterdam).

DOMINGUEZ-TENREIRO, R. and MARTINEZ, V.J. (1989): Multidimensional analysis of the large-scale segregation of luminosity. *ApJ*, **339**:L9.

DOMINGUEZ-TENREIRO, R., GOMEZ-FLECHOSO, M.A., and MARTINEZ, V.J. (1994): Scaling analysis of the distribution of galaxies in the CfA catalog. *ApJ*, **424**:42.

DRESSLER, A. (1980): Galaxy morphology in rich clusters - Implications for the formation and evolution of galaxies. *ApJ*, **236**:351.

DRINKWATER, M.J., ET AL. (2010): The WiggleZ Dark Energy Survey: survey design and first data release. *MNRAS*, **401**:1429. [`arXiv:0911.4246`].

EFRON, B. and TIBSHIRANI, R.J. (1993): *An Introduction to the Bootstrap.* Number 57 in Monographs on Statistics and Applied Probability (Chapman & Hall/CRC, Boca Raton).

EINASTO, J., KAASIK, A., and SAAR, E. (1974): Dynamic evidence on massive coronas of galaxies. *Nature*, **250**:309.

EINSTEIN, A. (1915): Zur allgemeinen Relativitätstheorie. *Sitzungsberichte der*

Königlich Preußischen Akademie der Wissenschaften (Berlin), pp. 778–786.

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

—— (1999): Power Spectra for Cold Dark Matter and Its Variants. *ApJ*, **511**:5. [`arXiv:astro-ph/9710252`].

Eisenstein, D.J., Hu, W., Silk, J., and Szalay, A.S. (1998a): Can Baryonic Features Produce the Observed $100\,h^{-1}$ Mpc Clustering? *ApJ*, **494**:L1+. [`arXiv:astro-ph/9710303`].

Eisenstein, D.J., Hu, W., and Tegmark, M. (1998b): Cosmic Complementarity: H_0 and Ω_M from Combining Cosmic Microwave Background Experiments and Redshift Surveys. *ApJ*, **504**:L57+. [`arXiv:astro-ph/9805239`].

Eisenstein, D.J., Seo, H., and White, M. (2007): On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter. *ApJ*, **664**:660. [`arXiv:astro-ph/0604361`].

Eisenstein, D.J., et al. (2001): Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample. *AJ*, **122**:2267. [`arXiv:astro-ph/0108153`].

—— (2005): Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, **633**:560. [`arXiv:astro-ph/0501171`].

—— (2011): SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, **142**:72. [`arXiv:1101.1529`].

Farge, M. (1992): Wavelet transforms and their application to turbulence. *Ann. Rev. Fluid Mech.*, **24**:395.

Feng, J.L. (2010): Dark Matter Candidates from Particle Physics and Methods of Detection. *ARA&A*, **48**:495. [`arXiv:1003.0904`].

Fernández-Soto, A., Lanzetta, K.M., and Yahil, A. (1999): A New Catalog of Photometric Redshifts in the Hubble Deep Field. *ApJ*, **513**:34. [`arXiv:astro-ph/9809126`].

Fernández-Soto, A., Lanzetta, K.M., Chen, H.W., Pascarelle, S.M., and Yahata, N. (2001): On the Compared Accuracy and Reliability of Spectroscopic and Photometric Redshift Measurements. *ApJS*, **135**:41. [`arXiv:astro-ph/0007447`].

Fernández-Soto, A., Lanzetta, K.M., Chen, H.W., Levine, B., and Yahata, N. (2002): Error analysis of the photometric redshift technique. *MNRAS*, **330**:889. [`arXiv:astro-ph/0111227`].

Fosalba, P., Gaztañaga, E., Castander, F.J., and Manera, M. (2008): The onion universe: all sky lightcone simulations in spherical shells. *MNRAS*, **391**:435. [`arXiv:0711.1540`].

Freedman, W.L. and Madore, B.F. (2010): The Hubble Constant. *ARA&A*, **48**:673. [`arXiv:1004.1856`].

Freedman, W.L., et al. (2001): Final Results from the Hubble Space Telescope Key Project to Measure the Hubble Constant. *ApJ*, **553**:47. [`arXiv:astro-ph/0012376`].

Frieman, J.A., Turner, M.S., and Huterer, D. (2008): Dark Energy and the Accelerating Universe. *ARA&A*, **46**:385. [`arXiv:0803.0982`].

Fry, J.N. and Gaztanaga, E. (1993): Biasing and hierarchical statistics in large-scale structure. *ApJ*, **413**:447. [`arXiv:astro-ph/9302009`].

Gabrielli, A., Joyce, M., and Sylos Labini, F. (2002): Glass-like universe: Real-space correlation properties of standard cosmological models. *Phys. Rev. D*, **65**(8):083523. [`arXiv:astro-ph/0110451`].

Garcia-Bellido, J. and Haugbølle, T. (2008): Confronting Lemaitre Tolman Bondi models with observational cosmology. *JCAP*, **4**:3. [`arXiv:0802.1523`].

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

Gaztañaga, E., Cabré, A., and Hui, L. (2009b): Clustering of luminous red galaxies - IV. Baryon acoustic peak in the line-of-sight direction and a direct measurement of H(z). *MNRAS*, **399**:1663. [`arXiv:0807.3551`].

Gaztañaga, E., Miquel, R., and Sánchez, E. (2009c): First Cosmological Constraints on Dark Energy from the Radial Baryon Acoustic Scale. *Physical Review Letters*, **103**(9):091302. [`arXiv:0808.1921`].

Gold, B., et al. (2011): Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Galactic Foreground Emission. *ApJS*, **192**:15. [`arXiv:1001.4555`].

Goto, T., Yamauchi, C., Fujita, Y., Okamura, S., Sekiguchi, M., Smail, I., Bernardi, M., and Gomez, P.L. (2003): The morphology-density relation in the Sloan Digital Sky Survey. *MNRAS*, **346**:601. [`arXiv:astro-ph/0312043`].

Gursky, H., Kellogg, E., Murray, S., Leong, C., Tananbaum, H., and Giacconi, R. (1971): A Strong X-Ray Source in the Coma Cluster Observed by UHURU. *ApJ*, **167**:L81+.

Guth, A.H. (1981): Inflationary universe: A possible solution to the horizon and flatness problems. *Phys. Rev. D*, **23**:347.

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

Hamilton, A.J.S. (1988): Evidence for biasing in the CfA survey. *ApJ*, **331**:L59.

—— (1993): Toward Better Ways to Measure the Galaxy Correlation Function. *ApJ*, **417**:19.

—— (1998): Linear Redshift Distortions: a Review. In: D. Hamilton (ed.), *The Evolving Universe*, *Astrophysics and Space Science Library*, volume 231, pp. 185–+. [`arXiv:astro-ph/9708102`].

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

Hamilton, A.J.S., Kumar, P., Lu, E., and Matthews, A. (1991): Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies. *ApJ*, **374**:L1.

Hawkins, E., et al. (2003): The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. *MNRAS*, **346**:78.

[`arXiv:astro-ph/0212375`].

Heinämäki, P., Suhhonenko, I., Saar, E., Einasto, M., Einasto, J., and Virtanen, H. (2005): Light-cone simulations: Evolution of dark matter haloes. *ArXiv Astrophysics e-prints*. [`arXiv:astro-ph/0507197`].

Hermit, S., Santiago, B.X., Lahav, O., Strauss, M.A., Davis, M., Dressler, A., and Huchra, J.P. (1996): The two-point correlation function and morphological segregation in the Optical Redshift Survey. *MNRAS*, **283**:709. [`arXiv:astro-ph/9608001`].

Hickson, P., Gibson, B.K., and Callaghan, K.A.S. (1994): Multinarrowband Imaging - a New Technique for Multi-Object Spectrophotometry. *MNRAS*, **267**:911.

Hogg, D.W. (1999): Distance measures in cosmology. *ArXiv Astrophysics e-prints*. [`arXiv:astro-ph/9905116`].

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

Holtzman, J.A. (1989): Microwave background anisotropies and large-scale structure in universes with cold dark matter, baryons, radiation, and massive and massless neutrinos. *ApJS*, **71**:1.

Horvath, J.E. (2009): Dark matter, dark energy and modern cosmology: the case for a Kuhnian paradigm shift. *Cosmos Hist.*, **5**:287. [`arXiv:0809.2839`].

Hoskin, M.A. (1976): The 'Great Debate': What Really Happened. *Journal for the History of Astronomy*, **7**:169.

Hu, W., Sugiyama, N., and Silk, J. (1997): The Physics of Microwave Background Anisotropies. *Nature*, **386**:37. [`arXiv:astro-ph/9604166`].

Hubble, E. (1925a): Cepheids in Spiral Nebulae. *Popular Astronomy*, **33**:252.

—— (1925b): NGC 6822, a remote stellar system. *ApJ*, **62**:409.

—— (1929): A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, **15**:168.

Huchra, J., Davis, M., Latham, D., and Tonry, J. (1983): A survey of galaxy redshifts. IV - The data. *ApJS*, **52**:89.

Hui, L., Gaztañaga, E., and Loverde, M. (2007): Anisotropic magnification distortion of the 3D galaxy correlation. I. Real space. *Phys. Rev. D*, **76**(10):103502. [`arXiv:0706.1071`].

Humason, M.L., Mayall, N.U., and Sandage, A.R. (1956): Redshifts and magnitudes of extragalactic nebulae. *AJ*, **61**:97.

Hütsi, G. (2006): Acoustic oscillations in the SDSS DR4 luminous red galaxy sample power spectrum. *A&A*, **449**:891. [`arXiv:astro-ph/0512201`].

—— (2010): Power spectrum of the maxBCG sample: detection of acoustic oscillations using galaxy clusters. *MNRAS*, **401**:2477. [`arXiv:0910.0492`].

Icke, V. and van de Weygaert, R. (1987): Fragmenting the universe. *A&A*, **184**:16.

Ilbert, O., et al. (2005): The VIMOS-VLT deep survey. Evolution of the galaxy luminosity function up to z = 2 in first epoch data. *A&A*, **439**:863. [`arXiv:astro-ph/0409134`].

—— (2006): Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey. *A&A*, **457**:841. [`arXiv:astro-ph/0603217`].

—— (2009): Cosmos Photometric Redshifts with 30-Bands for 2-deg^2. *ApJ*, **690**:1236. [`arXiv:0809.2101`].

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008): *Statistical Analysis and Modelling of Spatial Point Patterns.* Statistics in Practice (John Wiley & Sons, Chischester).

Jarosik, N., et al. (2011): Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *ApJS*, **192**:14. [`arXiv:1001.4744`].

Jedamzik, K. (2004): Did something decay, evaporate, or annihilate during big bang nucleosynthesis? *Phys. Rev. D*, **70**(6):063524. [`arXiv:astro-ph/0402344`].

Jing, Y.P., Mo, H.J., and Boerner, G. (1998): Spatial Correlation Function and Pairwise Velocity Dispersion of Galaxies: Cold Dark Matter Models versus the Las Campanas Survey. *ApJ*, **494**:1. [`arXiv:astro-ph/9707106`].

Jones, B.J., Martínez, V.J., Saar, E., and Trimble, V. (2005): Scaling laws in the distribution of galaxies. *Reviews of Modern Physics*, **76**:1211. [`arXiv:astro-ph/0406086`].

Jones, B.J.T. (2009): The Sea of Wavelets. In: V.J. Martínez, E. Saar, E. Martínez-González, and M.J. Pons-Bordería (eds.), *Data Analysis in Cosmology*, *Lecture Notes in Physics, Berlin Springer Verlag*, volume 665, pp. 3–50.

Jones, D.H., et al. (2004): The 6dF Galaxy Survey: samples, observational techniques and the first data release. *MNRAS*, **355**:747. [`arXiv:astro-ph/0403501`].

—— (2009): The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures. *MNRAS*, **399**:683. [`arXiv:0903.5451`].

Kaiser, N. (1984): On the spatial correlations of Abell clusters. *ApJ*, **284**:L9.

—— (1987): Clustering in real space and in redshift space. *MNRAS*, **227**:1.

Kaiser, N., et al. (2002): Pan-STARRS: A Large Synoptic Survey Telescope Array. In: J.A. Tyson and S. Wolff (eds.), *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 4836, pp. 154–164.

Kazin, E.A., Blanton, M.R., Scoccimarro, R., McBride, C.K., and Berlind, A.A. (2010a): Regarding the Line-of-sight Baryonic Acoustic Feature in the Sloan Digital Sky Survey and Baryon Oscillation Spectroscopic Survey Luminous Red Galaxy Samples. *ApJ*, **719**:1032. [`arXiv:1004.2244`].

Kazin, E.A., et al. (2010b): The Baryonic Acoustic Feature and Large-Scale Clustering in the Sloan Digital Sky Survey Luminous Red Galaxy Sample. *ApJ*, **710**:1444. [`arXiv:0908.2598`].

Kellogg, E., Gursky, H., Leong, C., Schreier, E., Tananbaum, H., and Giacconi, R. (1971): X-Ray Observations of the Virgo Cluster, NGC 5128, and 3C 273 from the UHURU Satellite. *ApJ*, **165**:L49+.

Kennicutt, Jr., R.C. (1992): A spectrophotometric atlas of galaxies. *ApJS*, **79**:255.

Kerscher, M. (1999): The geometry of second-order statistics - biases in common estimators. *A&A*, **343**:333. [`arXiv:astro-ph/9811300`].

Knebe, A., Green, A., and Binney, J. (2001): Multi-level Adaptive Particle Mesh (MLAPM): a C code for cosmological simulations. *MNRAS*, **325**:845.

Komatsu, E., et al. (2009): Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. *ApJS*, **180**:330. [`arXiv:0803.0547`].

—— (2011): Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *ApJS*, **192**:18. [`arXiv:1001.4538`].

Koo, D.C. (1985): Optical multicolors - A poor person's Z machine for galaxies. *AJ*, **90**:418.

—— (1986): Quests for primeval galaxies - A review of optical surveys. In: C. Chiosi and A. Renzini (eds.), *Spectral Evolution of Galaxies*, *Astrophysics and Space Science Library*, volume 122, pp. 419–438.

Korn, A.J., Grundahl, F., Richard, O., Barklem, P.S., Mashonkina, L., Collet, R., Piskunov, N., and Gustafsson, B. (2006): A probable stellar solution to the cosmological lithium discrepancy. *Nature*, **442**:657. [`arXiv:astro-ph/0608201`].

Kowalski, M., et al. (2008): Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets. *ApJ*, **686**:749. [`arXiv:0804.4142`].

Kravtsov, A.V., Berlind, A.A., Wechsler, R.H., Klypin, A.A., Gottlöber, S., Allgood, B., and Primack, J.R. (2004): The Dark Side of the Halo Occupation Distribution. *ApJ*, **609**:35. [`arXiv:astro-ph/0308519`].

Labatie, A., Starck, J., Lachièze-Rey, M., and Arnalte-Mur, P. (2011): Uncertainty in 2-point correlation function estimators and baryon acoustic oscillation detection in galaxy surveys. *Statistical Methodology*, **in press**. [`arXiv:1009.1232`].

Lahiri, S.N. (2003): *Resampling Methods for Dependent Data*. Springer Series in Statistics (Springer-Verlag, New York).

Landy, S.D. and Szalay, A.S. (1993): Bias and variance of angular correlation functions. *ApJ*, **412**:64.

Landy, S.D., Shectman, S.A., Lin, H., Kirshner, R.P., Oemler, A.A., and Tucker, D. (1996): The Two-dimensional Power Spectrum of the Las Campanas Redshift Survey: Detection of Excess Power on 100 H -1 MPC Scales. *ApJ*, **456**:L1+. [`arXiv:astro-ph/9510146`].

Larson, D., et al. (2011): Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters. *ApJS*, **192**:16. [`arXiv:1001.4635`].

Le Fèvre, O., et al. (2005): The VIMOS VLT deep survey. First epoch VVDS-deep survey: 11 564 spectra with $17.5 \leq I_{AB} \leq 24$, and the redshift distribution over $0 \leq z \leq 5$. *A&A*, **439**:845. [`arXiv:astro-ph/0409133`].

Leavitt, H.S. (1908): 1777 variables in the Magellanic Clouds. *Annals of Harvard College Observatory*, **60**:87.

Leavitt, H.S. and Pickering, E.C. (1912): Periods of 25 Variable Stars in the Small Magellanic Cloud. *Harvard College Observatory Circular*, **173**:1.

Lemaître, G. (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.

Lewis, A., Challinor, A., and Lasenby, A. (2000): Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. *ApJ*, **538**:473. [`arXiv:astro-ph/9911177`].

Lintott, C.J., et al. (2008): Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS*, **389**:1179. [`arXiv:0804.4483`].

López-Sanjuan, C., et al. (2011): The VIMOS VLT Deep Survey. The contribution of minor mergers to the growth of $L_B \gtrsim L_B^*$ galaxies since $z \sim 1$ from spectroscopically identified pairs. *A&A*, **530**:A20+. [`arXiv:1009.5921`].

Loveday, J., Maddox, S.J., Efstathiou, G., and Peterson, B.A. (1995): The Stromlo-APM redshift survey. II. Variation of galaxy clustering with morphology and luminosity. *ApJ*, **442**:457. [`arXiv:astro-ph/9410018`].

Lynds, R. and Petrosian, V. (1986): Giant Luminous Arcs in Galaxy Clusters. *BAAS*, **18**:1014.

Maddox, S.J., Efstathiou, G., Sutherland, W.J., and Loveday, J. (1990): The APM galaxy survey. I - APM measurements and star-galaxy separation. *MNRAS*, **243**:692.

Madgwick, D.S., Somerville, R., Lahav, O., and Ellis, R. (2003a): Galaxy spectral parametrization in the 2dF Galaxy Redshift Survey as a diagnostic of star formation history. *MNRAS*, **343**:871. [`arXiv:astro-ph/0210471`].

Madgwick, D.S., et al. (2002): The 2dF Galaxy Redshift Survey: galaxy luminosity functions per spectral type. *MNRAS*, **333**:133. [`arXiv:astro-ph/0107197`].

—— (2003b): The 2dF Galaxy Redshift Survey: galaxy clustering per spectral type. *MNRAS*, **344**:847. [`arXiv:astro-ph/0303668`].

Mallat, S. (2008): *A Wavelet Tour of Signal Processing, The Sparse Way* (Academic Press), 3rd edition.

Mandelbrot, B. (1975): Sur un modèle décomposable d'Universe hiérarchisé: déduction des corrélations galactiques sur la sphère céleste. *C.R. Acad. Sci. (Paris)*, **280**:1551.

Martínez, V.J. (1999): Is the Universe Fractal? *Science*, **284**:445.

Martínez, V.J. and Saar, E. (2002): *Statistics of the Galaxy Distribution* (Chapman & Hall/CRC, Boca Raton).

Martínez, V.J. and Trimble, V. (2009): Cosmologists in the Dark. In: J. Rubiño-Martín, J. Belmonte, F. Prada, and A. Alberdi (eds.), *Cosmology Across Cultures*, *Astronomical Society of the Pacific Conference Series*, volume 409, pp. 47–+. [`arXiv:0904.1126`].

Martínez, V.J., Paredes, S., and Saar, E. (1993): Wavelet analysis of the multifractal character of the galaxy distribution. *MNRAS*, **260**:365.

Martínez, V.J., Arnalte-Mur, P., Saar, E., de la Cruz, P., Pons-Bordería, M.J., Paredes, S., Fernández-Soto, A., and Tempel, E. (2009): Reliability of the Detection of the Baryon Acoustic Peak. *ApJ*, **696**:L93. [`arXiv:0812.2154`].

Martínez, V.J., Arnalte-Mur, P., and Stoyan, D. (2010): Measuring galaxy segregation with the mark connection function. *A&A*, **513**:A22+. [`arXiv:1001.1294`].

Masjedi, M., et al. (2006): Very Small Scale Clustering and Merger Rate of Luminous Red Galaxies. *ApJ*, **644**:54. [`arXiv:astro-ph/0512166`].

Mather, J.C., et al. (1994): Measurement of the cosmic microwave background spectrum by the COBE FIRAS instrument. *ApJ*, **420**:439.

Matsubara, T. (1999): Stochasticity of Bias and Nonlocality of Galaxy Formation: Linear Scales. *ApJ*, **525**:543. [`arXiv:astro-ph/9906029`].

McVittie, G.C. (1956): *General Relativity and Cosmology* (Chapman and Hall, London).

Meiksin, A., White, M., and Peacock, J.A. (1999): Baryonic signatures in large-scale structure. *MNRAS*, **304**:851. [`arXiv:astro-ph/9812214`].

Meisenheimer, K., et al. (1998): The Calar Alto Deep Imaging Survey for Galaxies and Quasars at $z > 5$. In: S. D'Odorico, A. Fontana, and E. Giallongo (eds.), *The Young Universe: Galaxy Formation and Evolution at Intermediate and High Redshift*, *Astronomical Society of the Pacific Conference Series*, volume 146, pp. 134–+.

Meléndez, J. and Ramírez, I. (2004): Reappraising the Spite Lithium Plateau: Extremely Thin and Marginally Consistent with WMAP Data. *ApJ*, **615**:L33. [`arXiv:astro-ph/0409383`].

Meneux, B., et al. (2009): The zCOSMOS survey. The dependence of clustering on luminosity and stellar mass at $z = 0.2 - 1$. *A&A*, **505**:463. [`arXiv:0906.1807`].

Miller, C.J., Nichol, R.C., and Chen, X. (2002): Detecting the Baryons in Matter Power Spectra. *ApJ*, **579**:483. [`arXiv:astro-ph/0207180`].

Miralda-Escudé, J. (2009): Comment on the claimed radial BAO detection by Gaztanaga et al. *ArXiv e-prints*. [`arXiv:0901.1219`].

Mo, H.J. and White, S.D.M. (1996): An analytic model for the spatial clustering of dark matter haloes. *MNRAS*, **282**:347. [`arXiv:astro-ph/9512127`].

—— (2002): The abundance and clustering of dark haloes in the standard ΛCDM cosmogony. *MNRAS*, **336**:112. [`arXiv:astro-ph/0202393`].

Mo, H.J., Jing, Y.P., and White, S.D.M. (1997): High-order correlations of peaks and haloes: a step towards understanding galaxy biasing. *MNRAS*, **284**:189. [`arXiv:astro-ph/9603039`].

Moles, M., et al. (2008): The Alhambra Survey: a Large Area Multimedium-Band Optical and Near-Infrared Photometric Survey. *AJ*, **136**:1325. [`arXiv:0806.3021`].

Nakamura, K. et al. (Particle Data Group) (2010): Review of particle physics. *J. Phys.*, **G37**:075021.

Neyman, J. and Scott, E.L. (1952): A Theory of the Spatial Distribution of Galaxies. *ApJ*, **116**:144.

Norberg, P., Baugh, C.M., Gaztañaga, E., and Croton, D.J. (2009): Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics. *MNRAS*, **396**:19. [`arXiv:0810.1885`].

Norberg, P., et al. (2001): The 2dF Galaxy Redshift Survey: luminosity dependence of galaxy clustering. *MNRAS*, **328**:64. [`arXiv:astro-ph/0105500`].

—— (2002): The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type. *MNRAS*, **332**:827. [`arXiv:astro-ph/0112043`].

Oke, J.B. and Sandage, A. (1968): Energy Distributions, K Corrections, and the Stebbins-Whitford Effect for Giant Elliptical Galaxies. *ApJ*, **154**:21.

Okumura, T., Matsubara, T., Eisenstein, D.J., Kayo, I., Hikage, C., Szalay, A.S., and Schneider, D.P. (2008): Large-Scale Anisotropic Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, **676**:889. [`arXiv:0711.3640`].

Olive, K.A., Steigman, G., and Walker, T.P. (2000): Primordial nucleosynthesis: theory and observations. *Phys. Rep.*, **333**:389. [`arXiv:astro-ph/9905320`].

Ostriker, J.P., Peebles, P.J.E., and Yahil, A. (1974): The size and mass of galaxies, and the mass of the universe. *ApJ*, **193**:L1.

Paczynski, B. (1987): Giant luminous arcs discovered in two clusters of galaxies. *Nature*, **325**:572.

Padmanabhan, N., et al. (2007): The clustering of luminous red galaxies in the Sloan Digital Sky Survey imaging data. *MNRAS*, **378**:852. [`arXiv:astro-ph/0605302`].

Peacock, J.A. (1999): *Cosmological Physics* (Cambridge University Press, Cambridge, UK).

Peacock, J.A. and Smith, R.E. (2000): Halo occupation numbers and galaxy bias. *MNRAS*, **318**:1144. [`arXiv:astro-ph/0005010`].

Peacock, J.A., et al. (2001): A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. *Nature*, **410**:169. [`arXiv:astro-ph/0103143`].

Peebles, P.J.E. (1974): The Gravitational-Instability Picture and the Nature of the Distribution of Galaxies. *ApJ*, **189**:L51+.

—— (1980): *The large-scale structure of the universe* (Princeton University Press, Princeton).

—— (1993): *Principles of physical cosmology* (Princeton Series in Physics, Princeton, NJ: Princeton University Press, —c1993).

Peebles, P.J.E. and Hauser, M.G. (1974): Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs. *ApJS*, **28**:19.

Peebles, P.J.E. and Ratra, B. (2003): The cosmological constant and dark energy. *Reviews of Modern Physics*, **75**:559. [`arXiv:astro-ph/0207347`].

Peebles, P.J.E. and Yu, J.T. (1970): Primeval Adiabatic Perturbation in an Expanding Universe. *ApJ*, **162**:815.

Penzias, A.A. and Wilson, R.W. (1965): A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, **142**:419.

Percival, W.J., Cole, S., Eisenstein, D.J., Nichol, R.C., Peacock, J.A., Pope, A.C., and Szalay, A.S. (2007a): Measuring the Baryon Acoustic Oscillation scale using the Sloan Digital Sky Survey and 2dF Galaxy Redshift Survey. *MNRAS*, **381**:1053. [`arXiv:0705.3323`].

Percival, W.J., et al. (2001): The 2dF Galaxy Redshift Survey: the power spectrum and the matter content of the Universe. *MNRAS*, **327**:1297. [`arXiv:astro-ph/0105252`].

—— (2007b): Measuring the Matter Density Using Baryon Oscillations in the SDSS. *ApJ*, **657**:51. [`arXiv:astro-ph/0608635`].

—— (2010): Baryon acoustic oscillations in the Sloan Digital Sky Survey Data Release 7 galaxy sample. *MNRAS*, **401**:2148. [`arXiv:0907.1660`].

Perlmutter, S., et al. (1999): Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, **517**:565. [`arXiv:astro-ph/9812133`].

Phillipps, S. and Shanks, T. (1987): On the variation of galaxy correlations with luminosity. *MNRAS*, **229**:621.

Phleps, S., Peacock, J.A., Meisenheimer, K., and Wolf, C. (2006): Galaxy clustering from COMBO-17: the halo occupation distribution at $\langle z \rangle = 0.6$. *A&A*, **457**:145. [`arXiv:astro-ph/0506320`].

Pollo, A., et al. (2006): The VIMOS-VLT Deep Survey. Luminosity dependence of clustering at z $\simeq$ 1. *A&A*, **451**:409. [`arXiv:astro-ph/0512429`].

Pons-Bordería, M.J., Martínez, V.J., Stoyan, D., Stoyan, H., and Saar, E. (1999): Comparing Estimators of the Galaxy Correlation Function. *ApJ*, **523**:480. [`arXiv:astro-ph/9906344`].

Postman, M. and Geller, M.J. (1984): The morphology-density relation - The group connection. *ApJ*, **281**:95.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2002): *Numerical Recipes in C++ : The Art of Scientific Computing* (Cambridge University Press, Cambridge), 2nd edition.

Rauzy, S., Lachièze-Rey, M., and Henriksen, R.N. (1993): Wavelet analysis of cosmic velocity fields. *A&A*, **273**:357.

Reid, B.A., et al. (2010): Cosmological constraints from the clustering of the Sloan Digital Sky Survey DR7 luminous red galaxies. *MNRAS*, **404**:60. [`arXiv:0907.1659`].

Riess, A.G., et al. (1998): Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, **116**:1009. [`arXiv:astro-ph/9805201`].

Rivolo, A.R. (1986): The two-point galaxy correlation function of the Local Supercluster. *ApJ*, **301**:70.

Rubin, V.C. and Ford, Jr., W.K. (1970): Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *ApJ*, **159**:379.

Rubin, V.C., Thonnard, N., and Ford, Jr., W.K. (1978): Extended rotation curves of high-luminosity spiral galaxies. IV - Systematic dynamical properties, Sa→Sc. *ApJ*, **225**:L107.

Saar, E. (2009): Multiscale Methods. In: V.J. Martínez, E. Saar, E. Martínez-González, and M.J. Pons-Bordería (eds.), *Data Analysis in Cosmology*, *Lecture Notes in Physics, Berlin Springer Verlag*, volume 665, pp. 523–563.

Saar, E., Martínez, V.J., Starck, J., and Donoho, D.L. (2007): Multiscale morphology of the galaxy distribution. *MNRAS*, **374**:1030. [`arXiv:astro-ph/0610958`].

Salzer, J.J., Hanson, M.M., and Gavazzi, G. (1990): The relative spatial distributions of high- and low-luminosity galaxies toward Coma. *ApJ*, **353**:39.

Sánchez, A.G. and Cole, S. (2008): The galaxy power spectrum: precision cosmology from large-scale structure? *MNRAS*, **385**:830. [`arXiv:0708.1517`].

Sánchez, A.G., Crocce, M., Cabré, A., Baugh, C.M., and Gaztañaga, E. (2009): Cosmological parameter constraints from SDSS luminous red galaxies: a new treatment of large-scale clustering. *MNRAS*, **400**:1643. [`arXiv:0901.2570`].

Saslaw, W.C. and Sheth, R.K. (1993): Nonlinear properties and time evolution of gravitational galaxy clustering. *ApJ*, **409**:504.

Saunders, W., Rowan-Robinson, M., and Lawrence, A. (1992): The spatial correlation function of IRAS galaxies on small and intermediate scales. *MNRAS*, **258**:134.

Saunders, W., et al. (2000): The PSCz catalogue. *MNRAS*, **317**:55. [`arXiv:astro-ph/0001117`].

Sawangwit, U., Shanks, T., Abdalla, F.B., Cannon, R.D., Croom, S.M., Edge, A.C., Ross, N.P., and Wake, D.A. (2009): Angular correlation function of 1.5 million LRGs: clustering evolution and a search for BAO. *ArXiv e-prints.* [`arXiv:0912.0511`].

Scalo, J.M. (1986): The stellar initial mass function. *Fund. Cosmic Phys.*, **11**:1.

Schlather, M. (2001): On the second-order characteristics of marked point processes. *Bernoulli*, **7**(1):99.

Scoccimarro, R., Sheth, R.K., Hui, L., and Jain, B. (2001): How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering. *ApJ*, **546**:20. [`arXiv:astro-ph/0006319`].

Seljak, U. (2000): Analytic model for galaxy and dark matter clustering. *MNRAS*, **318**:203. [`arXiv:astro-ph/0001493`].

Seljak, U. and Zaldarriaga, M. (1996): A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, **469**:437. [`arXiv:astro-ph/9603033`].

Seo, H.J. and Eisenstein, D.J. (2005): Baryonic Acoustic Oscillations in Simulated Galaxy Redshift Surveys. *ApJ*, **633**:575. [`arXiv:astro-ph/0507338`].

Shane, C.D. and Wirtanen, C.A. (1967): The Lick catalog of galaxies. *Publ. Lick Obs.*, **22**:1.

Shectman, S.A., Landy, S.D., Oemler, A., Tucker, D.L., Lin, H., Kirshner, R.P., and Schechter, P.L. (1996): The Las Campanas Redshift Survey. *ApJ*, **470**:172. [`arXiv:astro-ph/9604167`].

Sheth, R.K. and Saslaw, W.C. (1996): Scale Dependence of Nonlinear Gravitational Clustering in the Universe. *ApJ*, **470**:78.

Sheth, R.K. and Tormen, G. (1999): Large-scale bias and the peak background split. *MNRAS*, **308**:119. [`arXiv:astro-ph/9901122`].

Sheth, R.K., Connolly, A.J., and Skibba, R. (2005): Marked correlations in galaxy formation models. *ArXiv Astrophysics e-prints.* [`arXiv:astro-ph/0511773`].

Skibba, R.A., et al. (2009): Galaxy Zoo: disentangling the environmental dependence of morphology and colour. *MNRAS*, **399**:966. [`arXiv:0811.3970`].

Smith, R.E., et al. (2003): Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, **341**:1311. [`arXiv:astro-ph/0207664`].

Snethlage, M., Martínez, V.J., Stoyan, D., and Saar, E. (2002): Point field models for the galaxy point pattern. Modelling the singularity of the two-point correlation function. *A&A*, **388**:758. [`arXiv:astro-ph/0204211`].

Soneira, R.M. and Peebles, P.J.E. (1978): A computer model universe - Simulation of the nature of the galaxy distribution in the Lick catalog. *AJ*, **83**:845.

Soucail, G., Fort, B., Mellier, Y., and Picat, J.P. (1987): A blue ring-like structure, in the center of the A 370 cluster of galaxies. *A&A*, **172**:L14.

Springel, V. (2005): The cosmological simulation code GADGET-2. *MNRAS*, **364**:1105. [`arXiv:astro-ph/0505010`].

Starck, J.L. and Murtagh, F. (2006): *Astronomical Image and Data Analysis* (Springer, Berlin), 2nd edition.

Starck, J.L., Moudden, Y., Abrial, P., and Nguyen, M. (2006): Wavelets, ridgelets and curvelets on the sphere. *A&A*, **446**:1191. [`arXiv:astro-ph/0509883`].

Starck, J.L., Murtagh, F., and Fadili, M. (2010): *Sparse Image and Signal Processing* (Cambridge University Press).

Stefanon, M. (2011): *Multi-wavelength surveys: object detectability and NIR luminosity function of galaxies*. Ph.D. thesis, Universitat de València.

Stoughton, C., et al. (2002): Sloan Digital Sky Survey: Early Data Release. *AJ*, **123**:485.

Stoyan, D. (1984): On correlations of marked point processes. *Math. Nachr.*, **116**:197.

Stoyan, D. and Penttinen, A. (2000): Recent Applications of Point Process Methods in Forestry Statistics. *Statistical Science*, **15**(1):61.

Stoyan, D. and Stoyan, H. (1994): *Fractals, Random Shapes and Point Fields*. Probability and Mathematical Statistics (John Wiley & Sons, Chischester).

Stoyan, D., Kendall, W.S., and Mecke, J. (1995): *Stochastic Geometry and its Applications*. Probablity and Mathematical Statistics (John Wiley & Sons, Chischester, England), 2nd edition.

Strauss, M.A., et al. (2002): Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Main Galaxy Sample. *AJ*, **124**:1810. [`arXiv:astro-ph/0206225`].

Swanson, M.E.C., Tegmark, M., Hamilton, A.J.S., and Hill, J.C. (2008): Methods for rapidly processing angular masks of next-generation galaxy surveys. *MNRAS*, **387**:1391. [`arXiv:0711.4352`].

Sylos Labini, F. (2010): Characterizing the large scale inhomogeneity of the galaxy distribution. In: J.M. Alimi and A. Fuözfa (eds.), *Invisible Universe*, *American Institute of Physics Conference Series*, volume 1241, pp. 981–990. [`arXiv:0910.3833`].

Sylos Labini, F., Vasilyev, N.L., Baryshev, Y.V., and López-Corredoira, M. (2009): Absence of anti-correlations and of baryon acoustic oscillations in the galaxy correlation function from the Sloan Digital Sky Survey data release 7. *A&A*, **505**:981. [`arXiv:0903.0950`].

Szalay, A.S., Broadhurst, T.J., Ellman, N., Koo, D.C., and Ellis, R.S. (1993): Redshift Survey with Multiple Pencil Beams at the Galactic Poles. *Proceedings of the National Academy of Science*, **90**:4853.

Tegmark, M., et al. (2004): The Three-Dimensional Power Spectrum of Galaxies from the Sloan Digital Sky Survey. *ApJ*, **606**:702. [`arXiv:astro-ph/0310725`].

—— (2006): Cosmological constraints from the SDSS luminous red galaxies. *Phys. Rev. D*, **74**(12):123507. [`arXiv:astro-ph/0608632`].

Thomas, S.A., Abdalla, F.B., and Lahav, O. (2011): Excess Clustering on Large Scales in the MegaZ DR7 Photometric Redshift Survey. *Physical Review Letters*, **106**(24):241301. [`arXiv:1012.2272`].

Tian, H.J., Neyrinck, M.C., Budavári, T., and Szalay, A.S. (2011): Redshift-space Enhancement of Line-of-sight Baryon Acoustic Oscillations in the Sloan Digital Sky Survey Main-galaxy Sample. *ApJ*, **728**:34. [`arXiv:1011.2481`].

Tomita, K. (2001): A local void and the accelerating Universe. *MNRAS*, **326**:287. [`arXiv:astro-ph/0011484`].

Totsuji, H. and Kihara, T. (1969): The Correlation Function for the Distribution of Galaxies. *PASJ*, **21**:221.

Trimble, V. (1995): The 1920 Shapley-Curtis Discussion: Background, Issues, and Aftermath. *PASP*, **107**:1133.

Tucker, D.L., et al. (1997): The Las Campanas Redshift Survey galaxy-galaxy autocorrelation function. *MNRAS*, **285**:L5. [`arXiv:astro-ph/9611206`].

van de Weygaert, R. and Icke, V. (1989): Fragmenting the universe. II - Voronoi vertices as Abell clusters. *A&A*, **213**:1.

Vielva, P., Martínez-González, E., Barreiro, R.B., Sanz, J.L., and Cayón, L. (2004): Detection of Non-Gaussianity in the Wilkinson Microwave Anisotropy Probe First-Year Data Using Spherical Wavelets. *ApJ*, **609**:22. [`arXiv:astro-ph/0310273`].

Vogeley, M.S., Geller, M.J., and Huchra, J.P. (1991): Void statistics of the CfA redshift survey. *ApJ*, **382**:44.

Wälder, O. and Stoyan, D. (1996): On variograms in point process statistics. *Biometrical Journal*, **38**(8):895.

Wolf, C., Meisenheimer, K., Rix, H.W., Borch, A., Dye, S., and Kleinheinrich, M. (2003): The COMBO-17 survey: Evolution of the galaxy luminosity function from 25 000 galaxies with $0.2 < z < 1.2$. *A&A*, **401**:73. [`arXiv:astro-ph/0208345`].

Xu, X., et al. (2010): A New Statistic for Analyzing Baryon Acoustic Oscillations. *ApJ*, **718**:1224. [`arXiv:1001.2324`].

Yoo, J. and Miralda-Escudé, J. (2009): Gravitational Lensing Effects on the Baryonic Acoustic Oscillation Signature in the Redshift-Space Correlation Function. *ArXiv e-prints*. [`arXiv:0901.0708`].

York, D.G., et al. (2000): The Sloan Digital Sky Survey: Technical Summary. *AJ*, **120**:1579. [`arXiv:astro-ph/0006396`].

Zaldarriaga, M. and Seljak, U. (2000): CMBFAST for Spatially Closed Universes. *ApJS*, **129**:431. [`arXiv:astro-ph/9911219`].

Zehavi, I., et al. (2002): Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data. *ApJ*, **571**:172. [`arXiv:astro-ph/0106476`].

—— (2004): On Departures from a Power Law in the Galaxy Correlation Function. *ApJ*, **608**:16. [`arXiv:astro-ph/0301280`].

—— (2005a): The Intermediate-Scale Clustering of Luminous Red Galaxies. *ApJ*, **621**:22. [`arXiv:astro-ph/0411557`].

—— (2005b): The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, **630**:1. [`arXiv:astro-ph/0408569`].

—— (2011): Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *ApJ*, **736**:59. [`arXiv:1005.2413`].

Zheng, Z., Zehavi, I., Eisenstein, D.J., Weinberg, D.H., and Jing, Y.P. (2009): Halo Occupation Distribution Modeling of Clustering of Luminous Red Galaxies. *ApJ*, **707**:554. [`arXiv:0809.1868`].

Zucca, E., et al. (2009): The zCOSMOS survey: the role of the environment in the evolution of the luminosity function of different galaxy types. *A&A*, **508**:1217. [`arXiv:0909.4674`].

Zwicky, F. (1933): Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, **6**:110.

Zwicky, F., Herzog, E., and Wild, P. (1961): *Catalogue of galaxies and of clusters of galaxies, Vol. I* (California Institute of Technology, Pasadena).

List of Figures

List of Tables

Acknowledgements

En primer lloc, vull agrair als meus directors Vicent i Alberto l'oportunitat que em donaren per embarcar-me en aquest projecte, i tot el recolzament i l'esforç que m'han oferit durant aquests anys perquè aquesta tesi anara endavant. A més de que aprenguera (un poc) què és això de fer ciència, han aconseguit que puguera tindre el luxe (amb els alts i baixos normals) de disfrutar i d'il·lusionar-me amb allò què feia.

During this time, I have also worked with many people from whom I have learned a lot, and who deserve a large part of the merit for this thesis. I thank Enn Saar for many discussions and advices, and for his insight to address so many problems (or what I thought were problems). I also want to thank Jean-Luc Starck for giving me the opportunity to spend a few months in Saclay and, together with Nicolas Clerc and Antoine Labatie, for a fruitful (despite the difficulties) collaboration on wavelets and BAOs. I also spent four very interesting months in Hawai‘i thanks to the hospitality of István Szapudi and his group, specially Ben Granett. Finally, I would like to thank the whole Alhambra Team for taking me onboard the project, and for the joint effort on making it go forward. In particular, I thank Toño Montero and Alberto Molino for their collaboration on the design of the survey masks.

Durante estos años he estado trabajando con la "gran familia" de l'Observatori Astronòmic. Muchas gracias a todos los que habéis pasado por aquí (Alberto, Amelia, Elisa, Fernando, Javio, Jorge, Juan, Julia, Lara, Laura, Lupe, María, Mauro, Miquel, Pablo de la Cruz, Pascual, Sofía, Vicent, V. Peris, Xusa, y alguno que me dejaré) por el ambiente que habéis creado, por las risas en el café, por estar siempre dispuestos a echar una mano o resolver alguna duda, y por vuestra pasión por la astronomía.

Moltes gràcies també als meus amics. Als amics de la carrera (Paula, Pere, Xavi Alabau, Regi, Tito, Montse, Maria, Marina, Joel, ...) per què començàrem junts aquest viatge, i seguiu sent una referència per a mi. I per totes les paelles, sopars i discussions mítiques. I per les que en queden. I gràcies a Carles per saber simplificar les coses quan més ho necessites.

Gràcies a Lluís per tots els moments viscuts, per estar sempre ahí, i per tantes i tantes converses. I a la resta d'amics del Puig (Sento, Núria, Eva) i als meus Arquitectillos (Sile, Marta, Bea, Lola, Marién, . . .), moltes gràcies pels bons moments i per donar-me sempre ànims encara que no acabàreu d'entendre de què anava això de la tesi.

Gracias a Raquel por estar ahí todo este tiempo. Y por el apoyo y el cariño. Y por sufrir este último año de la tesis probablemente más que yo. Y por hacerme feliz.

Gracias a mi hermana Laura, por estar siempre presente, aunque estuviera a muchos kilómetros de distancia. Y gracias a mis padres por la confianza y el apoyo continuos.