UNIVERSITA' DEGLI STUDI ROMA TRE

Dipartimento di Matematica e Fisica

Scuola Dottorale in Scienze Matematiche e Fisiche
XXIX ciclo

Cosmological analysis of the angular correlation properties of 2MPZ galaxies

Tutor:
Prof. Enzo Branchini

Student:
Adriana Postiglione

Coordinator:
Prof. Roberto Raimondi

Anno Accademico 2016/2017

Contents

Introduction

A century has passed since Einstein (1916) published the Theory of General Relativity, and yet his theory provides the theoretical framework to understand our Universe. Today, a *concordance model* consisting in an Universe that is accelerating its expansion, approximately flat, in which matter contributes to about 1/3 of the total energy density budget, agrees with an impressively large number of very different, independent observations. About the 15% of the total matter content of the Universe seems to be in the form of non relativistic ordinary particles, called **baryonic matter**, while the remaining 85% is in the form of a dark component made of particles not predicted by the Standard Model of particle physics. This additional component, commonly referred to as **dark matter**, has a very small cross section with electromagnetic radiation, and dominates the local gravitational potential that governs the evolution of density fluctuations and the dynamics of virialized structures.

The existence of dark matter was first proposed by Zwicky (1933), after having analyzed the motion of the galaxies in the Coma cluster. The velocity dispersion within the cluster, he noted, implies a virial mass of the cluster well *above* that of the visible galaxies. About fifty years later, Rubin et al. (1980) discovered that the rotation velocity profiles of stars and gas in spiral galaxies didn't follow the model expected by gravity theory: only the existence of an extra component, not interacting electromagnetically but only gravitationally, could explain the observed discrepancy between model and data. In more recent years, the existence of yet another

dark component, called **dark energy**, was introduced in order to explain a number of additional cosmological observations. Firstly, the discovery of cosmic acceleration, confirmed by two independent studies about the relation between distance of Supernovae Ia type and redshift (Riess et al. (1998) and Perlmutter et al. (1999)). Then, independent evidences came from the abundance of clusters of galaxies, the spatial correlation properties of the Large Scale Structure and the temperature fluctuations in the Cosmic Microwave Background. If the dark matter dominates the total amount of matter in the Universe, the dark energy dominates the total energy budget, and is responsible for the accelerated expansion of the Universe. Actually, dark energy is not the only possible explanation to cosmic acceleration. It is also possible that the General Relativity ceases to be valid on cosmological scales, so that a modified theory of gravity is needed. At present, neither the nature of dark matter nor the one of dark energy is clear, and since the existence of these components has deep implications for fundamental physics, it is not surprising that an increasing fraction of the scientific community is putting a large effort in addressing these problems.

In this picture, the study of **Large Scale Structure** of the Universe, i.e. the spatial distribution of matter on cosmological scales, plays a crucial role. For this reason, in the last decades a significant number of *galaxy surveys* have been carried out to trace the spatial distribution of extragalactic objects over an ever increasing fraction of the observable Universe. Over times different redshift surveys have been performed. They were tailored to address specific scientific questions. So, for example, the 2MASS redshift survey (Huchra et al., 2012) aimed at exploring the nearby Universe is rather local but has a very large sky coverage. The more recent SDSS redshift survey (Abazajian et al., 2009) and its BOSS (Dawson et al., 2013) follow-up have been designed to encompass regions large enough to unambiguously detect the barionic acoustic oscillations imprinted in the spa-

tial distribution of galaxies (Eisenstein et al., 2005). This was also the main goal of the large but sparse WiggleZ survey (Blake et al., 2011) in contrast with deeper and denser surveys (VIPERS, Guzzo et al. (2014)) designed to both trace the large scale structures and investigate the properties of the observed galaxies. Next generation surveys like the one that will be carried out by the Euclid satellite mission (Laureijs et al., 2011), will push the limits even further by tracing the large scale structure over a significant fraction of the observable Universe, effectively probing the transition epoch in which the Universe has begun its accelerated expansion.

The surveys mentioned here are examples of *spectroscopic redshift surveys*, in which the redshift of a galaxy is used as a proxy to its distance and measured through spectroscopy, i.e. by measuring the wavelenght in the absorption or emission lines in the energy spectrum of extragalactic objects. A second, less precise proxy to galaxy distance is the **photometric redshift**. In this case the redshift is estimated by measuring the magnitude of the objects in a number of different energy bands. This allows one to sample the energy spectrum and infer the redshift of the objects. Photometric redshift are less precise than spectroscopic one but can be measured for a comparatively much large number of objects. The radial distance can therefore be estimated less precisely but for many more galaxies, allowing one to investigate galaxy clustering using a tomographic approach that consists in analyzing the angular correlation properties of galaxies in different radial shells.

In fact, assuming that galaxies trace the distribution of matter, most of which is dark, redshift surveys have shown that the galaxies and the underlying mass are not randomly distributed in the Universe, but instead organized in a coherent pattern of large scale structures called the *cosmic web*. This pattern is believed to be the result of the evolution of primordial perturbations on the early quasi-homogeneous and isotropic density field under the effect of gravity. As we will see in the following sections,

the dark matter density fluctuations above a characteristic scale, the *Jeans length*, started to grow by gravitational instability until they formed the cosmic structure we observe in our Universe today. On the other hand, the baryonic component, after decoupling itself from radiation, started to fall into the dark matter potential wells, until tracing the same large scale structures as the dark matter. This scenario has been confirmed by the detection of the temperature fluctuations in the Cosmic Microwave Background (COBE, Smoot et al. (1992)) and the Baryonic Acoustic Oscillation (BAO, Eisenstein et al. (2005)).

The evolution of structures can be well described by the theory using a perturbative approach in the so-called quasi-linear regime, that remains valid as long as perturbations are small, which occurs at early epochs or when averaging over large scales. On small scales, when density fluctuations become large, the linear approximation is not valid anymore and numerical approaches are needed to explore the full non-linear evolution. This can be done by means of N-body simulations. Irrespective of the approach used, whether analytic or numerical, the capability to predict the statistical proprieties of the mass distribution and their relation with the fundamental parameters of the model (the so-called *cosmological parameters*) is certainly one of the biggest successes of theoretical cosmology. For this reason, *observational* estimates of such statistical proprieties are crucial to constrain the properties of our universe and estimate the cosmological parameter that characterize its best fitting model.

2-point statistics, either in Fourier or in configuration space, represent the most important tool in the analysis of the large scale structure. In particular, when dealing with photometric redshift surveys, the angular power spectrum C_ℓ is the most widely used statistical tool. It represents the Legendre transform of the angular correlation function, which, in turn, expresses the probability of finding two galaxies separated by an angle θ. Because of the large photometric redshift errors, clustering along the radial

directions is preserved only at very large separations. For this reason, the best way of characterizing the large scale structures is by measuring the angular 2-point statistics within spherical shells. Slicing the sample along the radial direction is possible if photometric redshifts are available, hence the idea to use the aforementioned tomographic approach when dealing with photo-z surveys.

The aim of this thesis is to measure, for the first time, the angular power spectrum of the 2MPZ galaxy catalogue (Bilicki et al., 2014) in three different redshift slices, and estimate the dark matter density and the baryon fraction as well as the characteristic amplitude of the density fluctuations in the sample.

For this purpose, we develop a C++ code that allows us to measure the angular spectrum of the galaxy distribution over some fraction of the celestial sphere. In the code we implement the estimator introduced by Peebles (1973), who was among the first to tackle the problem of obtaining an unbiased estimator for the angular multipoles C_ℓ in the realistic case of a partial sky coverage. Not having a full-sky survey leads to a mixing between the multipoles of the spectrum, that needs to be taken into account when comparing observations with model predictions. This is done by convolving the latter by the so-called **mixing matrix**, also estimated by the C++ code used to compute the angular spectrum. The correlation between modes induced by a partial sky coverage (and by the non-linear evolution) prevents the estimation of the error to be done analitically. For this reason, several numerical methods like the *Jack-knife* method (Tukey, 1958) and the *Bootstrap* (Efron, 1979) technique, have been developed. Both approaches, however, are approximated as they estimate errors by rearranging the objects of the very same datasets one wants to analyze. For this reason, in our work we estimated errors using a suite of 2MPZ mock catalogues.

The creation of many independent and realistic mock photometric-redshift catalogs is one of the main goals of this thesis. This is indeed

a computationally challenging task. Fortunately, for the type of analysis we wish to perform, we can make several approximations. Indeed, the main sources of errors here are the partial source coverage and the large photo-z errors. The partial source coverage can be accurately mimicked in simplified mock catalogs that can be obtained by applying a lognormal transformation of a Gaussian fluctuation field sampled with Monte Carlo techniques. We developed an automated procedure to build these mocks that we used to quantify both systematic and random errors and their covariance and to optimize the strategy to estimate the cosmological parameters. The photo-z errors can be estimated modeling the true spectroscopic redshift distribution of galaxies. Equipped with this knowledge, we estimate the angular power spectrum of the galaxies in the 2MPZ catalogue (Bilicki et al., 2014) and use a Markov Chain Monte Carlo likelihood analysis to estimate the mean dark matter density, the spectral amplitude, the baryon fraction and their errors.

The structure of the thesis is as follows.
In Chapter 1 and 2 we introduce the key elements of the Standard Cosmological Model, starting from the Einstein's equations and the gravitational instability theory. Then, we introduce the concept of power spectrum of matter, analyze its dependence on the cosmological parameters and present its angular counterpart, the angular power spectrum. At the end, we face the issue of measuring 2-point statistics in real datasets consisting of spectroscopic and photometric galaxy redshift surveys.

In Chapter 3 we extend the formalism introduced in Chapter 2 to account for incomplete sky coverage. We introduce the angular power spectrum estimator we use in this thesis and the numerical code we developed to estimate it. We conclude by presenting the results of sanity checks and performance tests.

In Chapter 4 we present the galaxy catalogue that we analyze in this thesis, namely the 2MASS Photometric Catalogue (2MPZ, Bilicki et al.

(2014)). We describe the different photometric catalogues that have been cross-matched in order to obtain the final 2MPZ one, and then we compute the redshift distribution of the photo-z galaxies and of spectroscopic subsamples.

The Chapter 5 is dedicated to the creation of lognormal fake catalogues that mimic the properties of 2MPZ. We analyze their properties and estimate their angular spectra and their uncertainties and use these results to compute the covariance matrix that is used to estimate the covariant errors in the real data analysis. We also estimate the true redshift distribution, $dN/dz_{\rm spec}$, from the observed $dN/dz_{\rm phot}$, in order to correct for the photometric redshift errors.

The Chapter 6 is devoted to the methods used to extract cosmological information by comparing the measured spectra in different redshift shells with theoretical predictions. This is done in two ways: a standard χ^2 approach and a MCMC method. In the second part of the chapter we apply these statistical tools to the mock datasets to assess the goodness of the methods, to reveal potential sources of systematic errors and to highlight potential parameter degeneracies.

In Chapter 7 we focus on the final analysis, providing our estimation of some of the cosmological parameters from the 2MPZ catalogue.

In Chapter 8 we discuss the results and summarize the conclusion of this thesis.

Chapter 1

Background Cosmology

In this chapter we present the key elements of the Standard Cosmological Model, that represents the reference model of this work. After introducing the Cosmological Principle and the Einstein's equations, we will focus on the gravitational instability theory. Then, we will introduce the concept of power spectrum of the matter and analyze the information it encodes. From this, we derive the angular power spectrum, which represents the main subject of this thesis. At the end we will focus on the issues that arise when measuring the 2-point statistics from the galaxy redshift surveys.

1.1 Basic concepts and equations

The fundamental principle of the Standard Cosmological Model is the so-called **Cosmological Principle**, which states that the Universe is *isotropic* (i.e. rotational invariant) and *homogeneous* (i.e. traslational invariant). However, as the modern galaxy redshift surveys - like 2dF and SDSS - have proved, this hypothesis is only valid on *sufficiently large* scales, i.e. on scales larger than the scale of coherence in the spatial distribution of galax-

ies (more than hundreds of Mpc[1] In other words, the distribution of galaxies appears to be homogeneous and isotropic when smoothed on these scales. In this sense, each component of the Universe (matter, radiation etc) can be thought of an ideal fluid of non-interacting particles characterized by an energy (or mass) density, an equation of state and a temperature that don't vary across space, i.e. that have no spatial fluctuations. The ith component will have a pressure $p = p_i$ and energy density $\rho = \rho_i$, such that the total pressure and energy density will be:

$$\begin{aligned} p &= \sum_i p_i \\ \rho &= \sum_i \rho_i \end{aligned} \tag{1.1}$$

For a Universe subject to the Cosmological Principle, there is only one allowed metric, the *Friedmann-Robertson-Walker* one:

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu = -c^2 dt^2 + a^2(t) \left[dr^2 + S_k(r)^2 (d\theta^2 + sin^2(\theta) d\phi^2) \right] \tag{1.2}$$

where the Greek indices μ and ν vary from 0 to 3, with the 0 index related to the time coordinate ($dt = dx^0$), while the $1, 2, 3$ indices are related to the spatial coordinates. The term $g_{\mu\nu}$ is the *metric tensor*, while c represents the speed of light and t the cosmological *proper time*. The coordinates r, θ and ϕ are *spatial comoving coordinates*, i.e. they don't vary over time, whereas the physical distances do vary because of the **scale factor** $a(t)$, which can increase or decrease over time, meaning expansion or contraction (these last two being however homogeneous and isotropic, so that the Cosmological Principle stay valid over cosmic epocs). On the right side of **Eq. 1.2**, the term $S_k(r)$ is equal to $Rsin(r/R)$ in case of positive curvature (*close* Universe), r in case of null curvature (*flat* Universe), $Rsinh(r/R)$ in case of negative curvature (*open* Universe - see **Fig. 1.1**). Here R is the *curvature*

[1]The parsec is a unit of distance, which corresponds to 3.086×10^{13} km. The Megaparsec (Mpc) corresponds to 10^6 parsecs.

radius, while the subscript k refers to the *curvature parameter*, $k = k'/R^2$, where k' can be equal to $-1, 0, 1$ reflecting the sign of the curvature respectively for a close, flat and open geometry. While the comoving coordinates are adimensional, the scale factor has the dimension of a lenght; its value at present epoch, $a(t_0)$, is conventionally set to 1. One can also define the curvature parameter k, so that **Eq. 1.2** becomes:

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{dx^2}{1-kx^2} + x^2(d\theta^2 + sin^2(\theta)d\phi^2)\right] \tag{1.3}$$

with $x = S_k(r)$.

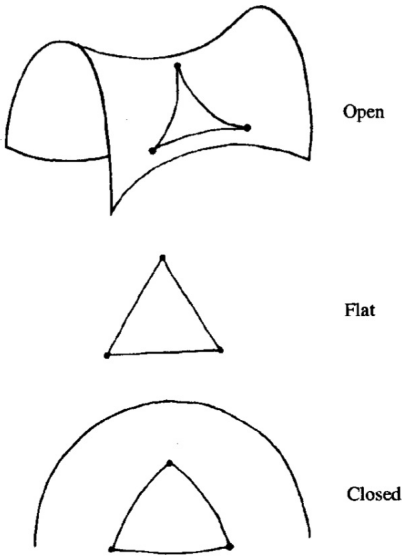

Figure 1.1: Examples of curved spaces in two dimensions. Figure from Coles and Lucchin (1995).

If the metric holds the information about a particular space-time geometry, the link between this latter and the energy content of the Universe is expressed, in general, by the *Einstein's field equations*:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{1.4}$$

where the *Einstein tensor* $G_{\mu\nu}$ describes the geometry; $R_{\mu\nu}$ is the *Ricci tensor*, which depends on the metric and its derivative; R is the *Ricci scalar*,

obtained contracting the Ricci tensor: $R = g^{\mu\nu}R_{\mu\nu}$; G is the gravitational constant, while $T_{\mu\nu}$ is the *energy-momentum tensor*, that describes the energy content of the Universe. Since we have assumed the Cosmological Principle, on large scales the energy-momentum tensor becomes the one of a perfect fluid:

$$T_{\mu\nu} = (p + \rho)\frac{u_\mu u_\nu}{c^2} + p g_{\mu\nu} \tag{1.5}$$

Here u_μ is the quadrivelocity of the fluid, ρ is its energy density and p its pressure. Considering the 00 tensor component of the Einstein's equations and the FRW metric, one can derive the *first Friedmann equation*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\sum_i \rho_i - \frac{kc^2}{a^2} \tag{1.6}$$

here the index i refers to the ith component of the Universe, such that the sum $\sum_i \rho_i$, as we have seen, represents the mean total energy density in the Universe, as the **Eq. 1.1** shows. The left-hand side of the **Eq. 1.6** is usually referred to as the *Hubble parameter*:

$$H = \frac{\dot{a}}{a} \tag{1.7}$$

which measures how rapidly the scale factor changes. The present epoch value of the Hubble parameter is called *Hubble constant* $H_0 \equiv H(t_0)$. Through the Hubble parameter one can re-write the energy density ρ_i for the ith component and define the **density parameter**:

$$\Omega_i \equiv \frac{\rho_i}{\rho_c} = \frac{8\pi G\rho_i}{3c^2H^2} \tag{1.8}$$

where ρ_c is the **critical density**:

$$\rho_c = \frac{3c^2H^2}{8\pi G}, \tag{1.9}$$

defined such that the **total density parameter**, defined as follow:

$$\Omega(t) = \sum_i \Omega_i(t), \tag{1.10}$$

is equal to 1 when the total mean density is equal to the critical one. One can also define the **curvature parameter**:

$$\Omega_k = -\frac{kc^2}{a^2H^2} \tag{1.11}$$

With these definitions the **Eq. 1.6** becomes:

$$1 = \Omega + \Omega_k \tag{1.12}$$

From the **Eq. 1.12** it is clear that the values of Ω and Ω_k are not independent: if $\Omega < 1$ then $k < 0$ and the Universe is open, if $\Omega = 1$ then $k = 0$ and the Universe is flat and if $\Omega > 1$ then $k > 0$ and the Universe is closed. The first Friedmann equation relates the energy density with the Hubble parameter and the curvature. But if one wants information about the time evolution of the scale factor another equation is needed. Setting to 0 the covariant derivative of the energy-momentum tensor, one can obtain the conservation equation for the ith component:

$$\dot{\rho}_i + 3\frac{\dot{a}}{a}(\rho_i + p_i) = 0 \tag{1.13}$$

The **Eq. 1.6** and the **Eq. 1.13** can be combined to give another important equation, the *acceleration equation* (often called *second Friedmann equation*):

$$\frac{\ddot{a}}{a} = \frac{4\pi G}{3c^2}\sum_i(\rho_i + 3p_i) \tag{1.14}$$

where again the sum is performed over all the components of the Universe. Note that the ordinary components, characterized by a positive pressure, like radiation and non relativistic matter, determine a decelerated expansion ($\ddot{a} < 0$). The only way to have an accelerated expansion is to have a dominant component with negative pressure ($p_i < -(1/3)\rho_i$, valid if the **Eq. 1.1** holds).

Equations **1.6** and **1.13** can be both solved once is specified the *equation of state* of the i-th component, which relates the pressure with the energy density. In many cases of physical interest the equation of state is written in the following form:

$$p_i = \omega_i\rho_i \tag{1.15}$$

with ω_i adimensional parameter. Here we have used natural units in which $c = 1$. Substituting this relation in **Eq. 1.13** one can obtain the energy density as a function of the scale factor (for constant ω_i):

$$\rho_{\omega_i} = \rho_{\omega_i,0} a^{-3(1+\omega_i)} \tag{1.16}$$

where $\rho_{\omega,0}$ is the present day energy density. From this equation one can describe the evolution of each component of the Universe. For example the non-relativistic matter, simply referred as *matter*, is characterized by a parameter $\omega \sim 0$, therefore, according to **Eq. 1.16**, matter density evolves as a^{-3}, while the relativistic matter, the *radiation*, has $\omega \sim 1/3$, so it evolves as a^{-4}. This means that in the limit $a \to 0$ the component with the higher value of ω dominates the energy-density budget, while for $a \to \infty$ is the component with the lower value of ω that dominates. However, this behavior does not explicit the time dependence of matter or radiation density. To specify it, one has to know the time evolution of the scale factor $a(t)$, which constitutes the subject of the next section.

1.1.1 The expanding Universe

In 1929 Edwin Hubble (Hubble, 1929), measuring the redshift z of known lines in the spectra of external galaxies, discovered that they receed from us with a velocity proportional to their distance. Such experimental law agrees with the hypothesis of a Universe that is expanding in a homogeneous and isotropic way, i.e. is fully consistent with the FRW metrics, where, as we have seen, a positive evolution of scale factor would determine a homogeneous and isotropic expansion of the system. This evidence also supports the Cosmological Principle, and suggests the existence of a Big Bang, i.e. that in the past the Universe was in a very dense state. As time went by, the temperature associated to each component (which was a single temperature when these components were coupled) decreased, and this means that, according to this model, there was an epoch in which the physical state of the Universe was similar to that of a stellar interior, and

a *cosmological* Nucleosynthesis could have happened. During this epoch, non negligible quantities of light elements would have been produced (especially He and D), and we have indeed observed them at high redshifts in the intergalactic medium, providing an additional observational support to the standard cosmological model. Furthermore, observational evidence is also provided by the *Cosmic Microwave Background* (CMB) radiation. At high density and temperature, baryons and photons are highly coupled. When the temperature dropped enough, hydrogen become neutral and atoms decoupled from radiations. This happens at redshift $z \sim 1100$ (that corresponds to about 3×10^5 years after the Big Bang): the baryons decouple from photons, the Universe becomes transparent to radiation that, because of the adiabatic cooling, we observe now as the much colder CMB radiation.

Today, the most commonly accepted model for the present day Universe is an expanding, approximately flat, accelerating ($\ddot{a} > 0$) one, as suggested by a variety of cosmological observations. In the total energy density budget at the present epoch, the *matter* contributes to $\sim 1/3$ of the total, of which about the 15% is ordinary matter (*baryonic* matter), while about the 85% is *cold dark matter*, a kind of collisionless, non-relativistic matter which interact only gravitationally but not electromagnetically. A small contribution (less then 10^{-4}%) is due to the relativistic matter (photons and neutrinos), generically referred to as *radiation*. Finally, the remaining $\sim 2/3$ of the total energy density is believed to be due to *dark energy*, a component with negative pressure responsible for acceleration of the expansion (see **Eq. 1.14** with $\omega < -1/3$). The simplest model of dark energy is the *cosmological constant* Λ, characterized by a constant energy density (ρ_Λ) and an equation of state parameter $\omega = -1$. The cosmological model, with a cosmological constant and a cold dark matter, is called ΛCDM. **Table 1.1** shows the values of the cosmological parameters of this model, in units of density parameter Ω (the subscript b refers to baryons, the CDM to the dark matter, M to the total matter and Λ to the cosmolog-

ical constant); h is the Hubble constant in units of 100 Km s^{-1} Mpc^{-1}.

Table 1.1: The measured values of the cosmological parameters and their 1σ errors obtained from the analysis of the CMB maps assuming a ΛCDM model (Planck Collaboration et al., 2014).

	best fit and 68% limits
H_0	67.4 ± 1.4
$\Omega_b h^2$	0.02207 ± 0.00033
$\Omega_{CDM} h^2$	0.1196 ± 0.0031
Ω_M	0.314 ± 0.020
Ω_Λ	0.686 ± 0.020
A_s	$(2.142 \pm 0.049) \times 10^{-9}$

Now that we have specified a reference cosmological model, we can have a better insight on **Eq. 1.16**. Since Λ has $\omega = -1$, its energy density remains constant with time. Moreover, since the Universe is expanding, and according to what we have seen in the last part of the previous section, $a \to 0$, i.e. at early times the energy budget is dominated by the radiation. On the contrary, when $a \to \infty$ (late times) is the dark energy to be dominant. In between, the Universe has been dominated by the non-relativistic matter, including both baryons and dark matter (see **Fig. 1.2**). The epoch at which the energy density of radiation and that of baryonic matter are equal is called *equivalence*, and occurs at $a_{eq} \cong 2.8 \times 10^{-4}$ (approximately 47000 years after Big Bang), as can be derived from the present day values of energy densities of matter and radiation. In fact, recalling that from **Eq. 1.16** we deduced that $\rho_r = \rho_r(t_0)a^{-4}$ and $\rho_m = \rho_m(t_0)a^{-3}$, one can write:

$$a_{eq} = \frac{\rho_r(t_0)}{\rho_m(t_0)} \tag{1.17}$$

We can now finally understand the time evolution of the scale factor $a(t)$ using **Eq. 1.6** and **Eq. 1.16**: during the radiation epoch $a \propto t^{1/2}$, during the

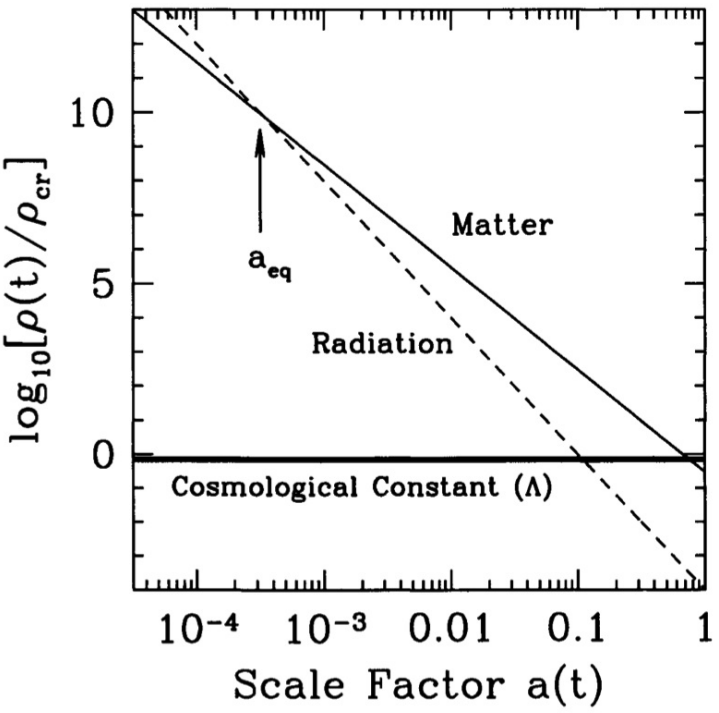

Figure 1.2: Energy density as a function of the scale factor for the different components of a flat Universe: matter, radiation, and a cosmological constant. In each epoch the evolution of the scale factor with time is driven by the dominant component. The energy density is expressed in units of the critical density today. The epoch in which the content in matter and radiation was equal is called *equivalence* a_{eq} (see the text). Figure from Dodelson (2003).

matter domination epoch $a \propto t^{2/3}$ while, when the a cosmological constant dominates the expansion, the dependence is exponential: $a \propto \exp(Ht)$.

1.1.2 Redshift and distances

If the Universe is expanding in a homogeneous and isotropic way, as we have just seen, then the distance between an observer and a generic space-time location, ideally measured instantaneously (the so-called *proper distance*), increases with time because of the expansion. This distance can be

computed from **Eq. 1.3**; if the observer is at the origin of the comoving coordinates (r, θ, ϕ), then:

$$d(t) = a(t) \int_0^r \frac{dx}{\sqrt{1-kx}} \equiv a F_k(r) \tag{1.18}$$

with r comoving radial coordinate of the object. Suppose now that at the time t_e a light signal is emitted at the radial comoving coordinate r and is observed at the center of coordinates at the time t_0. Since in a time dt light travels a *comoving distance* $dx/(1-kx^2) = cdt/a$ (see **Eq. 1.2**) then the total comoving distance light could have traveled (in the absence of interactions) is:

$$\int_{t_e}^{t_0} \frac{c\, dt}{a(t)} \tag{1.19}$$

At t_0, proper and comoving distance coincide:

$$d(t_0) = F_k(r) = \int_{t_e}^{t_0} \frac{c\, dt}{a(t)} \tag{1.20}$$

Since the scale factor at the time of emission t_e can be expressed in terms of the redshift of the light emitted:

$$1 + z = \frac{1}{a(t_e)} \tag{1.21}$$

one can write the comoving distance as a function of redshift, Hubble parameter and density parameter using **Eq. 1.6**:

$$d(z) = \frac{c}{H_0} \int_0^z X(z')^{-1/2} dz' \tag{1.22}$$

where X is the generic solution of the Friedmann equations:

$$\begin{aligned} X(z') = \Omega_R (1+z')^4 + \Omega_M (1+z')^3 + \Omega_K (1+z')^2 + \\ + \Omega_\Lambda \exp\left\{ -3 \int_0^{z'} \frac{dz''}{z''} \left[1 + \omega_\Lambda(z'') \right] \right\} \end{aligned} \tag{1.23}$$

Here ω_Λ is the state parameter of the dark energy component assumed to be time-independent. **Eq. 1.23** assumes the existence of a Dark Energy

that generalizes the concept of cosmological constant to the case of time-varying equation of state of a fluid with negative pressure. **Eq. 1.22** shows that the redshift-distance relation depends on the density parameters and the equation of state of the various components of the Universe, while the Hubble parameter H_0 sets the time and length scales of the Universe. This means that, ideally, all cosmological parameters can be derived from measuring comoving distances of objects at various epochs. Unfortunately, the comoving distance cannot be measured experimentally since it would requires performing instantaneous measurements. However, we can define other type of distances that can be measured and that are related to the comoving proper distance by simple relations: the *luminosity distance* d_L and the *angular diameter distance* D.

Forgetting for a moment the expansion of the Universe, the flux F observed at a distance d from a source of known intrinsic luminosity L is $L/4\pi d^2$ since the total luminosity through a spherical shell with area $4\pi d^2$ is constant. In an expanding Universe is worth working with the comoving grid with the source placed in the origin, so that one can define the luminosity distance d_L:

$$d_L \equiv \sqrt{\frac{L}{4\pi F}} = S_k(r)(1+z) \tag{1.24}$$

where d here is now the comoving distance expressed by **Eq. 1.22**.

Analogously, one can relate the distance of an object with the angle θ subtended by it knowing its physical size D'. In this case one defines the angular diameter distance D:

$$D(z) = \frac{D'}{\theta} = \frac{S_k(r)}{1+z} \tag{1.25}$$

Note that in the limit $z \to 0$, i.e. small distances from the observer, the three distances coincide:

$$d(z) \approx d_L \approx D(z) \approx \frac{c}{H_0} z \tag{1.26}$$

1.2 Gravitational instability

As we have seen, the Cosmological Principle is valid only on large scales: galaxies, clusters, superclusters, filaments and voids (the *cosmic web*) break the homogeneity and isotropy on increasingly large scales as the Universe evolves. The formation of these structures can be explained as the result of the growth of initial perturbation of the smooth background.

The basic mechanism of this growth is the gravitational instability that sets in when the physical scale of the fluctuations is beyond a characteristic scale called *Jeans length*. The same mechanism can explain the formation of stars in galaxies, but in a cosmological framework things are complicated by the coupling of matter perturbations with perturbations of the other components of the Universe and of the metric itself. Fortunately, the theory of gravitational instability is greatly simplified if one considers small fluctuations: in this case, the equations can be linearized. Indeed, the fluctuations *were* small at early epochs and still are on large scales. On small scales, however, the linear perturbation theory is not valid anymore and higher order perturbation theory or numerical simulations are needed in order to describe the evolution of the density contrast in the non-linear regime. Here we focus on linear perturbation theory. For simplicity we set $\hbar = c = k_B = 1$.

Starting from the FRW metric (**Eq. 1.2**) we can linearly perturb the metric tensor:

$$\begin{aligned} g_{00}(\vec{x}, t) &= -1 - 2\psi(\vec{x}, t) \\ g_{0i}(\vec{x}, t) &= 0 \\ g_{ij}(\vec{x}, t) &= a^2 \delta_{ij}(1 + 2\phi(\vec{x}), t) \end{aligned} \tag{1.27}$$

The perturbation to the metric are ψ, the *Newtonian potential*, and ϕ, the *perturbation to the spatial curvature*; since they are assumed to be small, all quadratic terms have been neglected so that one has only the linear perturbations to the metric. Note that one can classify perturbations in three

categories: scalar, vector and tensor perturbation. In **Eq. 1.27** only the scalar perturbations have been included. The reason is that while vector and tensor perturbations may also exist (tensor perturbations describe gravitational waves), only the scalar ones couple to perturbation in the matter component. Since we are interested in cosmic structures and their evolution, we shall focus on the latter. Moreover, these equations are written in *conformal Newtonian gauge* that corresponds to set to 0 two of the four scalar degrees of freedom: the remaining two are ψ and ϕ. Finally, in the absence of expansion (i.e. $a = 1$) this metric describes a weak gravitational field.

We will not go into details, but the **Eq. 1.27** allows us to derive, from Einstein's field equations, two independent equations analogous to the Friedmann equation. These equations, together with the Boltzmann equations, form the differential equations system that describes the evolution of the fundamental variables used to model the linear perturbation theory. These variables are the following: the temperature fluctuation of the photon fluid $\Theta = \delta T/T$, decomposed in its angular multipole components ${\Theta_l}^2$; the distribution function for massless neutrinos ν; the overdensity field of dark matter $\delta(\vec{x}, t) = \rho(\vec{x}, t)/\bar{\rho}(\vec{x}, t) - 1$; the peculiar velocity of dark matter $v(\vec{x}, t)$ and the analogous for baryons $\delta_b(\vec{x}, t) = \rho_b(\vec{x}, t)/\bar{\rho_b}(\vec{x}, t) - 1$ and $v_b(\vec{x}, t)$.

Let us now focus on dark matter perturbations. In principle, the evolution equations for them are the full set of Boltzmann *and* Einstein equations (see Dodelson (2003) for details). In practice, the full set of equations is not needed. First of all, since the baryon mass density is negligible with respect to that of the dark matter, their contribution to fluctuations of the

[2]We will not go into details, but it is worth noting that the quantity $\delta T/T$ can be only measured in correspondence to the last scattering surface, i.e. the period in which the light decoupled with matter and started to travel freely.

metric can be neglected[3]. Moreover, for the purposes of the matter distribution, the photon are relevant only before the decoupling, i.e. during the radiation dominated epoch. During this epoch photons were coupled with baryons, and scattering Compton was so efficient that the photon distribution can be characterized by only two moments, the monopole Θ_0 and the dipole Θ_1: all other moments are suppressed. It can be shown (Dodelson, 2003) that if there are no quadrupole moments $\psi \to -\phi$. Under the above approximations the four relevant Boltzmann equations and the time-time component of the Einstein equations become[4]:

$$\dot{\Theta}_{r,0} + k\Theta_{r,1} = -\dot{\phi} \tag{1.28}$$

$$\dot{\Theta}_{r,1} + \frac{k}{3}\Theta_{r,0} = -\frac{k}{3}\phi \tag{1.29}$$

$$\dot{\delta} + ikv = -3\dot{\phi} \tag{1.30}$$

$$\dot{v} + \frac{\dot{a}}{a}v = ik\phi \tag{1.31}$$

$$k^2\phi + 3\frac{\dot{a}}{a}\left(\dot{\phi} + \phi\frac{\dot{a}}{a}\right) = 4\pi G a^2\left[\rho_{CDM}\delta + 4\rho_R\Theta_{r,0}\right] \tag{1.32}$$

Here the subscript R refers to radiation, including both neutrinos and photons[5], while the subscripts 0 and 1 indicate the monopole and dipole moments of the temperature fluctuations in the CMB map. These equations are written in Fourier space: in fact, in linear theory the Fourier transform produces a set of independent ordinary differential equations that can be solved for each k-mode separately. In our case the Fourier transform con-

[3]This, as we have seen, is an approximation, since the baryons represent the 15% of the total mass budget. We nevertheless assume that $\Omega_b \ll \Omega_{CDM}$ here, but keep in mind that the role of baryon cannot be neglected if one wishes to account for certain features like Baryonic Acoustic Oscillation in galaxy cluster proprieties.

[4]Note that while Θ is defined as a 3D quantity, Θ_r is a 2D quantity, with a fixed r.

[5]It is not obvious that both species follow the same evolution equations, but this is true at least in the limit of small baryon density, and only for the purposes of following the matter evolution, as we are doing here.

vention is the following:

$$\tilde{\Theta}(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}}\, \Theta(\vec{k}) \tag{1.33}$$

Note that all the variables, Θ, δ, v and ϕ are function of the *comoving wavenumber* $\vec{k}$ and the *conformal time* η, defined as:

$$\eta \equiv \int_0^t \frac{dt'}{a(t')} \tag{1.34}$$

η is used as a time variable because it's monotonically increasing with times. It also coincides (for $c = 1$) with the *comoving horizon*, i.e. the comoving distance that a photon could have traveled from the Big Bang to the time t (recall **Eq. 1.19**). In other words, no information could have propagated further than η since the beginning of time: regions separated by distance *greater* than η are not casually connected. The derivatives in the **Eqs. 1.28 - 1.32** have to be estimated respect to the conformal time.

There are no analitic solutions for the system of equations **1.28 - 1.32** valid on all scales at all times. However, it's possible to make approximations and derive solutions valid only for certain scales at certain times.

1.2.1 Solutions to the linear perturbation theory

As the comoving horizon defined by the **Eq. 1.34** grows with time, it incorporates larger and larger comoving scales, which become casually connected.

Let's first focus on large scale modes, i.e. the wavenumbers k which enter the horizon at late times, well after the matter-radiation equivalence. These scales enter the horizon during the matter dominated epoch, while lying beyond the horizon during radiation dominated epoch. As long as the modes are outside the horizon, they are not casually connected, the gravitational potential[6] ϕ remains approximately constant and the density

[6]The perturbation in the curvature ϕ is equal to the gravitational potential until the scale stays much smaller than the horizon and the Newtonian limit remains valid.

perturbations of dark matter δ are constant too. These modes enter the horizon during the matter dominated epoch, when the density fluctuations are small and the Universe is approximately flat (i.e. *Einstein-de Sitter*). In this epoch the time derivative of the gravitational potential is zero and it can be shown that the density fluctuations within the horizon have the following behavior as a function of $\vec{k}$ and a:

$$\delta(\vec{k}, a) = \frac{k^2 \phi(\vec{k}, a) a}{(3/2)\Omega_M H_0^2} \tag{1.35}$$

Note that $\dot{\phi} = 0$ implies that the density perturbations grow proportionally to the scale factor only when matter dominates: when dark energy dominates this is not true anymore, because the amplitude of the gravitational potential decreases ($\dot{\phi} < 0$).

The small scales are the one that cross the horizon during the radiation era. In this case the solution of the **Eq.1.28 - 1.32** for the dark matter density fluctuations is:

$$\delta(\vec{k}, \vec{\eta}) \propto \ln(\vec{k}, \eta) \tag{1.36}$$

In fact, using the **Eq. 1.34** together with the fact that, as we have already seen, $a \propto t^{1/2}$ during the radiation dominated epoch, it can be shown that during this period $\eta = a$. In other words, if the mode enters the horizon during the radiation era, the dark matter perturbations grow logarithmically with the scale factor because of the radiation pressure. Later on, when the radiation perturbations become negligible, the equation that governs the evolution of the dark matter perturbation can be derived again from **Eq.1.28 - 1.32** and has two independent solutions: a growing solution $D_1(\eta)$, called *growth factor*, and a decaying solution $D_2(\eta)$. The general solution for dark matter perturbations will be a linear combination of the two:

$$\delta(\vec{k}, \eta) = C_1 D_1(\eta) + C_2 D_2(\eta) \tag{1.37}$$

where C_1 and C_2 depend on $\vec{k}$ and are set from the initial conditions. At late times, when matter starts to dominate, the growing solution scales with a

while the decaying mode vanishes. This means that, after a certain period of time, fluctuations on both large and small scales grow linearly with a. Consequently, the evolution of the matter perturbations on all the modes of cosmological interest become *scale indipendent* at late times, and can be described by a single time dependent factor: the growth factor D_1. At even later times ($a > 0.1$ and $z < 10$), when dark energy starts to dominate and the Universe experiences an accelerating expansion, the growth of fluctuations slows down. The precise scale time dependence in this case depends on the DE equation of state.

1.2.2 The growth factor

As we have just seen, in the matter dominated epoch all modes have experienced the same growth factor. In practice, this is the result of the fact that if the dark matter is cold, which we assume here, it has zero pressure: once a mode enters the horizon, there is no way for pressure to smooth out the inhomogeneities and all modes evolve identically. The expression of D_1 as a function of a is therefore particularly relevant. It can be obtained solving the equations **Eq.1.28 - 1.32**. Here we show the solution that includes the possibility of having energy other than matter and radiation:

$$D_1(a) = \frac{5\Omega_M}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{(a'H(a')/H_0)^3} \tag{1.38}$$

where $H(a)$ is given by the **Eq. 1.23**. **Fig. 1.3** shows the behavior of the growth factor for different cosmologies. For a matter-dominated Universe without cosmological constant (solid top line) the growth factor is equal to the scale factor (see **Eq.1.6** with $\Omega_M = 1$, $k = 0$ and matter domination). If we include a cosmological constant ((long dashed curve) the growth is suppressed at late times ($z \simeq 2$) because of the dark energy. Moreover, if we consider an open Universe (short dashed curve) the suppression acts over all cosmic times.

We can then draw the following qualitative conclusion: structure in an open or dark energy dominated Universe developed much earlier than

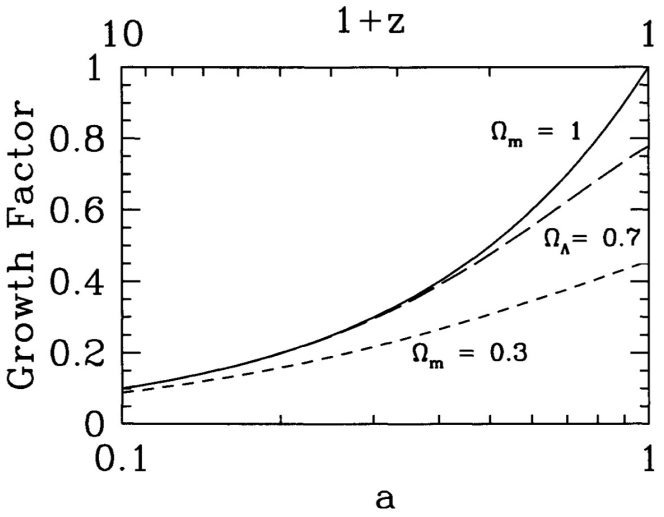

Figure 1.3: The growth factor in three cosmologies. The solid line on top represents a flat universe without cosmological constant; the dashed intermediate line a flat universe with cosmological constant (the ΛCDM model); the bottom dashed curve describes an open Universe. Figure from Dodelson (2003).

in a flat, matter-dominated universe. Moreover, since different kinds of dark energy are described by different curves, i.e. different growth histories, an accurate estimate of the growth factor can constrain $D_1(a)$ and consequently $H(a)$, i.e. the amount of dark matter and dark energy and its equation of state.

To give an even more concise summary of the effect of cosmology on the growth of perturbations, it is helpful to introduce the quantity f, called the *growth rate*, defined by:

$$f \equiv \frac{d \ln D_1}{d \ln a} = \frac{a}{D_1} \frac{dD_1}{da} \tag{1.39}$$

An accurate approximation to the growth rate is given by the expression:

$$f = \Omega_M(z)^\gamma \tag{1.40}$$

where the exponent γ depends on the type of dark energy (for the ΛCDM model is $\gamma \sim 0.55$). Measuring f, therefore, allows to distinguish between different dark energy models.

Chapter 2

Statistical properties of mass and galaxy clustering

As we have seen in the previous chapter, it is possible to derive solutions for the **Eq. 1.28 - 1.32**, meaning the equations that govern dark matter perturbations when linear theory applies. In this regime the Fourier transforms of all the fields (density, velocity, potential etc.) produce a set of independent ordinary equations that can be solved for each k-mode separately. However, the solution to the equations requires specifying initial conditions, which are not trivial to guess.An effective way of compressing the information of a perturbed field is by estimating its *statistical properties*. Let us focus on the matter density fluctuations.

Density fluctuations in a specified volume are a realization of the overdensity random field δ at the generic position $\vec{x}$: $\delta(\vec{x}) = \delta_i$. All possible realizations δ_i constitute an ensamble. We assume that the Universe we observe is a fair sample of this ensamble. This implies that the mean overdensity in the sample averages to zero. A significant fraction of the information is encoded in 2-point statistics, i.e. the *two point correlation function*, defined as:

$$\xi(\vec{r}) = \langle \delta(\vec{x}) \delta^*(\vec{x} + \vec{r}) \rangle \tag{2.1}$$

where $\langle\rangle$ indicates ensamble average over all pairs with separation $\vec{r}$ (the $*$ symbol indicates the operation of complex conjugation) . The 2-point correlation function expresses the excess probability (compared with that expected for a random distribution) of finding a pair of galaxies at a separation $\vec{r}$ (see **Fig. 2.1**). Together with this quantity, it is useful to define

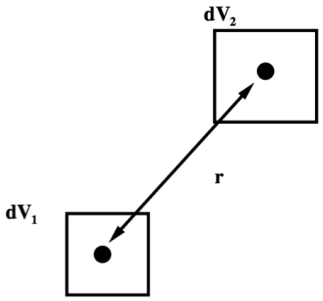

Figure 2.1: The correlation function expresses the excess probability (compared with that expected for a random distribution) of finding a pair of galaxies at a separation $\vec{r}$.

(and we will soon understand why) its Fourier transform, the *power spectrum* $P(\vec{k})$:

$$\langle \delta(\vec{k})\delta^*(\vec{k'})\rangle = (2\pi)^3 P(\vec{k})\delta_D^3(\vec{k}-\vec{k'}) \tag{2.2}$$

where δ_D^3 is the 3D Dirac delta. If we consider a large enough volume so that the assumption of isotropy is valid in the clustering properties, then the power spectrum does not depend on the direction of the wavenumber[1], and we can express the spectrum as function of the module of $\vec{k}$: $P(k)$. The great advantage of defining the power spectrum over the 2-point correlation function is that the different Fourier coefficients δ_k are uncorrelated until the evolution of the perturbations remains linear, while the number of pairs at different separation become correlated because of gravity. We will refer to the power spectrum of the density fluctuations in

[1]If we rotate or translate a given point δ_i the probability of having a correlation should be the same.

the matter component as *matter power spectrum*.

The 2-point statistics is sufficient to fully describe the statistical proprieties of the field if the density fluctuations are Gaussian distributed. The theoretical prejudice is that the primordial density field should indeed obey a Gaussian statistics. Such are in fact the properties of the density fluctuations generated during the *inflation era*. The cosmological inflationary theory states that the Universe has undergone, in the very early epochs, a period of accelerated expansion, after which the Universe appears highly homogeneous and flat on sub-horizon scales. The present day cosmic structures formed from the quantum fluctuations that have survived the inflationary expansion and that have been promoted from microscopic to cosmological scales in the process. As already noted, a variety of inflationary models predict that primordial perturbations obey to Gaussian statistics, and even if deviations from Gaussianity are possible, they are constrained to be very small by Cosmic Microwave Background observations (Komatsu et al., 2011). At later times, when density fluctuations grow larger, the nonlinear evolution destroys the primordial Gaussianity. It has been shown that, well into the nonlinear regime, the probability of density fluctuations is well described by a *lognormal* statistics rather than a Gaussian one (Coles and Jones, 1991). We shall elaborate further on the properties of the lognormal model in **Chapter 5**.

Coming back to **Eq. 1.35**, it expresses the dark matter density fluctuation field δ in terms of the gravitational potential ϕ for modes that enters the horizon during the matter dominated epoch, i.e. after the equivalence epoch, a_{eq}. If ϕ has an analytical expression, one could substitute **Eq. 1.35** into **Eq. 2.2** and thus obtain an analytical expression for the matter power spectrum (in linear theory); however, in order to do this, we have to specify the gravitational potential during the matter dominated epoch. Moreover, as already said, at later times ($a >> a_{eq}$) the linear evolution of the perturbations does not depend on k, and all Fourier modes evolve in the

same way: let's define a_{late} the value of the scale factor when this happens. If $\phi_P(\vec{k})$ is the potential at the end of inflationary epoch (the *primordial* potential), we can write schematically:

$$\phi(\vec{k}, a) = \phi_P(\vec{k}) \times \{\text{Transfer function(k)}\} \times \{\text{Growth Factor(a)}\} \quad (2.3)$$

here the Transfer Function describes the evolution of perturbations through the epoch of horizon crossing and radiation/matter transition, while the Growth Factor, as we have said, accounts for late time evolution ($a > a_{late}$). In particular, the transfer function is defined as:

$$T(k) \equiv \frac{\phi(k, a_{late})}{\phi_{large-scale}(k, a_{late})} \quad (2.4)$$

here $\phi_{large-scale}(k, a_{late}) \sim 9/10\ \phi_p$. Following its definition, the Transfer Function is equal to unity on large scales (the ones that never cross the horizon). Regarding the Growth Factor, it is defined as:

$$\frac{\phi(a)}{\phi(a_{late})} \equiv \frac{D_1}{a} \qquad a > a_{late} \quad (2.5)$$

with D_1 the growth factor already defined. The gravitational potential at late times is then:

$$\phi(\vec{k}, a) = \frac{9}{10}\phi_P T(k)\ \frac{D_1(a)}{a} \quad (2.6)$$

Substituting in the **Eq. 1.35** we obtain:

$$\delta(\vec{k}, a) = \frac{3}{5}\frac{k^2}{\Omega_M H_0^2}\phi_P(\vec{k})T(k)D_1(a) \quad (2.7)$$

Assuming that $\phi_p \propto k^{4+n_s}$, the power spectrum of matter density fluctuation at late times can then be obtained:

$$P(k, a) \propto \frac{k^{n_s}}{H_0^{n_s+3}}T^2(k)\left(\frac{D_1(a)}{D_1(a=1)}\right)^2 \quad (2.8)$$

This is the *linear power spectrum of matter*, and has dimensions of $[\text{length}]^3$. The term n_s is called *scalar spectral index*, because it expresses any deviation from *scale-invariance* (that implies $k^3P(k) = constant$); if $n_s = 1$ the

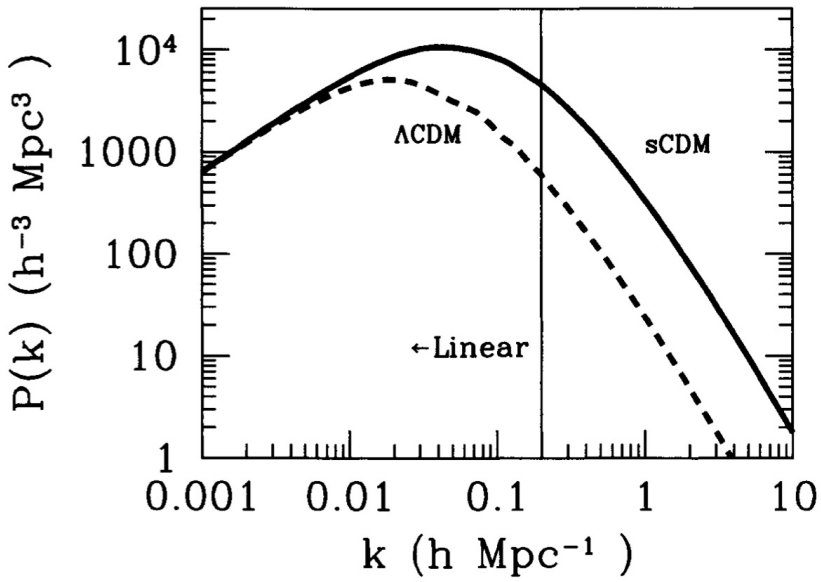

Figure 2.2: The power spectrum in two Cold Dark Matter models, with (ΛCDM) and without (sCDM) a cosmological constant. They have been normalized to agree on large scales (small k). The vertical line evidences the validity of linear regime (to the left). See text for details. Figure from Dodelson (2003).

fluctuations in the gravitational potential do not depend on scale. In **Fig. 2.2** we show the typical shape of the linear power spectrum at present time ($a = 1$ or $z = 0$) for two different flat cosmological models: with (ΛCDM) and without (sCDM) a cosmological constant. On large scales (small k) in both models $P(k) \propto k^{n_s}$ since the Transfer Function here is 1. Below a given scale, the power spectrum turns over. This is due to the fact that the small-scale modes enters the horizon well before the matter-radiation equality, during the radiation era, where their evolution is either delayed or obliterated. Here the power decreases because of the free streaming, radiation drag and all types of dissipation processes, so the transfer function is much smaller than unity. Modes that enter the horizon even earlier undergo more suppression. Thus, the power spectrum is a decreasing function of k on small scales. This means that there will be a turnover in the

shape of the power spectrum at the scale that corresponds to modes that enter the horizon *at* matter-radiation equality. The vertical line in figure refers to the scale above which non-linearities at the present epoch cannot be ignored, and can be obtained when $(k^3 P(k)/2\pi) \sim 1$, corresponding to $k \simeq 0.2\, h\, \mathrm{Mpc}^{-1}$.

In **Fig. 2.2** two models are shown. The main difference between them is that sCDM has more matter ($\Omega_M = 1$) and hence the equality occurs at an earlier epoch. This means that only the very small scales succeed in entering the horizon before the matter-radiation equivalence, and therefore the turnover in the spectrum occurs on smaller scales.

So far we have considered only dark matter perturbations, but we can't forget that the baryon also contribute to the mass budget. They have a secondary role, but anyway affects the transfer function shape. Unlike the dark matter, the baryonic matter decouple from photon much later, at the last scattering epoch. After, the baryons couple to dark matter through gravity and, being subdominant, trace the spatial distribution of the dark component. Before the recombination the perturbations of the photo-baryon plasma don't grow significantly with time, but either propagate as acoustic waves or remain frozen in the plasma when they're not dissipated by diffusion processes. Yet, density fluctuations in the baryon component do survive the plasma epoch and, since the contribution to the total mass budget is not negligible, leave their imprint in the spatial distribution of the matter in the form of *baryon acoustic oscillations*. The BAO signature is quite characteristics and has been detected in the 2-point correlation statistics of the galaxy distribution ((Cole et al., 2005) and (Eisenstein et al., 2005)).

The results of the complex interplay between baryon and plasma physics, dark matter properties and cosmological background are encoded in 2-point statistics that, therefore, can be used to infer cosmological param-

eters and physical properties. To illustrate this issue we show different model power spectra obtained by changing some of these parameters (**Fig. 2.3**). Among the relevant quantities that can be inferred from the power spectrum, there are:

- The *amplitude* of the primordial power spectrum of the gravitational potential ϕ_P. This quantity is indicated as A_s and is related to the typical amplitude of the mass density fluctuation by **Eq. 2.7**.

- The *shape* of the primordial power spectrum, which should be preserved on large scales, is uniquely characterized by one single parameter: the primordial spectral index n_s that can be inferred by the spectral slope at small k's.

- The matter density parameter, Ω_M, defines, together with the radiation density parameter, the value of the scale factor at equivalence (a_{eq}), which in turns defines the point of turn over in the power spectrum.

- The dark energy density parameter, Ω_Λ, affects the evolution of the perturbation at late times, mainly suppressing their growth. As we have seen, its specific impact depends on the specific type of dark energy. In addition, since in a flat Universe ($\Omega = 1$) Ω_Λ and Ω_M are not independent, changing the former modifies the scale of the spectral turnaround scale.

- The baryonic density parameter, Ω_b, also suppresses the power on small scales. Moreover, The larger Ω_b the more pronounced is the BAO feature. Therefore a large Ω_b values would increase the amplitude of the characteristic wiggles of the power spectrum.

- Another important parameter is σ_8 the root mean square `rms` fluctuation in total matter in spheres of 8 Mpc/h at $z = 0$ computed in

linear theory. The `rms` is:

$$\sigma_R^2 = \int \frac{dk}{2\pi^2} k^2 P(k) \left[\frac{3j_1(kR)}{kR}\right]^2 \tag{2.9}$$

where $R = 8$ Mpc$/h$ and j_1 is the spherical Bessel function of order 1. This means that σ_8 measures the normalization of the power spectrum, and is a derived parameter: σ_8 and A_s are related through **Eq. 2.9** which, in turn, depends on the other cosmological parameters.

- The mass density in neutrinos Ω_ν and the number of neutrinos families N_ν modify the value of a_{eq} and suppress the power on scales smaller than that of the turnover.

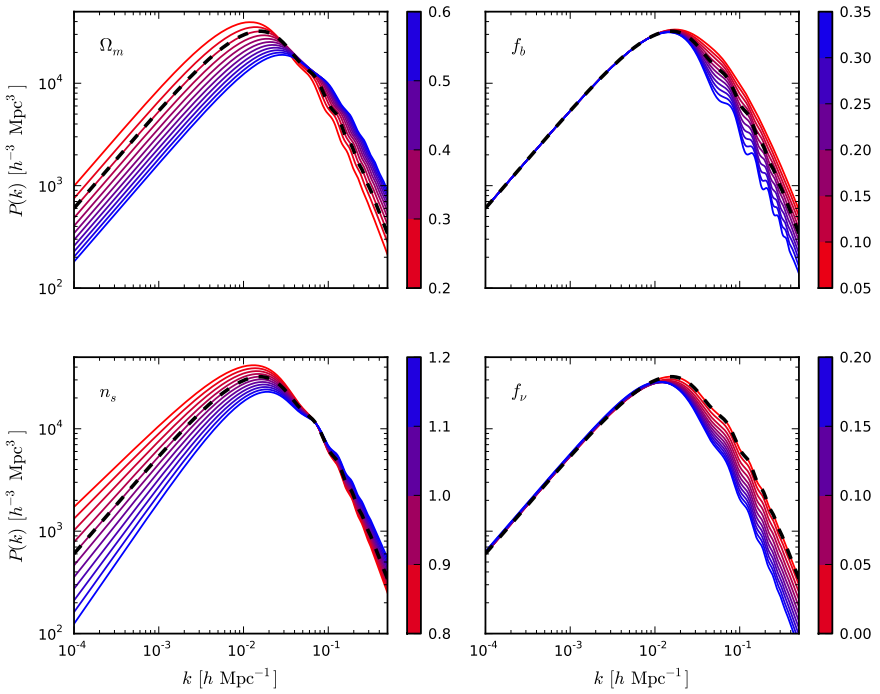

Figure 2.3: The matter power spectrum plotted for different values of the cosmological parameters (color lines): the matter density parameter Ω_M, the baryon fraction $f_b \equiv \Omega_b/\Omega_M$, the spectral index n_s and the neutrino density fraction $f_\nu \equiv \Omega_\nu/\Omega_M$. The colour bar indicates the value of the parameter. The dashed black curve is the power spectrum for the fiducial cosmology considered in Parkinson et al. (2012), and is the same in all four plots. Figure from Parkinson et al. (2012).

2.1 The angular power spectrum

The 2-point correlation function and its Fourier transform, the power spectrum $P(\vec{k})$, encode information on the statistical properties of the mass distribution. More often than not, astronomical datasets only provide information of the angular position and magnitude of a source and not its redshift. In this case we can still extract cosmological information by ana-

lyzing the *angular* rather than the spatial 2-point correlation properties. In fact, it is possible to define 2-point statistics analogous to those introduced in the previous section for a bi-dimensional dataset. The first one is the angular 2-point correlation function $\omega(\theta)$, i.e. the probability of finding two galaxies in small elements of solid angle $d\Omega_1$ and $d\Omega_2$ separated by an angle θ on the celestial sphere (see **Fig. 2.4**):

$$dP = \bar{n}(1 + \omega(\theta))d\Omega_1 d\Omega_2 \tag{2.10}$$

where $\bar{n}$ is the number of objects per steradian (i.e., the object surface density).

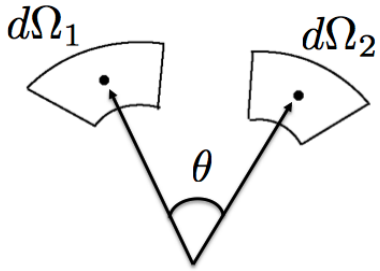

Figure 2.4: The angular correlation function is the probability of finding two galaxies in small elements of solid angle $d\Omega_1$ and $d\Omega_2$ separated by an angle θ on the celestial sphere.

We can further define the *angular power spectrum* as the Legendre transform of the angular correlation function:

$$C_\ell = 2\pi\bar{n}\int_{-1}^{1}\omega(\theta)P_\ell(\cos(\theta))\mathrm{d}(\cos)\theta \tag{2.11}$$

The rigorous definition of the angular power spectrum involves a continuous bidimensional density field $\sigma(\theta,\phi)$, which, being defined on the celestial 2D sphere can be expanded in spherical harmonics $Y_{\ell m}(\theta,\phi) \propto e^{im\phi}P_\ell(\cos\theta)$:

$$\sigma(\theta,\phi) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} a_{\ell m}Y_{\ell m}(\theta,\phi) \tag{2.12}$$

The expansion in spherical harmonics can be thought of a generalized Fourier transform: whereas the complete set of eigenfunctions for the Fourier transform are $e^{i\vec{k}\cdot\vec{x}}$, here the complete set of eigenfunctions for expansion on the surface of a sphere are $Y_{\ell m}(\theta,\phi)$. All of the information contained in the surface density field is also contained in the spherical harmonics coefficients $a_{\ell m}$.

Note that the orthonormality properties of the $Y_{\ell m}(\theta,\phi)$ implies that the last equation can be reversed:

$$a_{\ell m} = \int \sigma(\theta,\phi) Y_{\ell m}(\theta,\phi) d\Omega \tag{2.13}$$

If we focus on the ℓth multipole, the angular power spectrum expresses its amplitude:

$$C_\ell \equiv \langle |a_{\ell m}|^2 \rangle \tag{2.14}$$

where the brackets imply an averaging over many realizations of the density field. In other words, the angular power spectrum is the variance of the $a_{\ell m}$'s[2]. Since the assumption of isotropy ensures that $\langle |a_{\ell m}|^2 \rangle$ depends only on ℓ and not on m, then for a given ℓ each $a_{\ell m}$ has the same variance (see **Fig. 2.5**).

To show this point let us consider the case of $\ell = 100$. The number of coefficients associated to this multipole is $2\ell + 1 = 201$. Their amplitudes are drawn from the same distribution. So the value of C_ℓ for $\ell = 100$ is obtained by sampling the distribution 201 times. In the case of the quadrupole ($\ell = 2$) the distribution is sampled only 5 times ($m = -2, -1, 0, 1, 2$), hence providing a noisy estimate of its variance. This means that there is an intrinsic uncertainty in the C_ℓ which depend on the correlation properties and which we call **cosmic variance**. For a Gaussian random field, the cosmic variance is simply:

$$\left(\frac{\Delta C_\ell}{C_\ell}\right)_{\text{cosmic variance}} = \sqrt{\frac{2}{2\ell+1}} \tag{2.15}$$

[2]As for the density perturbations, we cannot make predictions about any particular $a_{\ell m}$ but only about the distribution from which they are drawn.

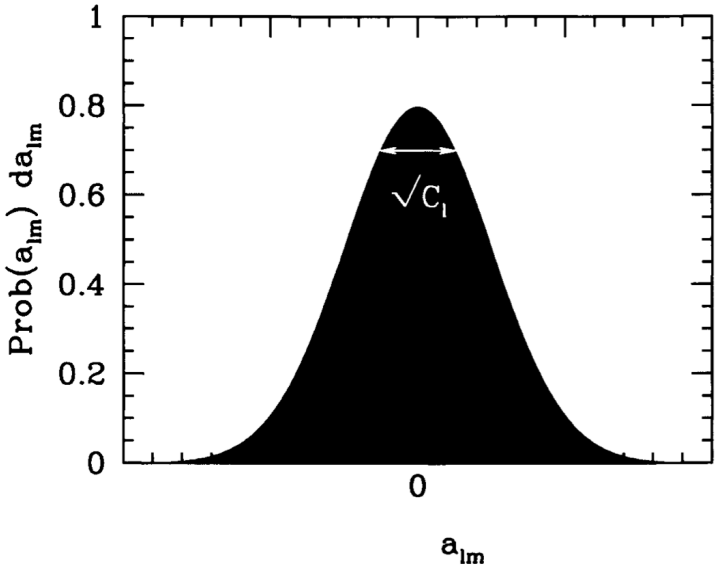

Figure 2.5: The distribution from which the $a_{\ell m}$'s are drawn. The distribution has expectation equal to zero and a width equal to $\sqrt{C_\ell}$. Figure from Dodelson (2003).

2.1.1 From $P(k)$ to C_ℓ

In order to use the angular power spectrum as a cosmological probe, we need to unveil its dependence on the cosmological parameter through its relation with the *spatial* power spectrum.

Let us consider some mass tracers, let us say galaxies, and their over-density δ. If $\hat{q}$ is a unit vector that specifies a certain direction on the sky, it is possible to define the density perturbation seen on the sky as (Peacock, 1999):

$$\delta(\hat{q}) = \int_0^\infty \delta(\vec{y}) y^2 \phi(y) dy \tag{2.16}$$

where y is the comoving distance and $\phi(y)$ is the so-called **selection function** of the sample, which quantify the probability of an object at y. This probability is normalized, i.e. $\int y^2 \phi(y) dy = 1$. The integral in **Eq. 2.16** is performed along the line of sight and ϕ represents a statistical weight to the density fluctuation at y. In general, the selection function depends

both on the radial and on the angular distance. However, the angular dependence is typically taken care of by the geometry mask, as we shall see, so that the selection function can be defined only radially. In the cases that we explore, the selection effect is induced by a flux cut in the galaxy sample, effectively inducing a deterministic relation between the selection function and the luminosity function of the galaxy catalog:

$$\phi(y) = \frac{\int_{\mathrm{Max}(L(y),L_{min})}^{\infty} \Phi(L)dL}{\int_{L_{min}}^{\infty} \Phi(L)dL} \tag{2.17}$$

The selection function can also be written in terms of the number of objects one have in a particular redshift bin dz:

$$\phi = \frac{1}{\mathcal{N}} \frac{dN}{dz} \tag{2.18}$$

where $\mathcal{N}$ is the total number of galaxies in the survey:

$$\int_0^{\infty} \frac{dN}{dz} dz \equiv \mathcal{N} \tag{2.19}$$

The following step to link $P(k)$ to C_ℓ is to Fourier transform **Eq. 2.16** using spherical harmonics. This is possible thanks to the relation between plane waves and spherical bessel functions $j_n(x) = (\pi/2x)^{1/2} J_{n+1/2}(x)$ (i.e. the Rayleygh expansion of plane waves, e.g. Press et al. (1992)):

$$e^{iky\cos\theta} = \sum_{\ell=0}^{\infty} (2\ell+1) P_\ell(\cos\theta) j_l(ky) i^\ell \tag{2.20}$$

Using the latter formula, the *addition theorem*:

$$P_\ell(\cos\theta) = \frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{m=\ell} Y^*_{\ell m}(\hat{q}) Y_{\ell m}(\hat{q}') \tag{2.21}$$

and the completeness relation:

$$\sum_\ell \sum_m Y^*_{\ell m}(\hat{q}) Y_{\ell m}(\hat{q}') = \delta(\hat{q} - \hat{q}') \tag{2.22}$$

one can obtain, starting from the definition of the angular correlation function $\omega(\theta) = \langle \delta(\hat{q})\delta(\hat{q}') \rangle$, the following important relation:

$$C_\ell = \langle |a_{\ell m}|^2 \rangle = 4\pi \int \Delta^2(k) \frac{dk}{k} \left[\int y^2 \phi(y) j_\ell(ky) dy \right]^2 \tag{2.23}$$

where:

$$\Delta^2(k) \equiv \frac{1}{2\pi^2} k^3 P(k) \tag{2.24}$$

Eq. 2.23 clarifies the link between the angular power spectrum, which is the quantity that we will compute in this thesis, and the 3D power spectrum, which is what Theory typically predict.

Analogously to **Fig. 2.3**, **Fig. 2.6** shows the impact on the angular power spectrum of the variation of four cosmological parameters. In this case, we consider the angular spectrum of CMB temperature variations ΔT.

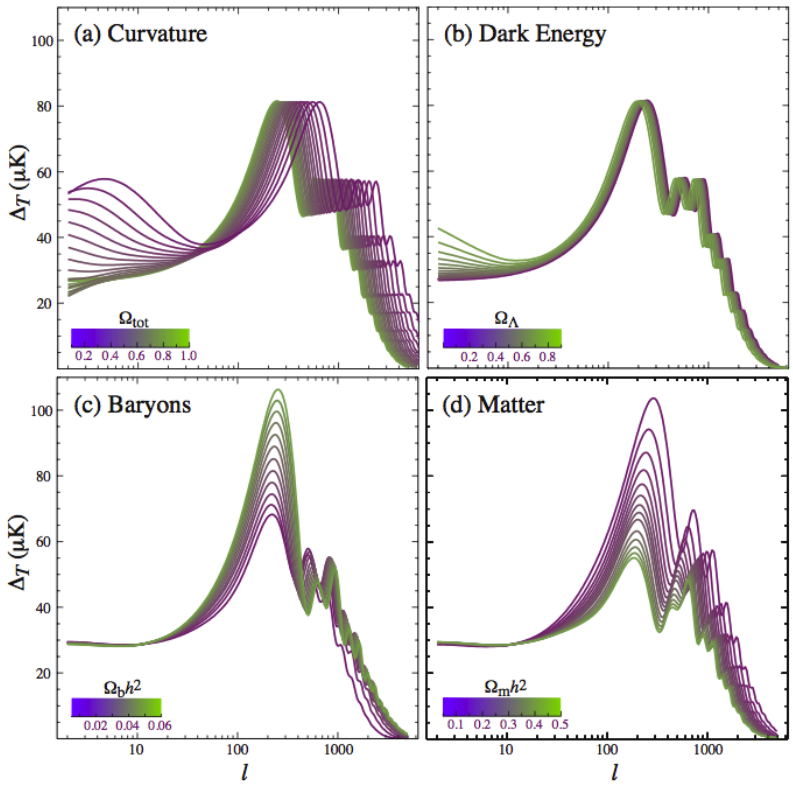

Figure 2.6: Example of the sensitivity of the angular power spectrum to four cosmological parameters. In this case, the angular spectrum of CMB temperature variations ΔT is considered. The parameters are: the curvature, quantified by Ω_{tot}, the dark energy density parameter Ω_Λ, the baryon density parameter Ω_b and the matter density parameter Ω_M. Figure from Hu and Dodelson (2002).

2.2 Measuring the matter power spectrum from real datasets

Now that we have defined the spatial and the angular power spectra and their dependence on the cosmological parameters, let us face the problem on how to *measure* it. First of all, let us point out that 2-point statistics that we have defined for continuous fields, can also be defined for a set of

discrete mass tracers. All one has to do is to describe the density field as a sum of Dirac delta functions centered on the objects positions. It is then possible to estimate the matter power spectrum from the spatial or angular positions of extragalactic objects provided that they trace the underlying density field. Large collections of such objects, that specify their angular and radial positions over large volumes, can therefore be used to estimate the power spectrum of the matter. In practice, however, two issues need to be addressed to achieve this goal. First of all: are galaxies adequate tracers for the distribution of the total matter, which we know to be mostly dark? Secondly, how good is galaxy redshift as a proxy to galaxy distance?

In fact, the relation between galaxies and dark matter is far from trivial. Galaxies are though to form within gravitational potential wells dominated by dark matter and, more specifically, within virialized structures called *dark matter haloes*. The physical processes that govern the evolution of the haloes and the galaxies within are very complex. On one hand, dark matter haloes are nonlinear structures, so that their evolution can not be described by the linear equations we have considered in the previous chapter. On the other hand, the galaxies themselves are highly elaborate structures, whose formation and growth is determined by the complex physics of the baryons and by stellar feedback processes.

In addition, there is another problem related to the use of galaxy redshift surveys. Even if one ignores the distortion effects induced by peculiar velocities, errors in the redshift measurements *do* introduce an uncertainty in the radial position of the object. The magnitude of the error depends on the technique used to measure the redshift. Accurate redshift estimate from galaxy spectra are time consuming. If the goal is to gather as many objects as possible then another, less precise, method, called *photometric redshift*, can be adopted. This comes at the price of increasing the uncertainty on the radial distance and, consequently, the fuzziness of the observed cosmic structures. This is where the angular correlation studies that we shall perform in this thesis come into play.

2.2.1 Galaxy bias

Let us elaborate on how well galaxies trace the underlying mass. The mapping is thought to be complex and not fully captured by current models of galaxy formation. Yet, this mapping, that we call **galaxy bias**, can be operationally defined and estimated, so that our ignorance on the galaxy evolution processes can be conveniently parametrized.

Let us focus on scales much larger than those affected by star formation and evolution processes. In this case, the fluctuations are still in the linear regime and we can assume that the bias relation between the galaxy density field and the mass field is *linear* too. If the galaxy density is uniquely determined by the underlying mass density, i.e. the bias relation is *local*, we can then write:

$$\delta_g = b\delta_M \tag{2.25}$$

here b is the linear galaxy bias. In this case the galaxy power spectrum and the linear matter power spectrum are related by a direct proportionality law:

$$P_g(k) = b^2 P(k) \tag{2.26}$$

Even if commonly adopted in literature, this relation is valid *only* on large scales. In this Thesis we shall assume that galaxy bias is indeed linear. The reason for this is twofold. First we shall avoid nonlinear scales and focus instead on angular scales where deviations from linear theory are not severe. Second, the 2D density field is obtained by projecting the 3D one on the celestial sphere. The projection effect effectively smooths out nonlinearity and, again, make the field close to linear.

Besides the linear bias model, in our analysis we shall also consider nonlinear ones in which nonlinear dynamics and bias are modeled in the framework of the *Halo Model* (HM hereafter), which provides a theoretical scheme to describe the clustering of dark matter, haloes and galaxies in

both linear and non-linear scales (Cooray and Sheth, 2002). As we have anticipated in the previous section, the basic idea of the HM is that galaxies can form only in the potential wells of virialized dark matter structures called *haloes*, that in turn are formed in correspondence of the peaks of the density field via gravitational collapse. As a result, the galaxy properties are strictly related to the mass and angular momentum of the parent halo.

If the Universe is thought to be partitioned up into distinct units like haloes, which are small compared with the typical distance between them, then the details of the internal structure of a single halo cannot be important on scales larger than a typical halo. In other words, on large scales the important ingredient is only the spatial distribution of the haloes. This lead to the realization that we can consider two different contribution to two point statistics: the 1-halo and the 2-halo components. Regarding the power spectrum, this separation can be quantified as:

$$P_g(k) = P_g^{1h}(k) + P_g^{2h}(k) \tag{2.27}$$

where the first term on rhs is the power relative to the structure within a single halo, while the second term is relative to the correlation between haloes. The latter term dominates on large scales, where the galaxy power spectrum is related to the linear matter power spectrum by an equation similar to **Eq. 2.26**, with b equal to:

$$b(z, m_g, L_g) = \int dm\; n(z,m)\; b_{1,\,\mathrm{halo}}(z,m)\; \frac{\langle N_{\mathrm{gal}}|m\rangle}{\bar{n}_{\mathrm{gal}}} \tag{2.28}$$

here m is the halo mass, $n(z,m)$ is the *halo mass function*, which gives the comoving number density of haloes with mass m at redshift z, $b_{1,\,\mathrm{halo}}(z,m)$ is the linear bias of the halo with mass m at redshift z, $\bar{n}_{\mathrm{gal}}$ is the mean number density of galaxies and $\langle N_{\mathrm{gal}}|m\rangle$ is the mean number of galaxies N_{gal} in a halo of mass m. The integral is to be performed over all halo masses.

In order to derive the halo mass function and the halo bias, one has to assume a dynamical model for the halo formation in a given cosmological framework. For example, a simple and useful approximation is the

spherical collapse of an initially tophat density perturbation (Gunn and Gott, 1972). Once the halo mass function and the halo bias are specified, one has to assume an halo occupation number $\langle N_{\rm gal}|m\rangle$, that generally depends on halo mass, redshift, type of galaxy, luminosity and of course mass of the galaxies. Qualitatively, one expects that galaxies with high mass and luminosity tend to populate the higher peaks of the density field of matter so that they are more clustered than less massive or faint galaxies, which are distributed more homogeneously (Kaiser, 1984). Moreover, galaxies with early morphological type, i.e. elliptical galaxies in the Hubble morphological classification, preferentially populate the massive galaxy clusters, whereas spiral galaxies are commonly found in lower density environments. The bias also depends on the redshift, because when galaxies started forming, in the past, they populated the highest density peaks that are statistically more clustered.

The HM model, like the various models of galaxy evolution, does not allow to predict the galaxy bias with sufficient accuracy in order to constrain cosmological models. Therefore, in practice, we estimate galaxy bias by comparing the measured 2-point statistics to theoretical prescription of the matter counterpart. In particular, as we will see in our analysis, the linear bias b is set as free parameter; the degeneracy with other cosmological parameters can only be broken through combinations with independent measurements like gravitational lensing or higher order statistics in galaxy clustering.

2.2.2 Non linearities

Different approaches can be used to address the issue of the non-linear evolution of density fluctuations.

The simplest way is to follow the evolution of a spherical overdensity starting from a small fluctuations that expands with the Hubble flow. The evolution of such fluctuation can be traced analytically. If the initial con-

trast is sufficiently large, it will eventually collapse and form a virialized structure.

Alternatively, one can push the perturbative approach described in Jeong and Komatsu (2006) to include second or higher order terms. This improves theoretical predictions in the mildly nonlinear regime at the price of an increasing mathematical complexity.

An elegant alternative approach is the so-called *Zel'dovich approximation* (Zel'dovich, 1970), that consists in following the trajectories of a particle distribution. Starting from the initial unperturbed Lagrangian coordinate $\vec{q}$, and assuming that the particles move along straight lines, the Eulerian position of each particle can be specified at the generic time as:

$$\vec{x} = \vec{q} + \psi(\vec{q}, \eta) \tag{2.29}$$

where $\psi(\vec{q}, \eta)$ is the displacement field. In other words, since the Zel'dovich approximation follows the evolution of the single particle, it is a first order Lagrangian perturbation theory, different from the Eulerian perturbation theory we used in **Section 1.2**, that instead follows the evolution of density and velocity fluctuations. This means that the Zel'dovich approximation consists in a purely kinematical approach and doesn't take into account close-range forces or pressure or shocks, and breaks down only when particles trajectories cross each other, when discontinuity in the mass density field, called *caustics*, are formed. However, for small displacements, the Zel'dovich approximation is an effective way to follow the evolution well beyond the linear regime.

No analytic approach can follow the evolution of a nonlinear self gravitating system. This is why numerical methods, the so called N-body simulations, have been developed and are now being used as standard tools to trace the evolution of density fluctuations and their spatial correlation properties.

2.2.3 Redshifts and peculiar velocities

Measured redshifts are used to infer the proper distance of an object once a cosmological model is assumed. This inference, however, is affected by systematic errors induced by the so called peculiar velocities. Peculiar velocities, i.e. fluctuations in the Hubble velocity flow, add to recession velocities so that the total velocity is:

$$\vec{v}_{obs} = \vec{v} + \hat{r} v_{cosm} \tag{2.30}$$

where $\vec{v}_{obs}$ is the total galaxy's velocity, v_{cosm} is the velocity of the expansion and $\hat{r}$ the radial versor. If we focus on the low redshift limit we have:

$$s_{obs} \equiv c z_{obs} = c z_{cosm} + \hat{r} \cdot \vec{v} = H_0 d + v \tag{2.31}$$

where d is the comoving distance and v is the component of the peculiar velocity in the radial direction. If one measures redshift as proxy to distances:

$$d_s = \frac{s_{obs}}{H_0} = \frac{c\, z_{obs}}{H_0} \tag{2.32}$$

then a systematic error is introduced that, because of the coherence of the velocity field, correlates with similar errors in the positions of nearby objects. The net effect is to induce a spurious, radial anisotropy in the clustering of objects that biases the estimate of 2-point statistics. The effect is usually referred to as **redshift space distortion**.

Redshift distortion affect the angular power spectrum too. To quantify their impact, let us consider **Eq. 2.23** and focus on the function (Fisher et al. (1994), Padmanabhan et al. (2007), Thomas et al. (2011)):

$$K_\ell(k) \equiv \int y^2 \phi(y) j_\ell(ky) dy \equiv \int f(y) j_\ell(ky) dy \tag{2.33}$$

where $f(y)$ is the weighting function related to the selection function of the sample. Writing the weight as a function of redshift distance $f(s)$ and assuming that the magnitude of the peculiar velocities are small compared with the depth of the sample, it is possible to perform a Taylor expansion:

$$f(s) \approx f(r) + \frac{\mathrm{d}f}{\mathrm{d}r} \left[\vec{v}(r\,\hat{r}) \cdot \hat{r}\right] \tag{2.34}$$

As a result the function $K_\ell(k)$ is the sum of two terms, $K_\ell(k)$ and $K_\ell^R(k)$. The first one accounts for the distribution of the objects in real space, whereas the second one includes redshift distortion effects. The Fourier transform of $\vec{v}$ is in turn related to the density perturbation through the linear continuity equation:

$$\vec{v}(\vec{k}) = -i\beta\delta_g(\vec{k})\frac{\vec{k}}{k^2} \tag{2.35}$$

where δ_g is the galaxy density contrast, while the constant of proportionality β is the ratio between the growth rate of density fluctuations and the linear bias parameter of the mass tracer and is called *redshift distortion parameter*. Substituting this into the expression of $K_\ell(k)$ and Legendre transforming (see Padmanabhan et al. (2007) for the detailed calculation) leads to:

$$\begin{aligned} K_\ell^R(k) =& \beta \int f(y) \left[\frac{(2\ell^2+2\ell-1)}{(2\ell+3)(2\ell-1)} j_\ell(ky) + \frac{\ell(\ell-1)}{(2\ell-1)(2\ell+1)} j_{\ell-2}(ky) + \right. \\ & \left. - \frac{(\ell+1)(\ell+2)}{(2\ell+1)(2\ell+3)} j_{\ell+2}(ky) \right] dy \end{aligned} \tag{2.36}$$

For large value of ℓ, the integral in the previous equation tends to zero, so $K_\ell(k)$ is reduced to the previous form. This means that, even with the inclusion of redshift distortions, the small angle approximation leads to an accurate estimate of the angular power spectrum at small scales. On the contrary, redshift distortions do have an impact of the power at small multipoles.

We shall account for redshift space distortions in modeling the angular power spectrum in the analysis of real data (**Chapter 7**).

Let us elaborate further on the methods of measuring galaxy redshift, their errors and the impact on clustering studies.

The most accurate way of measuring redshift is through spectroscopy, i.e. by measuring the wavelength of a well known line in the spectrum

of an object. The higher the dispersion of the spectrograph the better is precision of the measurement. This redshift, called *spectroscopic redshift*, can only be obtained for relatively bright objects, because it requires sufficient signal-to-noise ratio to detect individual emission and/or absorption lines. Large observational efforts are in progress (e.g. VIPERS, see Guzzo et al. (2014) and BOSS, see Dawson et al. (2013)) or planned (EUCLID, Laureijs et al. (2011)) to map the 3D galaxy distribution using spectroscopic redshifts. As this is a time-consuming technique, the number of objects in these surveys is significantly smaller than those in the photometric samples, in which one estimates redshift from the magnitude of objects measured in different bands. This photo-z technique provide an estimate of the redshift considerably less precise than the spectroscopic, but for a comparatively larger number of objects. The idea is that one can obtain a sampling of an object's Spectral Energy Distribution (SED)[3] by measuring the flux of the object in a relatively small number of wavebands. We will not go into details[4], but the SED of galaxies has some broad spectral features that change as a function of redshift. When compared with theoretical predictions obtained by repeating the same analysis over a sample of template galaxy spectra, one obtains an estimate of the redshift, which is what we call *photometric redshift*. Obviously, its precision is inferior to the spectroscopic one: photometric errors are typically of the order of 3–10% while the spectroscopic ones stay around 0.1%.

As we will see in the following chapters, in our analysis we will consider the 2MASS photometric catalogue (2MPZ) (Bilicki et al., 2014). We will then have to face the problem of inferring the radial position of the objects (or its spectroscopic redshift) from photometric redshifts, in order to

[3]The Spectral Energy Distribution, often called SED, is just the flux density as a function of the frequency or wavelength of light.

[4]For more details, see Mo et al. (2010), Chapter 2

have a deeper understanding of impact of photo-z errors in the clustering analysis. We shall investigate this issue in **Chapter 5**.

2.2.4 A code for the power spectrum model: CLASS

Once computed the 2-point statistics (in our case, the angular power spectrum) from a catalogue, the following step is to compare it to theoretical predictions. As we have discussed, a numerical treatment would require to perform N-body simulations of different cosmological models and measure their power spectra. This strategy would be too computationally expensive. Fortunately, when it comes to 2-point statistics for the matter density field, analytic or semi-analytic techniques have been developed to predict spatial and angular spectra with high accuracy. Consequently, several numerical code aimed at predicting 2-point stats are publicly available. Examples are `CMBFAST` (Seljak and Zaldarriaga, 1996) or `CMBEASY` (Doran, 2005). Perhaps, the most famous and complete code is surely `CAMB` (Lewis et al. (2000) and Howlett et al. (2012)), written in Fortran90. In our study we have instead chosen to use the code `CLASS` (*Cosmic Linear Anisotropy Solving System*) (Lesgourgues (2011) and Blas et al. (2011)), for two reason: first of all, `CLASS` is written in C/C++, the same language that we used to write our own code to estimate the angular spectrum; secondly, it offers a natural interface with the code `MontePython` (Audren et al., 2013) that we will use for the estimation of the cosmological parameters (see **Chapter 6**).

The code `CLASS` computes spatial and angular spectra for any cosmological models characterized by a number of parameters. This task is done in separate steps that are associated with different structures within the code. `CLASS` can be executed with a maximum of two input files, for example:

```
./class explanatory.ini chi2pl1.pre
```

The file with a .ini extension is the cosmological parameter input file, and

the one with a .pre extension is the precision file. Both files are optional: all parameters are set to default values corresponding to that of the standard cosmological model, and are eventually replaced by the parameters passed in the two input files. For example, in our case the default accuracy settings were sufficient, so we just run `./class explanatory.ini`. Within this file, each parameter that is not specified is automatically left to the default value.

`CLASS` can accommodate substantially all the parameters one could need for a cosmological study: those concerning photons, baryons, cold dark matter, massless neutrinos and other ultra-relativistic species. The user can decide to use the linear theory or `Halofit` to investigate the nonlinear regime, to add a particular selection function ϕ or a window function in order to define the redshift range under study. The code also accounts for the effect of redshift space distortions in both the 3D and the angular spectrum. Among the other outputs, the code provides the power spectrum $P(k)$ and, thanks to the extension `CLASSgal` (Di Dio et al., 2013), it also computes the angular correlation function and the angular power spectrum C_ℓ of the matter density. In our analysis we will use the angular spectrum computed by `CLASS` as a reference theoretical spectrum, as we will examine in depth in the following chapters.

Chapter 3

The angular power spectrum estimate

In the previous chapter we have introduced the concept of angular power spectrum in the ideal case of an all-sky sample. Here, we extend the formalism to account for the most common observational effects such as the incomplete sky coverage. We shall move from the theoretical definition of the angular spectrum to introduce the estimator, describe its properties and describe the associated numerical code. We conclude by testing the performance of the code.

3.1 Theory

From now on, let us define $\delta(\vec{r}, z_i)$ as the three-dimensional galaxy fluctuation at the position $\vec{r}$ in a given redshift bin z_i. As we have seen in **Eq. 2.16**, we can project the 3D fluctuation along the line of sight in order to obtain the angular fluctuation for the ith-redshift bin:

$$\delta_i(\hat{r}) = \int_0^\infty dr r^2 \phi_i(\vec{r}) \delta(\vec{r}, z_i) \tag{3.1}$$

where $\phi_i(\vec{r})$ (normalized such that $\int dr r^2 \phi(r) = 1$) is the selection function defined in **Sec. 2.1.1** and is equal to zero outside the z-bin. The angular

fluctuation field over the sphere can be expanded in spherical harmonics. The coefficients of such expansion are:

$$\delta^i_{\ell m} = \int d\hat{r}\delta_i(\hat{r})Y^*_{\ell m}(\hat{r}) = \int d^3r\ \phi_i(\vec{r})\delta(\vec{r}, z_i)Y^*_{\ell m}(\hat{r}) \tag{3.2}$$

Analogously to **Eq. 2.14** we can define the cross-angular power spectrum between two redshift bins i and j as the ensemble average of the product of the harmonic decomposition averaged over the m modes:

$$C^{ij}_\ell \equiv \frac{1}{2\ell+1}\sum_m \langle \delta^i_{\ell m}\delta^{j*}_{\ell m}\rangle \tag{3.3}$$

The quantity C^{ij}_ℓ indicates the cross-spectrum that reduces to the usual auto-power spectrum when $i = j$. As we have seen in **Sec. 2.1.1**, this angular power spectrum is related with the 3D power spectrum $P(k)$ by the **Eq. 2.23**. To distinguish between the theoretical and the estimated angular spectrum we indicate the latter as $\tilde{C}_\ell$.

Let us now consider the effect of the various observational biases on the measured spectrum. These effects are usually encoded in the so-called selection function ϕ, which quantifies the probability that a galaxy at $\vec{r}$ is included in the sample. The selection function can be conveniently factorized into a radial and an angular component. The angular component, or angular **mask** $M(\hat{r})$, quantifies the angular coverage of the sample:

$$\phi_i(\vec{r}) = \phi_i(r)M(\hat{r}) \tag{3.4}$$

The angular dependence of the mask can be quite complicated since the completeness of the sample may depend on the angular position. Here we shall consider the simple case of a sample 100% complete in some area of the sky (in which $M = 1$) and 0% complete in the remaining areas (corresponding to $M = 0$). The use of such binary mask is not unrealistic since in most cases one can extract a sub-sample characterized by constant completeness over some area and zero objects outside it. $\phi(r)$ represents the radial selection function which, for large enough galaxy sample, can be obtained from the redshift distribution of the objects $N(z)$ assuming a

cosmology dependent $z(r)$ relation. Using the addition theorem (**Eq. 2.21**) and the Fourier transform of the galaxy density field observed at redshift z, we can write:

$$\delta^i_{\ell m} = 4\pi \int d^3r \int \frac{d^3k}{(2\pi)^3} \delta(\vec{k}, z(r)) \sum_{\ell' m'} i^{-\ell'} j_{\ell'}(kr) \phi_i(r) Y^*_{\ell m}(\hat{r}) M(\hat{r}) Y_{\ell' m'}(\hat{r}) Y^*_{\ell' m'}(\hat{k}) \tag{3.5}$$

where $\delta(\vec{k}, z(r))$ is the Fourier transform of the fluctuation $\delta(\vec{r}, z(r))$ at the redshift z. The cross 3D galaxy-galaxy power spectrum is defined as:

$$\begin{aligned} \langle \delta(\vec{k}, r(z)) \delta^*(\vec{k'}, r(z')) \rangle &= (2\pi)^3 \delta^3(\vec{k} - \vec{k'}) \sqrt{P(k,z) P(k,z')} \simeq \\ &\simeq (2\pi)^3 \delta^3(\vec{k} - \vec{k'}) b(z) b(z') D(z) D(z') P_{\text{matter}}(k, z = 0) \end{aligned} \tag{3.6}$$

where $P(k,z)$ is the galaxy power spectrum at redshift z. The second line of **Eq. 3.6** is valid in the linear regime in which i) the galaxy bias $b(z)$ is assumed to be linear and ii) the growth of density fluctuation $D(z)$ is the one predicted by linear theory. In this case $P_{\text{matter}}(k, z = 0)$ would be the matter linear power spectrum. A more sophisticated model would be required to describe the cross-spectrum in the nonlinear regime. Here we shall simply assume that in the nonlinear regime **Eq. 3.6** is still valid but the matter linear power spectrum can be substituted by the nonlinear one. The goodness of this assumption is questionable. However we shall keep in mind that in our analysis we consider a 2D-projected density fluctuation field smoothed over some pixel size. The amplitude of its fluctuations are not large so that effective deviations from linearities are small and **Eq. 3.6** is a good approximation. In any case, the theoretical uncertainties in the model described by **Eq. 3.6** will be included in the error budget of our analysis. Putting together **equations 3.5 and 3.6**, we obtain:

$$\langle \delta^i_{\ell m} \delta^{*j}_{\ell m} \rangle = \sum_{\ell'} \Gamma^m_{\ell \ell'} C^{ij}_\ell \tag{3.7}$$

where, for a given value of m, Γ is a matrix that specifies how the power in the ℓ' mode contribute to that associated to the ℓ mode (their expression is the subject of the next section). C^{ij}_ℓ is the *theoretical* angular power

spectrum:

$$C_\ell^{ij} = \int_0^\infty k^2 P_{\rm matter}(k, z=0) F_\ell^i(k) F_\ell^j dk \tag{3.8}$$

Each kernel $F_\ell^i(k)$ can be shown to be equal to:

$$F_\ell^i(k) \equiv \sqrt{\frac{2}{\pi}} \int_0^\infty dr r^2 \phi_i(r) b(z(r)) D(z(r)) j_\ell(kr) \tag{3.9}$$

where again ϕ_i is zero outside the z-bin as in the **Eq. 3.1**. We can now rewrite the selection function in the ith bin in terms of the one of the full sample ($\phi(\vec{r})$) together with a function that characterizes the redshift bin, a **window function** W_i: $\phi_i(\vec{r}) = \phi(\vec{r}) W_i$. Moreover, using the **Eq. 2.18**, we can then write:

$$F_\ell^i(k) \sim \sqrt{\frac{2}{\pi}} \int_0^\infty dz \frac{dN}{dz} W_i(z) b(z) D(z) j_\ell(kr(z)) \tag{3.10}$$

In the limit of a very thin redshift slice, the selection function becomes:

$$\frac{dN_i}{dz} \approx \frac{\delta^K(z - z_i)}{N_{\rm gal}^i} \tag{3.11}$$

where δ^K is the Kronecker delta, the kernel becomes:

$$F_\ell^i(k) = \sqrt{\frac{2}{\pi}} \frac{1}{N_{\rm gal}^i} W_i(z) b(z_i) D(z_i) j_\ell(kr(z_i)) \tag{3.12}$$

and the corresponding power spectrum:

$$C_\ell^{ij} = \frac{b(z_i) D(z_i) b(z_j) D(z_j)}{N_{\rm gal}^i N_{\rm gal}^j} \int_0^\infty k^2 P_{\rm matter}(k) j_\ell(kr(z_i)) j_\ell(kr(z_j)) dk \tag{3.13}$$

3.1.1 Limber approximation

For $\ell \gg 1$, i.e. at small angular separations, the angular power spectrum can be written in an approximated form called **Limber approximation** (Limber, 1953). Under this condition, in fact, the spherical Bessel functions can be expressed in terms of their asymptotic limit:

$$\lim_{l \gg 1} j_l(x) = \sqrt{\frac{\pi}{2\ell + 1}} \delta^D(\ell + \frac{1}{2} - x) \tag{3.14}$$

where δ^D is the Dirac delta function. Using this equation the angular power spectrum becomes:

$$C_\ell^{ij} = \int_0^\infty dz \frac{dN_i}{dz} \frac{dN_j}{dz} b(z)^2 D(z)^2 P_{\text{matter}} \left(k = \frac{l + 1/2}{r(z)} \right) \tag{3.15}$$

For a very thin redshift interval (**Eq. 3.11**) this last relation can be written as:

$$C_l^{ij} \approx \frac{(D(z_i)b(z_i))^2}{N_{\text{gal}}^i N_{\text{gal}}^j} P_{\text{matter}} \left(k = \frac{l + 1/2}{r(z_i)} \right) \delta^K(z_i - z_j) \tag{3.16}$$

This means that under the Limber approximation there is no correlation between redshift bins.

3.1.2 The mixing matrix

The terms $\Gamma_{\ell\ell'}^m$ appearing in **Eq. 3.7** are defined as:

$$\Gamma_{\ell\ell'}^m = \sum_{m'} \int d\hat{r} Y_{\ell m}^*(\hat{r}) M(\hat{r}) Y_{\ell' m'}(\hat{r}) \int d\hat{r}' Y_{\ell m}^*(\hat{r}') M(\hat{r}') Y_{\ell' m'}(\hat{r}') \tag{3.17}$$

Taking the average over the m-modes, one obtains:

$$R_{\ell\ell'} = \frac{1}{2\ell + 1} \sum_m \Gamma_{\ell\ell'}^m \tag{3.18}$$

where $R_{\ell\ell'}$ is called **angular mixing matrix** and relates the measured angular power spectrum (**Eq. 3.3**) with the theoretical one through a convolution (see Peebles (1973) and Thomas et al. (2011)):

$$\tilde{C}_\ell^{ij} = \sum_{\ell'} R_{\ell\ell'} C_{\ell'}^{ij} \tag{3.19}$$

The mixing matrix can be rewritten by decomposing the mask in spherical harmonics with coefficients $M_{\ell m}$:

$$M(\hat{r}) = \sum_{\ell m} M_{\ell m} Y_{\ell m}(\hat{r}) \tag{3.20}$$

where:

$$M_{\ell m} = \int d\hat{r} M(\hat{r}) Y_{\ell m}^*(\hat{r}) \tag{3.21}$$

If one defines (not to be confused with the bin size W_i used previously):

$$W_\ell \equiv \frac{1}{2\ell+1}\sum_m |M_{\ell m}|^2 \tag{3.22}$$

then it can be shown (Hivon et al., 2002) that the mixing matrix can be written in terms of the Wigner symbols:

$$R_{\ell\ell'} = \frac{2\ell'+1}{4\pi}\sum_{\ell''}(2\ell''+1)W_{\ell''}\begin{pmatrix}\ell & \ell' & \ell''\\ 0 & 0 & 0\end{pmatrix}^2 \tag{3.23}$$

For a full-sky sample, the mask is equal to unity everywhere ($M(\hat{r}) = 1$), so that $M_{\ell m} = \sqrt{4\pi}\,\delta_{\ell 0}\delta_{m0}$, $W_\ell = 4\pi\delta_{\ell 0}$ and:

$$R_{\ell\ell'} = (2\ell'+1)\begin{pmatrix}\ell & \ell' & 0\\ 0 & 0 & 0\end{pmatrix}^2 = \delta^K_{\ell\ell'} \tag{3.24}$$

with $\delta^K_{\ell\ell'}$ the Kronecker delta. In this case:

$$\tilde{C}^{ij}_\ell = C^{ij}_\ell \tag{3.25}$$

The mixing matrix quantifies the alteration in the shape of the angular power spectrum due to a partial sky coverage. **Eq. 3.19** tells us that the presence of a mask introduces correlation between different ℓ modes because it mixes the otherwise orthonormal spherical harmonic coefficients. The mixing matrix only depends on the angular mask and not on the radial selection function whose effects are encoded in the kernels F_ℓ.

It is important to note that **Eq. 3.19** does not provide an unbiased estimate of the angular power spectrum. To illustrate this point let us consider the case of a constant underlying power spectrum C. In this case the **Eq. 3.19** becomes:

$$\tilde{C}^{ij}_\ell = C\sum_{\ell'} R_{\ell\ell'} \tag{3.26}$$

If we define:

$$J_{\ell m} \equiv \int d\hat{r} M(\hat{r})|Y_{\ell m}(\hat{r})|^2 = \int_{\Delta\Omega} d\hat{r}|Y_{\ell m}(\hat{r})|^2 \tag{3.27}$$

where $\Delta\Omega$ is the total surveyed area, then we have:

$$\tilde{C}_\ell^{ij} = \frac{C}{2\ell+1}\sum_m J_{\ell m} = C f_{\rm sky} \tag{3.28}$$

with $f_{\rm sky}$ sky fraction, i.e. the ratio between the observed (masked) area and the total area. The estimate provided by **Eq. 3.28** is biased by the constant factor $f_{\rm sky}$. To eliminate this bias we need to renormalize the mixing matrix $R_{\ell\ell'}$. Peebles (1973) and Hauser and Peebles (1973) suggested the following definition:

$$\langle \delta^i_{\ell m}\delta^{*j}_{\ell m}\rangle = J_{\ell m}\sum_{\ell'}\tilde{\Gamma}^m_{\ell\ell'}C^{ij}_{\ell'} \tag{3.29}$$

with:

$$\tilde{\Gamma}^m_{\ell\ell'} \equiv \frac{\Gamma^m_{\ell\ell'}}{J_{\ell m}} \tag{3.30}$$

we can define a pseudo-power spectrum $\tilde{D}_{\ell m}$ as:

$$\tilde{D}^{ij}_{\ell m} \equiv \frac{\langle \delta^i_{\ell m}\delta^{*j}_{\ell m}\rangle}{J_{\ell m}} = \sum_{\ell'}\tilde{\Gamma}^m_{\ell\ell'}C^{ij}_{\ell'} \tag{3.31}$$

such that its m-average is the convolution of the underlying power spectrum with a renormalized mixing matrix $\tilde{R}_{\ell\ell'}$:

$$\tilde{D}^{ij}_{\ell} \equiv \frac{1}{2\ell+1}\sum_m D^{ij}_{\ell m} = \sum_{\ell'}\tilde{R}_{\ell\ell'}C^{ij}_{\ell'} \tag{3.32}$$

The normalized mixing matrix is:

$$\begin{aligned}\tilde{R}_{\ell\ell'} &= \frac{1}{2\ell+1}\sum_{m=-\ell}^{m=+\ell}\frac{\Gamma^m_{\ell\ell'}}{J_{lm}} = \\ &= \frac{1}{2\ell+1}\sum_{m=-\ell}^{m=+\ell}\left(\frac{\sum_{m'}\int d\hat{r}Y^*_{\ell m}(\hat{r})M(\hat{r})Y_{\ell' m'}(\hat{r})\int d\hat{r}'Y_{\ell m}(\hat{r}')M(\hat{r}')Y^*_{\ell' m'}(\hat{r}')}{d\hat{r}M(\hat{r})|Y_{\ell m}(\hat{r})|^2}\right)\end{aligned} \tag{3.33}$$

with $\sum_{\ell'}\tilde{R}_{\ell\ell'} = 1$. Coming back to the example of the white noise, we obtain $\tilde{D}^{ij}_\ell = C$, i.e. the power spectrum $\tilde{D}$ is effectively unbiased. However, the unbiased estimator for the power spectrum comes with a price,

and it consists in losing the possibility of expressing the mixing matrix in terms of the Wigner symbols. To circumvent this problem Peebles (1973) proposed an approximated estimator for the mixing matrix:

$$\begin{aligned}\tilde{R}_{\ell\ell'} &= \frac{1}{2\ell+1}\sum_{m=-\ell}^{m=+\ell}\frac{\Gamma^m_{\ell\ell'}}{J_{lm}} \approx \\ &\approx \frac{1}{2\ell+1}\frac{\sum_{m=-\ell}^{m=+\ell}\Gamma^m_{\ell\ell'}}{\sum_{m=-\ell}^{m=+\ell}J_{\ell m}} = \frac{(2\ell'+1)}{4\pi f_{\rm sky}}\sum_{\ell''}(2\ell''+1)W_{\ell''}\begin{pmatrix}\ell & \ell' & \ell''\\ 0 & 0 & 0\end{pmatrix}^2\end{aligned} \tag{3.34}$$

We shall use this approximation in our analysis and quantify its accuracy in the following section.

3.2 The angular cross power spectrum estimator

Now that we have a deeper understanding of the modification induced on the angular power spectrum by a partial sky coverage, we can introduce our cross angular spectrum estimator.

Let $\sigma_i(\hat{r})$ be the (observed) galaxy surface density in a redshift bin i and $\bar{\sigma}_i(\hat{r})$ the ensamble average of the galaxy surface density, which we assume to be equal to the mean surface galaxy density in the direction $\vec{r}$ and in the redshift bin i. Let us define the quantity $F_i(\hat{r})$ as the fluctuation in the surface density:

$$F_i(\hat{r}) \equiv \frac{1}{\bar{\sigma}_i(\hat{r})}(\sigma_i(\hat{r}) - \bar{\sigma}_i(\hat{r})) \tag{3.35}$$

To expand this field into spherical harmonics we insert the **Eq. 3.35** into **Eq. 3.2**:

$$a^i_{\ell m} = \int d\hat{r} Y^*_{\ell m}(\hat{r}) F_i(\hat{r}) \equiv A^i_{\ell m} - I^i_{\ell m} \tag{3.36}$$

where the term $A^i_{\ell m}$ is the spherical harmonics decomposition of the surface density:

$$A^i_{\ell m} = \int d\hat{r}\frac{1}{\bar{\sigma}_i(\hat{r})}Y^*_{\ell m}(\hat{r})\sigma_i(\hat{r}) \tag{3.37}$$

while $I^i_{\ell m}$ is its ensamble average:

$$I^i_{\ell m} = \langle A^i_{\ell m} \rangle = \int d\Omega Y^*_{\ell m}(\hat{r}) \tag{3.38}$$

For a full sky sample, we have $I^i_{\ell m} = 4\pi\sqrt{4\pi}\delta_{\ell 0}\delta_{m0} = \sqrt{4\pi}\delta_{\ell 0}\delta_{m0}$ and $a_{\ell m} = A_{\ell m}$ for $\ell > 0$.

Let us now compute the ensamble average of the product of the harmonic coefficients:

$$\begin{aligned} &\tilde{C}^{ij}_{\ell m} \equiv \langle a^i_{\ell m} a^{*j}_{\ell m} \rangle = \\ &= \int d\hat{r} \frac{1}{\bar{\sigma}_i(\hat{r})\bar{\sigma}_j(\hat{r})} \hat{r}' Y^*_{\ell m}(\hat{r}) Y_{\ell m}(\hat{r}') \langle \, (\sigma_i(\hat{r}) - \bar{\sigma}_i(\hat{r}))(\sigma_j(\hat{r}') - \bar{\sigma}_j(\hat{r}')) \, \rangle \end{aligned} \tag{3.39}$$

When it comes to measured quantities, the surface density is estimated from discrete objects counts. To model the latter we assume that galaxies are a Poisson random process so that we can write:

$$\begin{aligned} \langle (\sigma_i(\hat{r}) - \bar{\sigma}_i(\hat{r}))(\sigma_j(\hat{r}') - \bar{\sigma}_j(\hat{r}')) \rangle = \bar{\sigma}_i(\hat{r}) \delta(\hat{r} - \hat{r}') \delta^K_{ij} + \\ + \bar{\sigma}_i(\hat{r}) \bar{\sigma}_j(\hat{r}) \, \omega_{ij} \left(\hat{r} \cdot \hat{r}' \right) \end{aligned} \tag{3.40}$$

where δ^K_{ij} is the Kronecker delta that arises from the assumption that the different redshift bins are disjoint; $\omega_{ij} \left(\hat{r} \cdot \hat{r}' \right)$ is the angular correlation function while $\delta(\hat{r} - \hat{r}')$ is the two dimensional Dirac delta function on a sphere. As we have seen in the previous Chapter, the (cross) angular correlation function is related to the (cross) angular power spectrum by a Legendre transform (the inverse of **Eq. 2.11**):

$$\omega_{ij} \left(\hat{r} \cdot \hat{r}' \right) = \frac{1}{4\pi} \sum_\ell (2\ell + 1) C^{ij}_\ell P_\ell(\hat{r}) \cdot \hat{r}' = \sum_{\ell m} C^{ij}_\ell Y_{\ell m}(\hat{r}) Y^*_{\ell m}(\hat{r}') \tag{3.41}$$

where P_ℓ is the Legendre polynomial and we have used the addition theorem (**Eq. 2.21**). If we assume that the angular dependence of the mean surface density is fully captured by the mask, i.e. if the mean galaxy density of the sample is constant within the surveyed area, then we have:

$$\bar{\sigma}_i(\hat{r}) = \bar{\sigma}_i M(\hat{r}) \tag{3.42}$$

where $\bar{\sigma}_i = N_i/\Delta\Omega$, with N_i number of objects in the ith redshift bin and $\Delta\Omega$ the area of the surveyed region in the sky. Since the value of the mask is equal to 1 inside the surveyed area and 0 outside it, then $\bar{\sigma}_i$ drops outside the integral and we can express **Eq. 3.39** as:

$$\tilde{C}^{ij}_{\ell m} = \frac{1}{\bar{\sigma}_i}\delta^K_{ij} J_{\ell m} + \sum_{\ell'} \Gamma^m_{\ell\ell'} C^{ij}_{\ell'} \tag{3.43}$$

where $\Gamma^m_{\ell\ell'}$ and $J_{\ell m}$ are defined in the **Eq. 3.7** and **3.27** respectively. As we have seen, **Eq. 3.43** provides a biased estimate of the angular spectrum. We can define an unbiased quantity:

$$D^{ij}_{\ell m} \equiv \frac{\tilde{C}^{ij}_{\ell m}}{J_{\ell m}} = \frac{1}{\bar{\sigma}_i}\delta^K_{ij} + \sum_{\ell'} \tilde{\Gamma}^m_{\ell\ell'} C^{ij}_{\ell'} \tag{3.44}$$

where $\tilde{\Gamma}^m_{\ell\ell'}$ expressed in **Eq. 3.30**. We can therefore form an unbiased estimator from the measured spherical harmonics coefficient, mean surface density and angular mask:

$$\tilde{D}^{ij}_{\ell m} \equiv \frac{|(a^i_{\ell m} a^{*j}_{\ell m})|}{J_{\ell m}} - \frac{1}{\bar{\sigma}_i}\delta^K_{ij} \tag{3.45}$$

The ensamble average of this estimator is:

$$\langle \tilde{D}^{ij}_{\ell m} \rangle = \sum_{\ell'} \tilde{\Gamma}^m_{\ell\ell'} C^{ij}_{\ell'} \tag{3.46}$$

Averaging over the m modes we obtain:

$$\tilde{D}^{ij}_{\ell} = \frac{1}{2\ell+1}\sum_m \frac{|(a^i_{\ell m} a^{*j}_{\ell m})|}{J_{\ell m}} - \frac{1}{\bar{\sigma}_i}\delta^K_{ij} \tag{3.47}$$

The second term of this equation is the *shot-noise*, which is zero if we consider different redshift bins ($i \neq j$) and equal to $\Delta\Omega/N_i$ otherwise. The ensamble average of $\tilde{D}$ is:

$$\langle \tilde{D}^{ij}_{\ell} \rangle = \sum_{\ell'} \tilde{R}_{\ell\ell'} C^{ij}_{\ell'} \tag{3.48}$$

where $\tilde{R}_{\ell\ell'}$ is the renormalized mixing matrix defined in **Eq. 3.33**.

Several authors (Blake et al. (2004), Blake et al. (2007), Thomas et al. (2011)) have implemented the estimator in **Eq. 3.45** using the mixing matrix in **Eq. 3.34**, which we have seen to be an approximated form of the exact mixing matrix in **Eq. 3.33**. Peebles (1973) and Hauser and Peebles (1973) have proposed an alternative estimator for the angular power spectrum. Let us call it $\tilde{K}$:

$$\tilde{K}_\ell^{ij} = \frac{1}{f_{\rm sky}(2\ell+1)} \sum_{m=-\ell}^{m=+\ell} |a_{\ell m}^i a_{\ell m}^{*j}| - \frac{1}{\bar{\sigma}_i}\delta_{ij}^K \tag{3.49}$$

The ensamble average in this case turns out to be:

$$\langle \tilde{K}_\ell^{ij} \rangle = \sum_{\ell'} \frac{R_{\ell\ell'}}{f_{\rm sky}} C_{\ell'}^{ij} \tag{3.50}$$

Here the mixing matrix is the one defined by **Eq. 3.23** with no approximations: in case of a simple binary mask all one needs to do is to divide $R_{\ell\ell'}$ by the observed sky fraction $f_{\rm sky}$, as the **Eq. 3.34** suggests.

The difference between the two estimators D and K can be appreciated in **Fig. 3.1** .The top panel shows the ratio between the mean over 10^4 realizations of the angular spectrum computed with the estimator D (**Eq. 3.47**) and the corresponding model obtained through the mean of the convolution of the theoretical spectrum (obtained from `CLASS`) with the mixing matrix following the **Eq. 3.45**. The middle panel shows the same but for the estimator K of **Eq. 3.49** and the expected spectrum (**Eq. 3.50**). Finally, the bottom panel shows the per cent difference between the two estimators. The mixing matrix has been computed with our code from the mask provided by Maciej Bilicki (Bilicki et al., 2014) (see **Fig. 3.8**). As expected, the two estimators generate a different pseudo power spectra because they are linked to the underlying power spectrum through a different mixing matrix. Nevertheless, it can be noted from the figure that the difference between the two estimates is smaller than the expected (gaussian) variance represented by the shadowed areas. In other words, the choice of using the D estimator with the approximated mixing matrix of **Eq. 3.34** is reasonable even more so if one consider that the exact version of the mixing

matrix is more complicated to compute. In this sense, the decision of Blake et al. (2004), Blake et al. (2007) and Thomas et al. (2011) can be justified. We thus decided to follow these authors and use the estimator D.

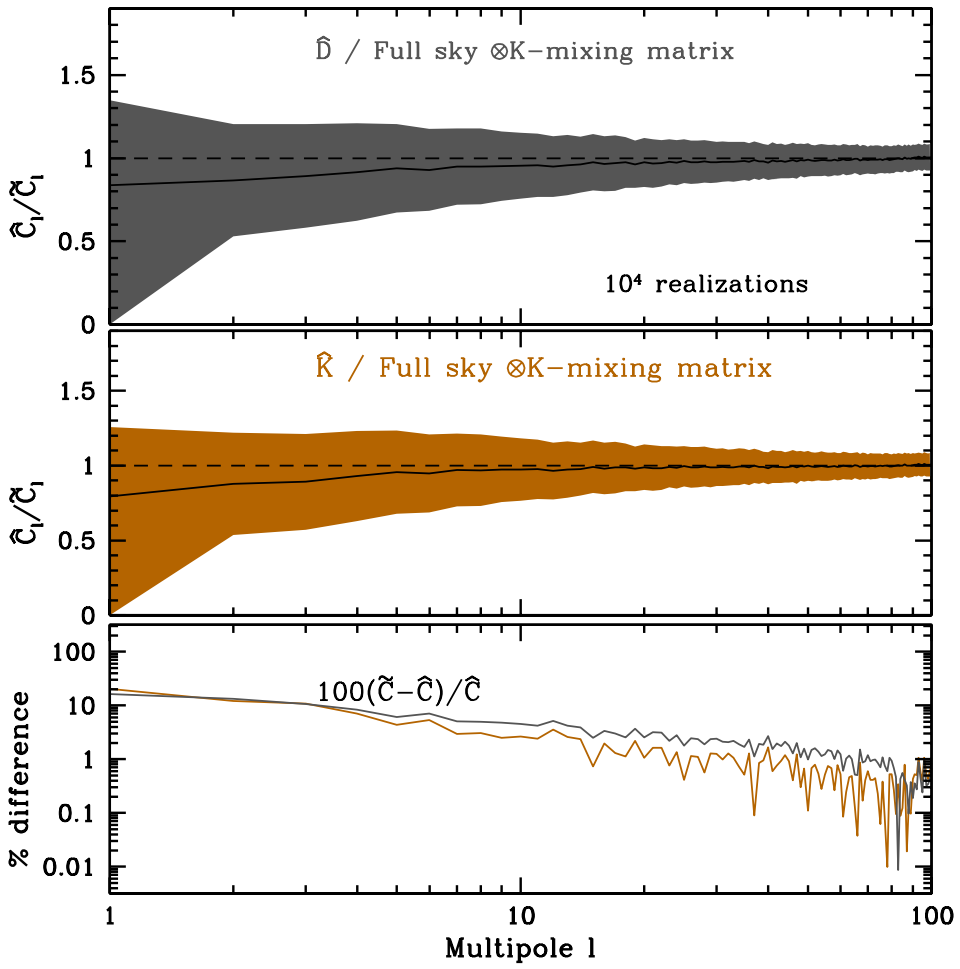

Figure 3.1: Comparison between the estimators of pseudo power spectrum D and K described in the text. The shadowed areas in the two upper panels represent the variance of the ensemble of 10^4 maps. The bottom panel shows the per cent difference between the estimators.

3.3 The numerical code to estimate the angular spectrum

The estimator for the angular spectrum that will be used in this thesis is the one specified by **3.47**, D_ℓ. The estimator applies to a continuous 2D field whereas in our case we have a distribution of discrete points on the 2-sphere. The commonly adopted solution is to pixelize the counts. The angular resolution of the resulting map is set by the pixels size (assumed to be equal across the sphere) and is determined by the trade-off between shot-noise error and the need to measure the spectrum down to the smallest possible scales to extract the maximum amount of information.

In our case, the pixelization is performed through the publicy available library `HEALPix` (Górski et al., 2005) that we briefly described in the following section[1].

3.3.1 HEALPix

The library `HEALPix` (acronym for Hierarchical Equal Area isoLatitude Pixelization of a sphere, see Górski et al. (2005)) has been developed by the JPL at NASA. Created to measure the cosmic microwave background anisotropy, this numerical toolkit has then been adapted for many more applications, and produces a subdivision of a spherical surface in which each pixel covers the same surface area as every other pixel. The resolution of the pixelized map is set by the parameter n_{side}, which provides the dimension of the single pixel. The number of pixels on the sphere is then set to $12n_{side}^2$. As the resolution increases, lower angular scales (i.e. higher multipole) can be analyzed, with the maximum multipole being $\ell_{\max} = 4n_{side}$ (see **Fig. 3.2**).

[1]More details on this library are available at `http://healpix.jpl.nasa.gov/`.

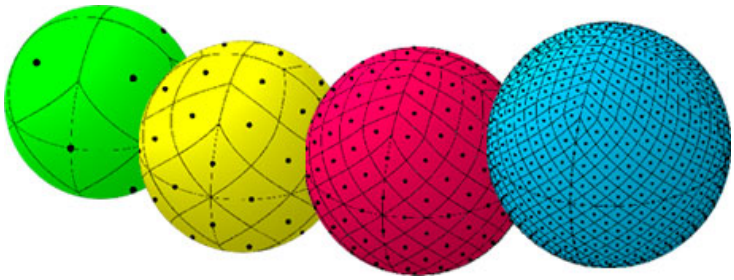

Figure 3.2: Different resolutions in the pixelization provided by `HEALPix`.

In **Table 3.1** the different resolutions allowed by `HEALPix` are shown.

Table 3.1: Different resolutions for the pixelization provided by `HEALPix` (see the text for details).

Resolution	n_{side}	Total number of pixels	$\ell_{\max}$
0	1	12	4
1	2	48	8
2	4	192	16
3	8	256	32
4	16	3072	64
5	32	12288	128
6	64	49152	256
7	128	196608	512
8	256	786432	1024
9	512	3145728	2048
10	1024	12582912	4096

In addition to the resolution, the `HEALPix` user should define the ordering scheme, i.e. the way in which the pixels are numbered. There are two possibilities: *ring* and *nested* scheme. Both numbering schemes map the two dimensional distribution of discrete area elements on a sphere into a one dimensional, integer pixel number array (in our case, for example, each pixel will contain the number of objects in that region). However,

while in the *ring* scheme the pixel number winds down from north to south pole through the consecutive isolatitude rings, in the *nested* scheme the pixel number grows with consecutive hierarchical subdivisions on a tree structure seeded by the twelve base-resolution pixels[2]. In our analysis we always use the *ring* scheme.

3.3.2 The angular spectrum estimator of a pixelized map

Let us first adapt the definition of the quantities defined in **Sec. 3.1** for a continuous field to the case of a pixelized map. The mean surface density is now:

$$\sigma_i(\hat{r}) = \sum_{k=1}^{N_{\rm pix}} \mathcal{N}_{ik}\, \delta_D^2(\hat{r} - \hat{r}_k) \tag{3.51}$$

where $\mathcal{N}_{ik}$ is the number of galaxies that belongs to the i-th redshift bin and the k-th pixel on the sphere, while δ_D^2 is the Dirac delta function on the 2-sphere. Similarly, the angular mask is:

$$M(\hat{r}) = \sum_{k=1}^{N_{\rm pix}} m_k\, \delta_D^2(\hat{r} - \hat{r}_k)\Delta\Omega p \tag{3.52}$$

where m_k is either 1 or 0 depending on whether the pixel belongs or not to the surveyed area; $\Delta\Omega p$ is the pixel area (equal for all the pixels, as we have seen). The term $A^i_{\ell m}$ in **Eq. 3.37** becomes:

$$A^i_{\ell m} = \frac{1}{\hat{\sigma}_i}\int d\hat{r} Y^*_{\ell m}(\hat{r})\sigma_i(\hat{r}) = \frac{1}{\hat{\sigma}_i}\sum_{k=1}^{N_{\rm pix}} \mathcal{N}_{ik} Y^*_{\ell m}(\hat{r}_k) = \Delta\Omega_p \sum_{k=1}^{N_{\rm pix}} \left(\frac{\mathcal{N}_{ik}}{\bar{\mathcal{N}}_i}\right) Y^*_{\ell m}(\hat{r}_k) \tag{3.53}$$

where $\hat{\mathcal{N}}_i$ is the mean number of objects per pixel in the i-th redshift bin. Analogously we can rewrite the terms $I_{\ell m}$ of **Eq. 3.38** and $J_{\ell m}$ of **Eq. 3.27**:

$$I^i_{\ell m} = \frac{1}{\hat{\sigma}_i}\int d\hat{r} Y^*_{\ell m}(\hat{r})\hat{\sigma}_i(\hat{r}) = \int d\hat{r} Y^*_{\ell m}(\hat{r}) M(\hat{r}) = \Delta\Omega_p \sum_{k=1}^{N_{\rm pix}} m_k\, Y^*_{\ell m}(\hat{r}_k) \tag{3.54}$$

[2]For more details, see *http://healpix.jpl.nasa.gov/html/intronode4.htm*

and:

$$J^i_{\ell m} = \int d\hat{r} M(\hat{r}) |Y_{\ell m}(\hat{r})|^2 = \Delta\Omega_p \sum_{k=1}^{N_{\rm pix}} m_k |Y_{\ell m}(\hat{r})|^2 \tag{3.55}$$

If the angular mask is the same in all redshift bins, which is the case that is relevant for this thesis, then neither $I_{\ell m}$ nor $J_{\ell m}$ depends on the redshift. Finally, the spherical harmonic coefficients of **Eq. 3.36** are:

$$a^i_{\ell m} = \Delta\Omega_p \sum_{k=1}^{N_{\rm pix}} \left(\frac{\mathcal{N}_{ik} - \bar{\mathcal{N}}_i}{\bar{\mathcal{N}}_i} \right) Y^*_{\ell m}(\hat{r}) \tag{3.56}$$

As a result, the estimator of the angular power spectrum, D, can be built from the measured harmonics coefficients **Eq. 3.56**, the $I_{\ell m}$ and $J_{\ell m}$ terms estimated as in **Eqs. 3.54** and **3.55** and the mean number density of objects defined in **Eq. 3.51**.

The library `HEALPix` not only provides a pixelization of the sphere, i.e. a one dimensional array whose elements corresponds to the value of the field in each pixel (which we call the `map`), but also includes several functions related to the decomposition of this field in spherical harmonics. One of the function we used in our code is called `map2alm`, and provides the spherical harmonic coefficients of the pixelized map. In our case, if the `map` is the one dimensional integer array obtained from the surface density field, then the `alm` would be those expressed in **Eq. 3.37**. If, instead, one consider the fluctuations in the surface density, then `map2alm` would directly provide the coefficients expressed in **Eq. 3.36**. However, in both cases we would add the contribution of the term $J_{\ell m}$ to take into account for the mask, as expressed by **Eq. 3.45** or equivalently by **Eq. 3.49**.

In this thesis we have used the estimator D to measure the angular cross spectrum of a galaxy density fluctuation field on a sphere. For this purpose we developed and implemented our own C++ code that exploits some of the `HEALPix` routines to estimate the angular power spectrum of galaxies distributed over some patch of the sky. The structure of the code is as follows.

Inputs:

1. The code reads in a galaxy catalog that specifies the angular position of each object in one or more redshift bins. It can also import a density or overdensity `HEALPix` map in fits format.

2. A binary angular mask in `HEALPix` format (with specified pixel size).

Execution steps:

1. The mean number density of objects in each redshift shell is computed.

2. For each shell, an `HEALPix` map is created that specifies the density fluctuation in the number of objects within each pixel. The pixel size is set to match that of the angular mask.

3. The `HEALPix` routine `map2alm` is called to estimate the spherical harmonics coefficients in each shell.

4. The terms $I_{\ell m}$ and $J_{\ell m}$ are computed through **Equations 3.54 - 3.55**. For each shell, the angular power spectrum is then obtained from **Eq. 3.53**.

5. The code also estimates the mixing matrix corresponding to the input mask (**Eq. 3.34**).

6. It is also possible to compute the angular spectrum and the mixing matrix in ℓ-bins.

Outputs:

1. The angular power spectrum

2. The mixing matrix associated to the geometric mask, if provided.

3.3.3 The pixel window function

Unavoidably, pixelization removes power below the angular resolution of the map. As a result, the shape of the angular power spectrum is modified on small scales. Although the exact impact is often difficult to treat, some assumptions can be made in order to quantify the effect. Our approach is to mimic the pixelization effect on the theoretical spectrum before comparing it with the measured one. In practice, we multiply the theoretical spectrum by the so called **pixel widow function** defined below.

The pixelized signal $f(p)$ is just the average signal within the pixel size $\Omega_{\rm pix}$:

$$f(p) = \int du\ \omega_p(u) f(u) \tag{3.57}$$

where $\omega_p(u)$ is equal to $1/\Omega_{\rm pix}$ within the pixel and 0 otherwise, so that the integral over all the pixels is normalized: $\int du\ \omega_p(u) = 1$. We expand $f(p)$ in spherical harmonics:

$$f(p) = \sum_{\ell=0}^{\ell_{\rm max}} \sum_m a_{\ell m} \omega_{lm}(p) \tag{3.58}$$

where $\omega_{\ell m}(p)$ is the spherical harmonic transform of the pixel p:

$$\omega_{\ell m}(p) = \int du\ \omega_p(u) Y_{\ell m}(u) \tag{3.59}$$

The exact structure of the pixels makes the **Eq. 3.59** computationally difficult to handle. However, since the pixel size is much smaller than the angular correlation scale of the sample, we can average over the m-modes and obtain:

$$\omega_\ell(p) = \left(\frac{4\pi}{2\ell+1} \sum_{m=-\ell}^{\ell} |\omega_{\ell m}(p)|^2 \right)^{1/2} \tag{3.60}$$

so that we can define:

$$\omega_{\ell m}(p) = \omega_\ell(p) Y_{\ell m}(p) \tag{3.61}$$

Note that the quantity $\omega_\ell(p)$ does not depend on the pixel location in the sky. Since pixels in the `HEALPix` maps have the same area, then the relation between the spectrum of the pixelized map, $C_\ell^{\rm pix}$, and the underlying

angular spectrum, $C_\ell^{\rm unpix}$, is simply:

$$C_\ell^{\rm pix} = \omega_\ell^2 C_\ell^{\rm unpix} \tag{3.62}$$

where the effective *pixel window function* is defined as:

$$\omega_\ell = \left(\frac{1}{N_{\rm pix}} \sum_{p=0}^{N_{\rm pix}-1} \omega_\ell^2(p) \right)^{1/2} \tag{3.63}$$

Eq. 3.62 provides the relation between the pixelized and the unpixelized angular spectrum: when comparing our measure with the model, we will have to multiply this latter by $\omega_\ell^2(p)$. The shape of the pixel window function for a `HEALPix` map with $n_{side} = 64$, corresponding to the resolution adopted in our analysis, is shown in **Fig. 3.3**. **Fig. 3.4** shows instead the impact of this pixel function on the angular power spectrum.

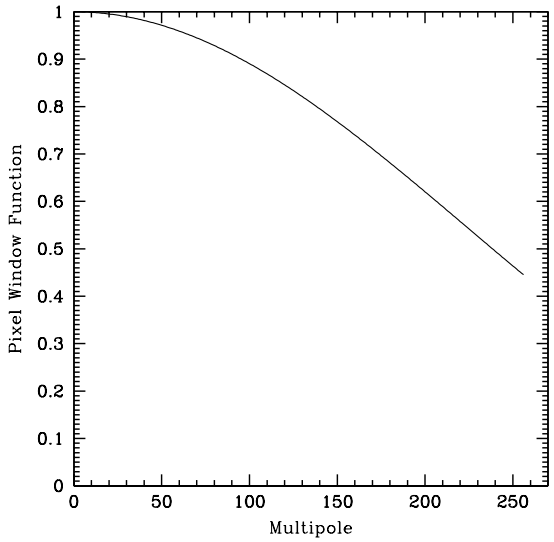

Figure 3.3: The pixel window function $\omega_l(p)$ provided by `HEALPix` for $n_{side} = 64$.

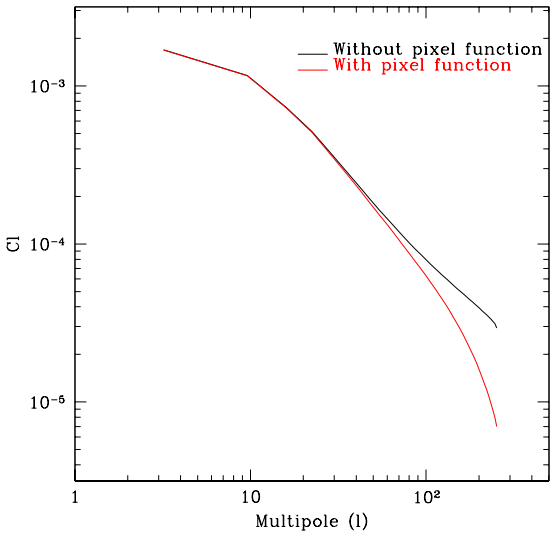

Figure 3.4: The impact of the pixel window function $\omega_l(p)$ on the angular power spectrum.

3.4 Assessing the code performance

Now that we have described the estimator and its numerical implementation, we assess its performance by running a series of tests to evaluate its precision and accuracy. In doing so we will focus on the ability to account for the effect of a mask on our measure.

3.4.1 Test 1: Recovering a known angular spectrum

In this first test we assess the ability of our estimator to recover the correct angular power spectrum for a 2D Gaussian realization of a fluctuation field characterized by a well known set of C_ℓ. This set is generated by the code `CLASS`, which solves the Boltzmann equation in a pre-defined cosmological scenario, as we have seen in **Section 2.2.4**.

The test is as follows:

1. Generate the angular power spectrum C_ℓ^{input} of a cosmological model

using `CLASS`. We consider a ΛCDM model with $\Omega_b = 0.04$, $\Omega_{CDM} = 0.21$, $\Omega_\Lambda = 0.75$, $T_{cmb} = 2.7255$, $h = 0.73$, $A_s = 2.46 \cdot 10^{-9}$ (that corresponds to a $\sigma_8 \simeq 0.825$);

2. Assuming a Gaussian random field, generate the 2D map of the associated overdensity field;

3. Apply our $\hat{C}_\ell$ estimator to measure the angular power spectrum of this map. The measured spectrum is not expected to match the theoretical one because of the cosmic variance.

4. For this reason, we repeat the steps 2 and 3 to generate many different realizations of the same angular spectrum;

5. Estimate the mean and the scatter among the C_ℓ measurement. The `rms` scatter quantifies the precision of the estimator. The discrepancy between the mean and the theoretical spectrum quantifies its accuracy.

In this first test we consider a full-sky realization; i.e. the effect of the mask is ignored here.

Given the input angular power spectrum ($C_\ell^{\rm input}$ hereafter), we have generated a set of **Gaussian distributed** spherical harmonic coefficients $a_{\ell m}$ with zero mean and variance equal to $\sqrt{C_\ell^{\rm input}}$ by random sampling the corresponding Gaussian distribution. This was done by using the appropriate routine in the `GSL` libraries:

```
alm(l,m).re = gsl_ran_gaussian(r, sqrt(0.5*cl_th[l]));
alm(l,m).im = gsl_ran_gaussian(r, sqrt(0.5*cl_th[l]));
```

where the first and second line refer to the real and imaginary part of the $a_{\ell m}$, `r` is the seed for the random number generator and `cl_th[l]` is the input power spectrum. Note that if the $a_{\ell m}$ are Gaussian distributed with a variance equal to $\sqrt{C_\ell^{\rm input}}$, then the real and imaginary part of the $a_{\ell m}$

have a variance equal to $\sqrt{C_\ell/2}$ (while for $m = 0$ the imaginary part is set equal to zero).

Once we obtain a set of $a_{\ell m}$, we generate one realization of the 2D map of the corresponding density field using the `HEALPix` function `alm2map`. The resolution of the map is set equal to $n_{side} = 64$. The overall procedure is repeated 10^4 times to obtain a corresponding number of independent realizations of the same Gaussian field and account for the effect of cosmic variance. For each of the 10^4 maps we use the `HEALPix` function `map2alm` (see **Section 3.3.1**) to estimate the $a_{\ell m}$ and then measure the angular spectrum using the **Eq. 3.45** with $J_{\ell m} = 1$ since we are considering the full-sky case. We also ignore the shot-noise term, since we are dealing with an overdensity map and not a catalogue. In this way we obtain the $C_\ell^{\rm fullsky}$ that can be compared with the $C_\ell^{\rm input}$ both for each realization and for the mean over all the realizations, that is:

$$\bar{C_\ell}^{\rm fullsky} = \frac{1}{N_r}\sum_r C_\ell^{{\rm fullsky},r} \tag{3.64}$$

where r runs over all the realizations N_r. Finally we compare the measured $C_\ell^{{\rm fullsky},r}$ and the mean $\bar{C_\ell}^{\rm fullsky}$ with the $C_\ell^{\rm input}$. The result is shown in the top **Panel (a)** of **Fig. 3.5**, where we can see the input angular power spectrum (solid black line) and the mean measured full-sky power spectrum $\bar{C_\ell}^{\rm fullsky}$ (dotted blue line). The shadowed area represents the `rms` scatter over the 10^4 realizations. The two spectra overlap each other, showing that the estimator is unbiased on all scales down to the largest multipole corresponding to the angular resolution of the map. The same result is illustrated in the middle **Panel (b)** of the same figure, where we plot the residual ratio to $C_\ell^{\rm input}$: discrepancies are really small, much smaller than the error on the single estimate quantified by the shaded region. In the top **Panel (a)** of **Fig. 3.5** we also show the results of a sanity check in which the routine `map2alm` is used to compute the harmonics coefficients measured from a map generated with the inverse healpix routine `alm2map`. As expected no discrepancy is found. This is also highlighted in the bot-

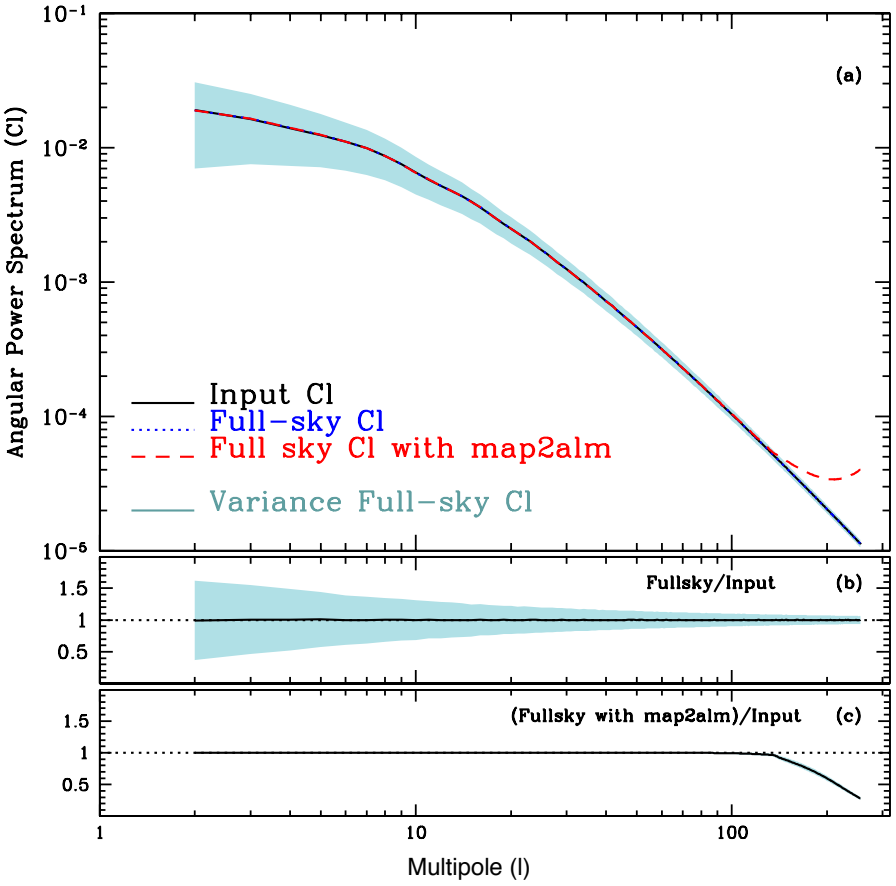

Figure 3.5: Top Panel (a): The solid black line shows the $C_\ell^{\rm input}$; the dotted blue line shows $\bar{C}_\ell^{\rm fullsky}$, while the dashed red line shows the full-sky spectrum measured with the `HEALPix` method `map2alm` as described in the text. The shadowed area represents the variance of the ensamble of the realizations for the full-sky C_ℓ. **Middle Panel (b)**: The mean of the relative ratio between the $C_\ell^{\rm fullsky}$ and the $C_\ell^{\rm input}$. **Bottom Panel (c)**: The relative ratio between the $C_\ell^{\rm fullsky}$ and the full-sky spectrum measured with the `HEALPix` method `map2alm` (see the text).

tom panel. The mismatch at large ℓs reflects the finite pixel size of the map.

To further investigate the effect of the pixalization we repeat the same tests using maps with different resolutions. As we have seen, increasing the resolution, i.e. increasing the size of the pixel (n_{side}), affects multipoles larger than $\ell_{max,\,\texttt{HEALPix}} = 4n_{side}$. In **Fig. 3.6** the $C_\ell^{\rm input}$ is shown, together with the measured power spectra with and without the `HEALPix` method `map2alm` for different resolutions of the map. The results clearly illustrate the effect of changing the resolution: maps with increasing resolutions allows one to measure the spectrum over increasingly smaller angular scales. Below $\ell_{\rm max}$ the estimated spectrum is flat, determined by white-noise like signal.

To better quantify the effect of the resolution and find a quantitative expression for the max value of ℓ out to which one can push the angular correlation analysis, we compare the difference between $C_\ell^{\rm input}$ and the $\bar{C}_\ell^{\rm fullsky}$ with the `rms` scatter over the 10^4 realizations. The upper panel of **Fig. 3.7** shows the $\ell_{\rm max}$ value at which:

$$|C_\ell^{\rm input} - \bar{C}_\ell^{\rm fullsky}| \leq \sigma_{C_\ell} \tag{3.65}$$

with σ_{C_ℓ} being the `rms` scatter, versus the `HEALPix` resolution. The solid red line shows a linear fit with equation:

$$n_{side} = a\ell_{max} + b;\ a \simeq -4.7;\ b \simeq 0.48 \tag{3.66}$$

We can thus use n_{side} or $\ell_{max,\,\texttt{HEALPix}}$ as a proxy to the minimum angular scale for analyzing the spectrum. **Eq. 3.66** clearly depends on the clustering properties of the field, i.e. on the cosmological model adopted. However, within the parameter region allowed by current observations, the dependence is expected to be mild and **Eq. 3.66** can be used heuristically to determine the ℓ_{max} associated to a pixelized map.

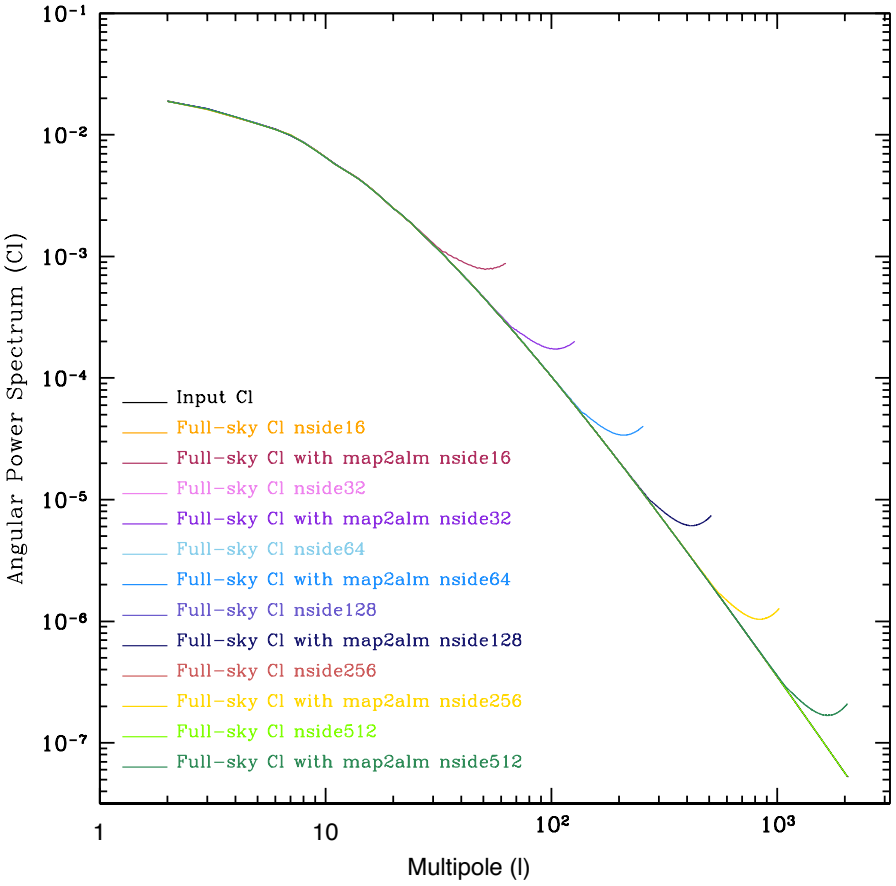

Figure 3.6: The solid black curve shows the C_ℓ^{input}; the other colored solid curves show the measured full-sky spectrum with and without the `HEALPix` method `map2alm` for different resolutions, as explained in the text. The case with $n_{side} = 512$ (dark green) is superimposed to the expected spectrum (black curve). On small scales (i.e. high ℓ values) the effect of the pixalization is more evident.

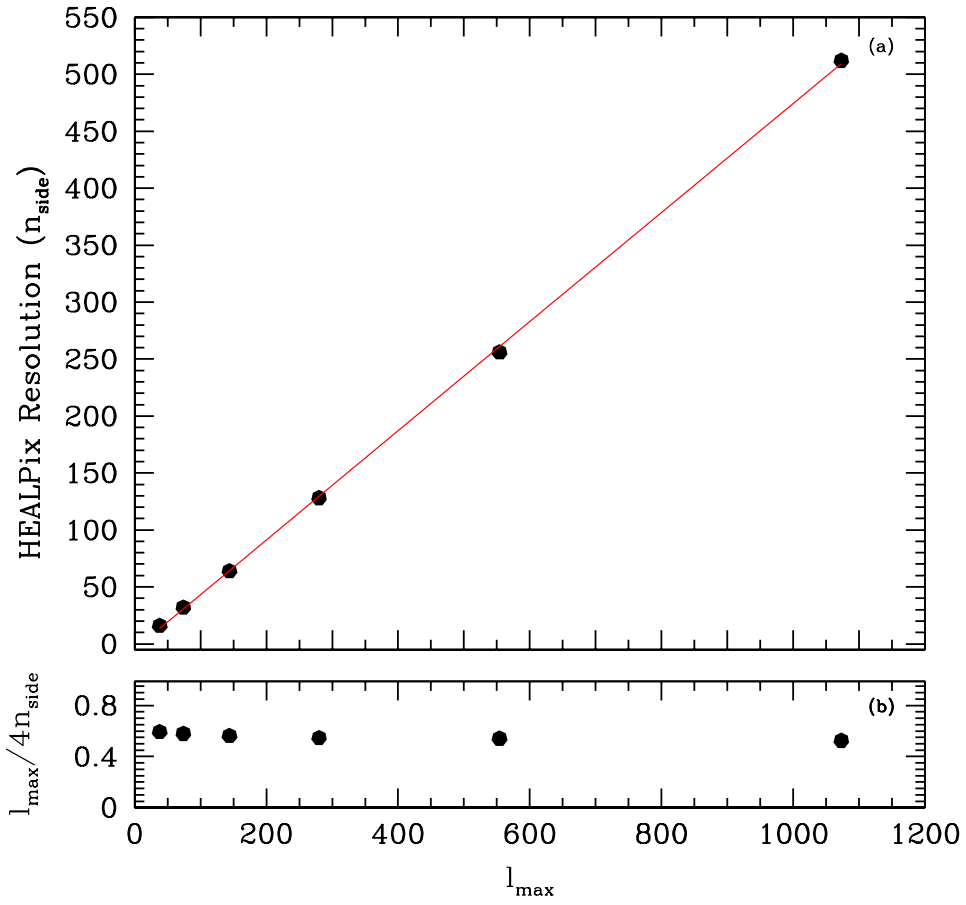

Figure 3.7: Panel (a): The `HEALPix` resolution n_{side} as a function of the ℓ_{max} up to which we can trust our measurement, i.e. the ℓ_{max} corresponding to which the difference between the C_ℓ^{input} and the $\bar{C}_\ell^{\text{fullsky}}$ measured with `map2alm` is within $1-\sigma$ of the expected `rms` scatter Gaussian distribution. The solid red line shows the linear fit with parameters $a = -4.7$; $b = 0.48$ **Panel (b)**: The ratio between the ℓ_{max} up to which we can trust our measurement and the maximum ℓ provided by `HEALPix`, i.e. $\ell_{max,\,\texttt{HEALPix}} = 4n_{side}$.

3.4.2 Test 2: The effect of the geometry mask

As we have seen in the previous sections, a realistic survey does not cover the whole sky, and this fact has a number of consequences. First of all, the surveyed area is reduced, so the cosmic variance is amplified. Secondly, the effect of a mask is to mix the phases of the $a_{\ell m}$ coefficients, that, therefore, are not independent. To account for these effects we have thus introduced the mixing matrix (**Eq. 3.18**), and we have also showed that in order to have an unbiased estimator of the angular power spectrum we should use **Eq. 3.47** with the mixing matrix showed in the **Eq. 3.34**, that provides a good approximation to its exact version **Eq. 3.23** for the K and the D angular spectrum estimators.

The goal of this test is to compare the angular power spectrum estimated in presence of an angular mask (to which we will refer to as C_ℓ^{masked}) with theoretical expectations provided by the reference power spectrum convolved with the estimated mixing matrix.

We performed the following test:

1. Generate a `HEALPix` map of the geometry mask; in particular, we use the mask provided by Maciej Bilicky (Bilicki et al., 2014) in `HEALPix` format with $n_{side} = 64$, showed in **Fig. 3.8**. The effect of masking out the region around the Galactic plane is that of removing power on large angular scales. The fraction of unmasked sky is $f_{sky} = 0.692708$.

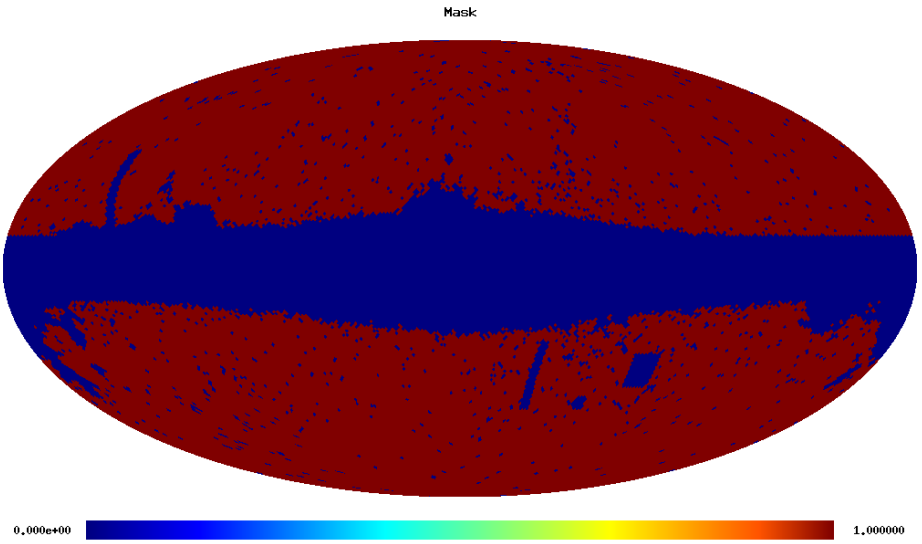

Figure 3.8: The geometry mask we use to perform the test described in this section. The surveyed area is associated with the mask value 1 (in red), while the pixels that are not observed are associated with the mask value 0 (in blue). This mask was provided by Maciej Bilicki (Bilicki et al., 2014) and has $f_{sky} = 0.692708$.

2. Compute the angular spectrum of the mask, $I_{\ell m}$ (**Eq. 3.38**);
3. Compute the mixing matrix $R_{\ell\ell'}$ (**Eq. 3.34**);
4. Generate an all sky map with an input power spectrum, C_ℓ^{input} using the `alm2map` routine;
5. Estimate C_ℓ^{fullsky} from the map using our code;
6. Mask out the blue regions in **Fig. 3.8**;
7. Estimate the C_ℓ^{masked} (**Eq. 3.47**) using our code;
8. Compute the theoretical spectrum convolved with the mixing matrix $R_{\ell\ell'}\, C_\ell^{\text{conv}}$ using (**Eq. 3.45**);
9. Go back to step 4 and repeat for 10^4 realizations;

10. Compare the mean of $C_\ell^{\rm conv}$ and $C_\ell^{\rm masked}$ and their scatter.

The results of the tests are summarized in **Fig. 3.9**, which is analogous to **Fig. 3.5**. It compares the mean measured spectrum from the 10^4 realization of the masked sky (green dot-dashed curve) with the theoretical expectation obtained by convolving the input power spectrum with the mixing matrix (black curve). The shadowed area represents the `rms` scatter of the $\bar{C}_\ell^{\rm masked}$ over all realizations. The effect of the mask is accounted for by the convolution reasonably well. The convolved theoretical spectrum turns out to be systematically larger than the estimated one. The systematic offset (**Panel c** of the **Figure 3.9**) increases with the angular scale, as expected. It is below 20% and significantly smaller than the random error (shadowed area). The large discrepacies at small reflect the map pixelization.

The scatter among the estimates, i.e. the random errors, is expected to be larger for the masked sky than for the full sky cases. To quantify this difference, for the type of mask we have considered here, we compare the `rms` scatter measured in these two cases in **Fig.3.10**. The solid blue curve, hidden by the continuous magenta curve, and the green dashed curve represent the amplitude of the `rms` scatter in the full-sky and masked-sky cases, respectively. The `rms` scatter for the masked case is larger than in the full-sky cases at high multipoles. At low multipoles the situation is reversed. The reason for this is that the application of the mask induces spurious dipoles and quadrupoles that are common to all realizations, effectively decreasing the cosmic variance on the corresponding angular scales. In the plot we also show the `rms` errors of a Gaussian, all-sky field (solid magenta curve) that, as expected, matches the measured `rms` of the full-sky spectra. We also show the case of a Gaussian field extended over an $f_{sky} = 0.692708$ fraction of the sky rather than on the whole sphere (dashed magenta line). In this case Gaussian errors are larger than in the actual realizations. The reason is that in the Gaussian case all modes are assumed to be independent whereas the mask induces a correlation among them.

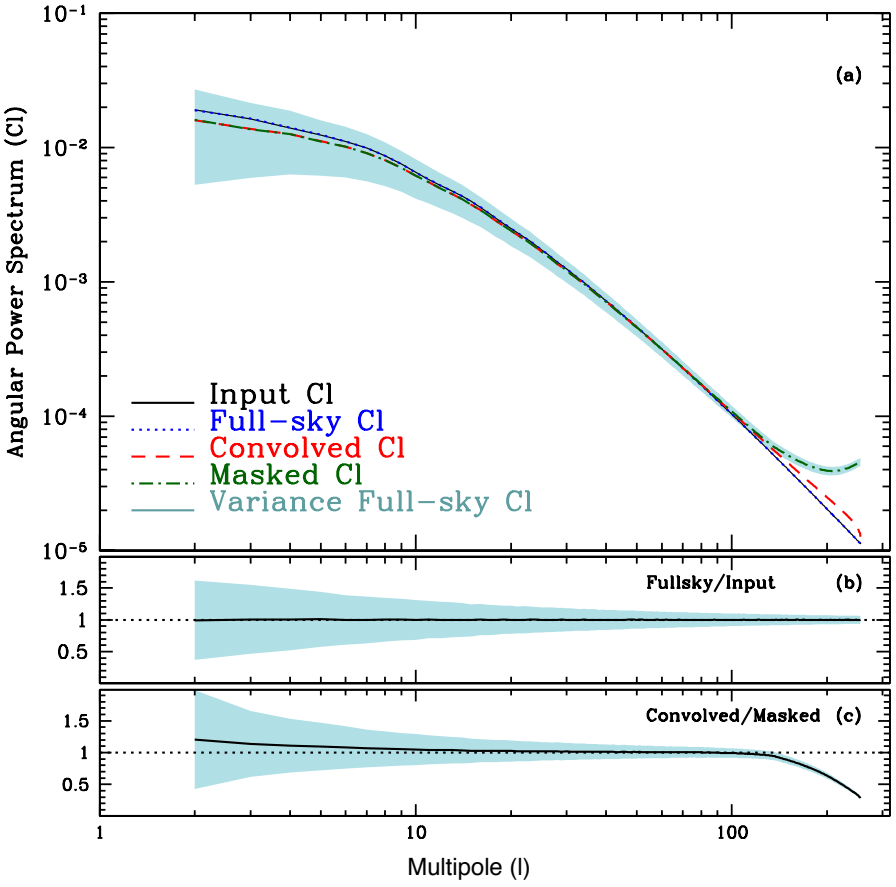

Figure 3.9: Panel (a): The solid black line shows the $C_\ell^{\rm input}$ i.e. the unmasked angular power spectrum used to generate the full-sky map; the dotted blue line,fully superimposed to the black one, shows the mean of the $C_\ell^{\rm fullsky}$, that is the power spectrum measured from the full-sky map also shown in **Fig. 3.5**. The dashed red line is the input C_ℓ convolved with the mixing matrix and should be compared with the dot-dashed green curve that represents the mean angular spectrum estimated from the 10^4 realizations of the masked map. The shaded region is the `rms` scatter among the realizations. **Panel (b)**: Is the same as the middle panel of **Fig. 3.5** and shows the mean of the ratio between the fullsky power spectrum and the input one compared with the corresponding variance (shadowed area). **Panel (c)**: The ratio between the mean measured spectrum in the 10^4 realizations of the masked sky and the convolved input spectrum with the corresponding variance (shadowed region).

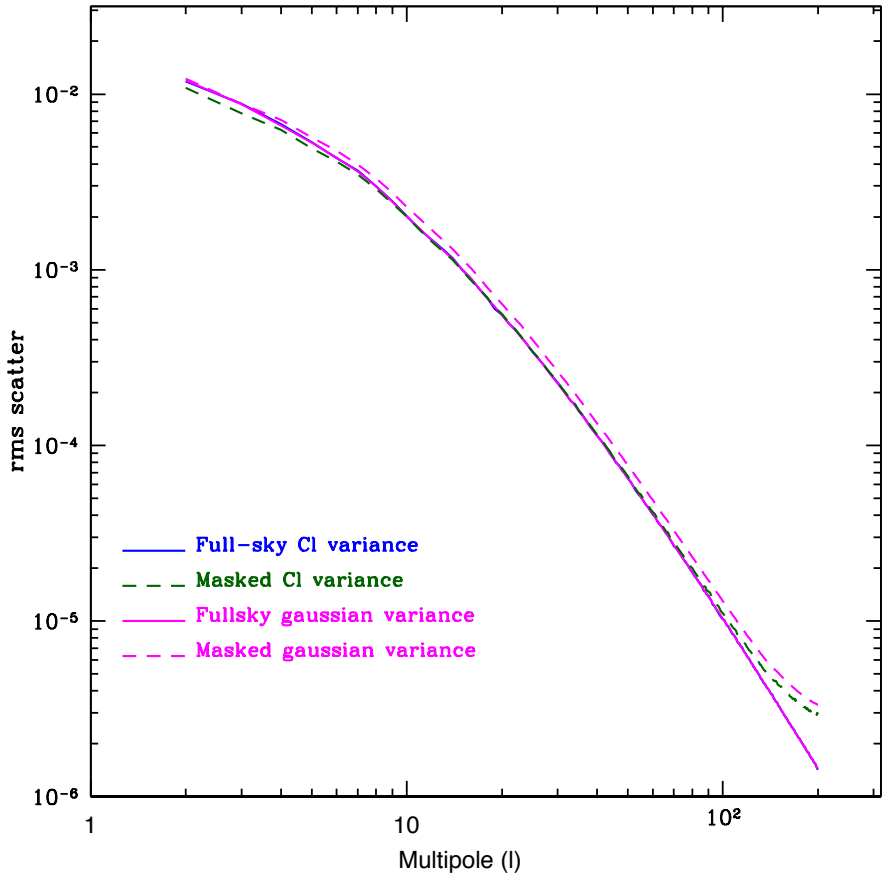

Figure 3.10: The variances of the different angular power spectra. The solid blue line shows the variance of the $C_\ell^{\rm fullsky}$, i. e. the power spectrum measured from the full-sky map obtained from random Gaussian distributed $a_{\ell m}$; the dashed green line shows the variance of the spectrum measured from the masked map, $C_\ell^{\rm masked}$. The corresponding solid and dashed magenta lines shows the Gaussian variances computed for two different cases: a full-sky map (continuous curve) and a field over a fraction of the sky, $f_{sky} = 0.692708$ (dashed curve).

3.4.3 Test 3: Recovering an angular spectrum: code performance in a realistic setup

The goal of this section is to assess the performance of the code under more realistic circumstances. So far we have considered angular power spectrum of a discretized Gaussian random field. To get closer to reality let us now consider a distribution of mass tracers characterized by a redshift distribution matching that of realistic galaxy catalogs. In particular, we shall compute the angular spectrum of a set of dark matter halos extracted from the BASICC N-body simulation (Angulo et al., 2008).

BASICC is the acronymous for Baryon Acoustic Simulation performed at the ICC using the publicly available GADGET-2 code (Springel, 2005). Initial conditions were set up from a Gaussian realization of the CAMB (Lewis et al., 2000, Howlett et al., 2012) power spectrum of density fluctuations. The particle distribution was then evolved using the Zel'dovich approximation (see **Sec. 2.2.2**) up to $z = 60$. The cosmological model adopted is the same as the Millennium Simulation (Springel et al., 2005) ΛCDM cosmology with $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$ and $h = 0.733$. The BASICC simulation is dark matter only and consists of a computational cube of side $L = 1340\ h^{-1}$Mpc loaded with 1448^3 particles of mass $m = 5.4910^{10}\ h^{-1}M_\odot$. We have considered all the halos detected by the Friend to Friend algorithm with more than 10 dark matter particles i.e. more massive than $5.4910^{11}\ h^{-1}M_\odot$. We place the observer at the center of the cube and consider only haloes within a sphere of radius $R = L/2$. Then we Monte Carlo extract a relatively shallow sample of halos with a redshift distribution modeled using the same functional form adopted by Crocce et al. (2011):

$$\frac{dN}{dz} \propto \left(\frac{z}{0.03}\right)^2 \exp\left[-\left(\frac{z}{0.03}\right)^{1.5}\right] \tag{3.67}$$

where 0.03 is the redshift corresponding to the maximum value of the

dN/dz (we have chosen this value since the maximum redshift of the BASICC simulation is $z \sim 0.235$). The redshift distribution is linked to the selection function through **Eq. 2.18**, with again $\mathcal{N}$ the total number of galaxies:

$$\int_0^{z_{\max}} \frac{dN}{dz} dz \equiv \mathcal{N} \tag{3.68}$$

where $z_{\max}$ is the redshift corresponding to $L/2$.

We then projected the halo distribution onto a sphere and obtain a pixelized map of the halos' surface over-density. We use our code to estimate the angular spectrum of this map and compare the result with the theoretical mass angular spectrum produced by CLASS. In this last step we consider two model spectra: i) the one predicted by linear theory ad ii) the nonlinear one predicted by the Halofit model (Smith et al., 2003). The resulting spectrum is shown in **Fig. 3.11**.

Since halos are biased tracers of the mass density field, we don't expect that the measured spectrum matches the one predicted by CLASS. On linear scales the two will differ by a multiplicative factor represented by the effective halo bias $b_{\rm eff}$, that we computed using the definition:

$$b_{\rm eff}(\mathcal{M}) = \frac{\int_{\mathcal{M}}^{\infty} dm\; n(m)\; b(m)}{\int_{\mathcal{M}}^{\infty} dm\; n(m)} \simeq 0.835 \tag{3.69}$$

where $n(\mathcal{M})$ is the halo mass function provided by Jenkins et al. (2001), $\mathcal{M} = 5.49 \times 10^{11} h^{-1} M_{\odot}$ is the minimum mass of the haloes of the simulation and $b(m)$ is the halo linear bias model of Sheth and Tormen (2002). The fact that we have extracted haloes ignoring their substructures corresponds to assuming that, in the halo model, only one galaxy is hosted in each halo, i.e. $\langle N_{\rm gal}|m\rangle/\bar{n}_{\rm gal} = 1$ in **Eq. 2.28**. The measured angular spectrum is shown in the upper panel of **Fig. 3.11** (solid black line) and compared with both linear (solid red) and nonlinear (dashed blue) CLASS predictions.

Our results are in good agreement with theoretical expectations at small ($\ell < 10$) multipoles where nonlinear effects are negligible. At higher mul-

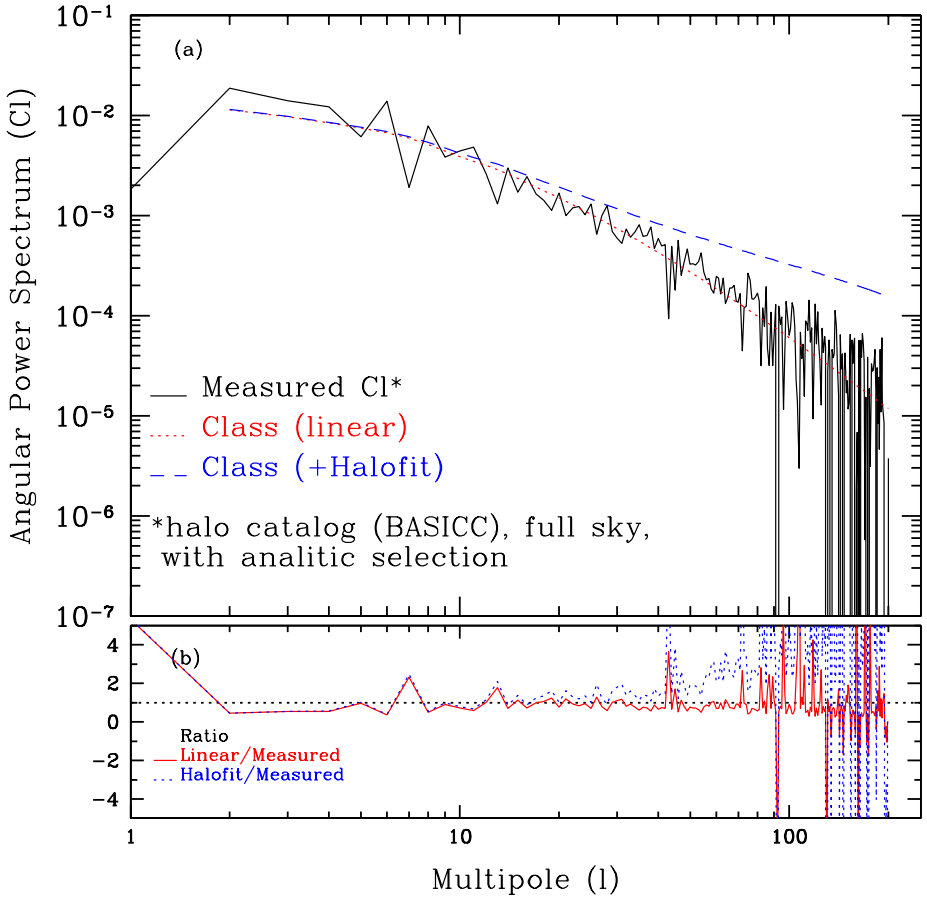

Figure 3.11: Panel (a): The comparison between the angular power spectrum measured from a mock catalogue built from the BASICC simulation and the one obtained from the code CLASS using the same cosmological parameter of the simulation. The dotted-red and dashed blue curves show, respectively, the linear theory and the HALOFIT model angular spectra from CLASS. The amplitude of the spectra are corrected by the bias as explained in the text. **Panel (b):** The ratio between the different spectra: in solid red the ratio between the CLASS linear spectrum and the measured one, in dotted blue the CLASS non-linear spectrum and the measured one.

tipoles linear theory provides a better match to our estimated spectrum. This seemingly contradictory result reflects the fact that, by ignoring the halo substructure, we are minimizing the impact of nonlinear effect or, to use the halo model, we are ignoring the 1-halo term and only consider the two-halo one that is well described in the framework of linear theory.

Chapter 4

The 2MPZ catalogue

In this Chapter we describe the galaxy catalogue that we analyze in this Thesis, namely the 2MASS Photometric Redshift (2MPZ) catalogue. In particular, we will focus on those observational proprieties that are relevant to our correlation study and that affect the selection criteria used to extract the subsamples used in our analysis.

The 2MPZ (Bilicki et al., 2014) has been built by cross-matching the Two Micron All Sky Survey Extended Source Catalog (2MASS XSC, Jarrett et al. (2000)), the Wide-field Infrared Survey Explorer (WISE, Wright et al. (2010)) and the SuperCOSMOS Sky Survey program (SSS, Hambly et al. (2001a) Hambly et al. (2001b), Hambly et al. (2001c)), in order to obtain an homogeneous estimate of the photometric redshifts over the best part of the celestial sphere. Three different spectroscopic samples have been used to calibrate the photometric redshifts in order to guarantee unbiased estimate over the redshift range probed by the 2MASS catalogue.

4.1 Photometry

The basic photometric dataset is the 2MASS Extended Source Catalog (XSC), which contains approximately 1 million objects from the 2MASS (Skrutskie et al., 2006), a ground-based survey performed between 1997 and

2001, covering 99.998% of the celestial sphere and delivering uniform, precise photometry in the three near infrared bands $J(1.25\ \mu\mathrm{m})$, $H(1.65\ \mu\mathrm{m})$, and $K_s(2.16\ \mu\mathrm{m})$. Observations were conducted from two dedicated 1.3 m diameter telescopes located at Mount Hopkins, Arizona and Cerro Tololo, Chile. Among the total objects of the complete 2MASS XSC, (Bilicki et al., 2014) removed a small fraction corresponding to artifacts and erroneous or null measurements, and then masked out the strip of galactic latitude $|b| < 8^\circ$ since prone to large extinction and, consequently, correction uncertainties.

The photometric data from 2MASS were supplemented by the measurements made within the WISE and SuperCOSMOS surveys. The former is a space-borne telescope that in 2010 mapped the entire sky in four bands centered at 3.4, 4.6, 12 and 22 μm. The two bands at shorter wavelengths, W_1 and W_2, have a much higher detection rate than the two other bands, so the expectation was that the large majority of 2MASS XSC galaxies should be also detected in W1 and W4. For this reason, only the photometry in these bands has been used, with a resulting matching rate of the '2MASS-good' sample with WISE W_1- and W_2-detected sources of over 99%. The coverage, however, is not uniform. Characteristic features in the sky map are the Galactic Plane and bulge, the Magellanic clouds, incompleteness along the Moon trail (Kovács et al., 2013), and a long strip at $\ell_{\mathrm{gal}} \simeq 140^\circ$ and $15^\circ < b < 35^\circ$ due to the spacecrafts' magnetic torque rods, whose activation resulted in a small jump in the telescope pointing with the effect of smearing the resulting image of the same point on each orbit[1]. All these areas have been masked out in the catalogue that we have used in our analysis.

The superCOSMOS catalogue is the outcome of a project aimed at digitizing the photographic plates measurements in the optical B, R and I

[1]See `http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/sec6_2.html#lowcoverage` for more details.

bands obtained from an automatic scanning machine. The source photographic material came from the United Kingdom Schmidt Telescope (UKST) in the south hemisphere and the Palomar Observatory Sky Survey-II (POSS-II) in the north one, observations having been taken during the last quarter of the twentieth century. These catalogs were given an accurate photometric calibration using the procedure described in Peacock et al. (2016). Although almost 189 million sources with B and R extended-source photometry has been measured over the whole sky, only those away from the Galactic Plane are applicable for extragalactic studies and have been considered in this analysis.

4.2 Spectroscopic and photometric redshifts

Photometric redshifts computed from the photometric material described in the previous section have been calibrated using a subsample of 2MPZ objects with spectroscopic redshifts. These were taken from 5 different spectroscopic surveys. In order of increasing depth they are: the 2MASS Redshift Survey (2MRS, Huchra et al. (2012)), the ZCAT compilation[2] based on the CfA Redshift Catalogue (Huchra et al., 1996), that constitute the most local sample, the 2dF Galaxy Redshift Survey (2dFGRS, Colless et al. (2001), Colless et al. (2003)), the 6dF Galaxy Survey Data Release 3 (6dFGS DR3, (Jones et al., 2009)) and the Sloan Digital Sky Survey Data Release 9 (SDSS DR9, Ahn et al. (2012)). These surveys have some overlap. After taking care of the redundancy we are left with about 46000 objects, i.e. the 32% of the total '2MASS-good' sample used to build 2MPZ.

This significant fraction of spectroscopic redshifts encouraged (Bilicki et al., 2014) to estimate photometric redshift through the machine-learning approach. Contrary to the SED fitting we have mentioned in **Chapter 1**, which bases its efficacy on the fit of the overall shape of spectra and on

[2] `http://www.cfa.harvard.edu/~dfabricant/huchra/zcat/`

the detection of strong spectral features, the machine-learning method derives an empirical relation between magnitudes (or possibly other galaxy parameters, such as colors) and redshifts using a subsample of objects with measured spectroscopic redshifts, known as the *training set*. The training algorithm is then applied to the rest of the sample for which no spectroscopic redshifts are known. In particular, the authors applied a particular implementation of the Artificial Neural Networks (ANN, Firth et al. (2003), Tagliaferri et al. (2003), Collister and Lahav (2004), Vanzella et al. (2004), Way et al. (2009), Cavuoti et al. (2012)) called ANNz (Collister and Lahav, 2004), a freely available, ready-to-use software[3].

The derived photometric redshifts have errors nearly independent of distance, with an all-sky precision of $\sigma_{\rm phot} = 0.015$ and a very small percentage of outliers ($\sim 3\%$). The error distribution $\delta z \equiv z_{\rm phot} - z_{\rm spec}$ for the all-sky sample is showed in **Fig. 4.1**. Since its shape is non-Gaussian in the wings, the authors preferred to use a more flexible fitting form (blue line in **Fig. 4.1**):

$$N(\delta z) \propto (1 + \delta z^2/2as^2)^{-a} \tag{4.1}$$

with $a \simeq 3.2$. For the whole sample, best-fit $s = 0.0125$ which is consistent with the average scatter of ~ 0.015. The Gaussian gives instead a worse approximation (dashed black line) since with the best fit $\sigma = 0.0138$ the central part is reasonably well matched but the wings are underestimated.

In this Thesis we shall ignore the non-Gaussian wings and assume that errors are Gaussian distributed with a variance of $\sigma_{\rm spec} = 0.015$ that represents our rms error.

4.3 The 2MPZ catalogue used in our analysis

The final 2MPZ catalogue obtained after masking out the problematic regions, including the zone of avoidance around the Galactic plane ($b < 10$),

[3] `http://www.homepages.ucl.ac.uk/~ucapola/annz.html`

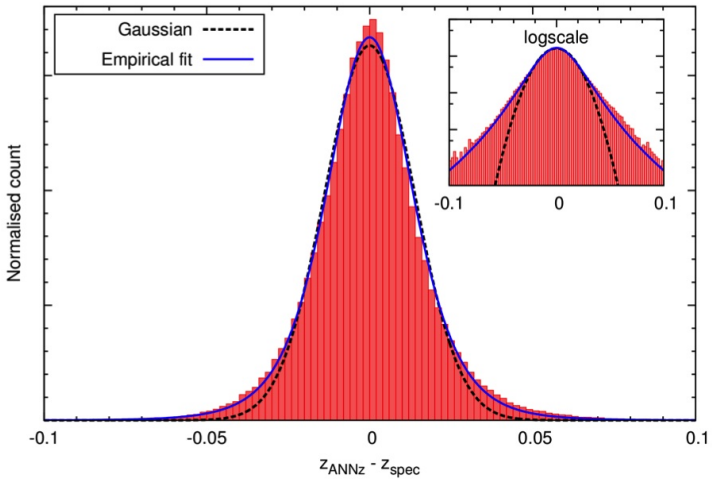

Figure 4.1: Error distribution between the photometric and spectroscopic redshifts for the 2MASS Photometric Redshift sample (red bars). The blue solid line is the empirical fit with $a = 3.2$ and $s = 0.0125$, the black dashed line shows a best-fit Gaussian with $\sigma = 0.0138$. Figure from Bilicki et al. (2014).

contains about 935000 galaxies distributed over 80% of the sky. The catalogue is 95% complete. The redshift distribution of its galaxies is shown in **Fig. 4.2** (green histogram). The distribution peaks at $z \sim 0.065$ with a prominent tail at redshifts as large as 0.3, which pushes the mean redshift at $\bar{z} = 0.08$. The figure also shows the distributions of 2MPZ galaxies with known spectroscopic redshift (red) and of those with photo-z redshifts only (blue).

In **Fig. 4.3** represents the Aitoff projection of the sky distribution of 2MPZ galaxies in Galactic coordinates. The Galactic plane is clearly evident. The color code refers to photometric redshifts. The blue color is used for the nearest objects and the red for the most distant ones. The spidery-like structure of the cosmic web is clearly visible despite the projection effect.

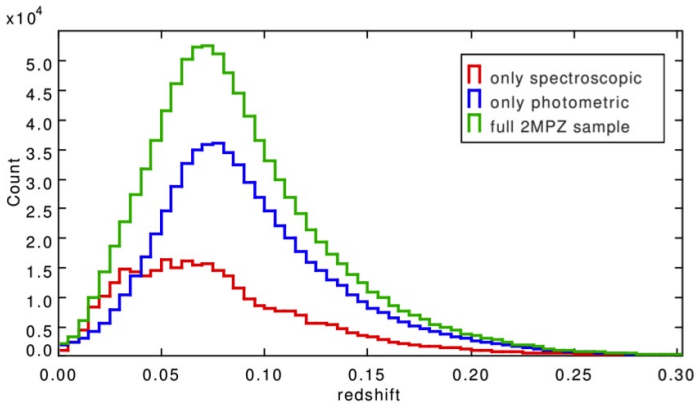

Figure 4.2: Comparison of redshift distributions in the all-sky 2MPZ catalogue: sources with spectroscopic redshifts in red, those with only photo-z's in blue and the full redshift sample in green. Figure from Bilicki et al. (2014).

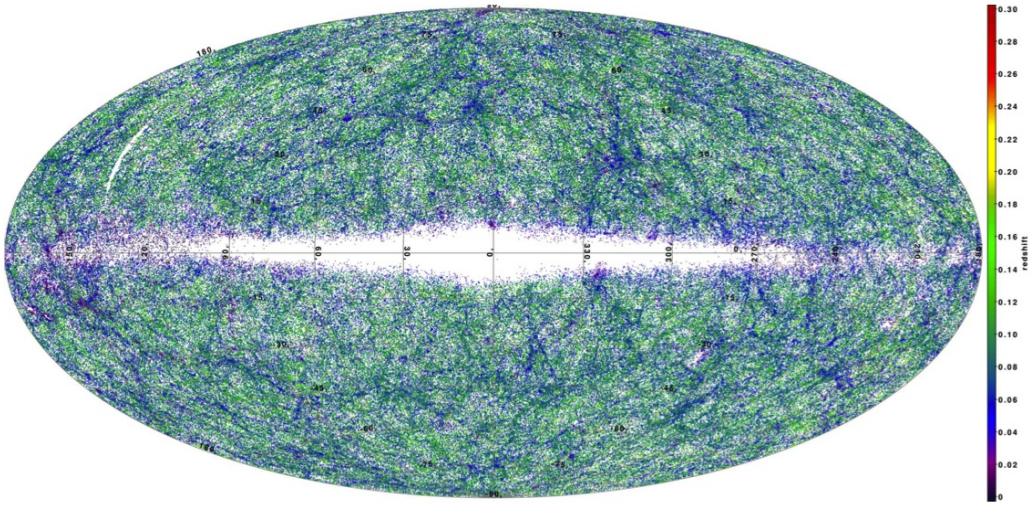

Figure 4.3: Aitoff projection in Galactic coordinates of the 2MASS Photometric Redshift catalog, color-coded by photometric redshift. The cosmic web is evident. Figure from Bilicki et al. (2014).

Chapter 5

Lognormal 2MPZ mock catalogues

In **Chapter 3** we have seen how in the case of a linearly evolved Gaussian field and all-sky coverage the errors can be estimated analytically. On the contrary, when the evolution enters the non-linear regime and/or geometry masks are present, the spherical harmonics coefficients are not independent anymore and errors must be evaluated with numerical methods.

One approach is to rearrange the objects in the dataset itself, as in the *Jackknife* (Tukey, 1958) or the *Bootstrap* (Efron, 1979) methods. However, these are approximated methods. To fully account for errors and their covariance one needs to access several independent realization of the same statistical sample. In our case this can be done by generating a set of independent mock catalogs mimicking the same characteristics as the real one. In this context, the best mock catalogues are the realistic ones that one can build from large N-body simulations complemented with adequate treatment of baryon physics (hydrodynamical simulations) or semi-analytic recipes to associate galaxies to the mass distribution. The second route is currently the most popular one and consists of assigning galaxies to dark matter halos extracted from a collisionless N-body experiment using some phenomenological recipe. However, the creation of a large

number of such mock catalogues is computationally expensive, since one would need to run many different N-body simulations. However, in many cases, one can rely on approximate methods to generate a large number of mocks. The goodness of this approach depends on the characteristics of the real sample. However, most often than not, such approximate methods are good enough to estimate uncertainties for specific model parameters and within a well defined range of scales. In this thesis we shall assume that the 2-dimensional density field traced by galaxies is **lognormal**. With this assumption, we are able to set up a fast procedure to generate a large number of 2MPZ mock catalogues, as detailed below.

In **Chapter 1** we have seen how different inflationary models predict that primordial density perturbations obey to Gaussian statistics. Moreover, if the distribution is Gaussian, the 2-point statistic is sufficient to fully describe the statistical proprieties of the density field, which justifies the interest of Gaussian fields in cosmology. Unfortunately, the Gaussian hypothesis is only valid at early epochs or at very large scales and breaks down on small, nonlinear scales. If one wishes to describe the statistical properties of the galaxy distribution in the local universe down to a few Mpc scales, then the Gaussian hypothesis cannot be valid.

An effective way to describe the statistical properties of non-linearly evolved cosmological density fields without resorting to N-body techniques is that of adopting the **lognormal model** (Coles and Jones, 1991), that provide a reasonably good description of the non-linearly evolved density field obtained from N-body experiments. After introducing this model and its proprieties, we will use it to generate a discrete distribution of object and compute its angular power spectrum. This spectrum is then compared with the one estimated in the previous Chapter and generated from a Gaussian field to check for possible discrepancies. Then, we will focus on building a mock 2MPZ catalogue of galaxies with a pre-defined angular spectrum, redshift distribution dN/dz and lognormal statistics. Each mock galaxy in the final catalogue will be characterized by the angular

position, spectroscopic redshift and photometric redshift.

5.1 The lognormal model

The 1-point probability distribution function (PDF) characterizes the statistical properties of the mass density field ρ. As long as the density fluctuations are in the linear regime, their PDF remains Gaussian, but once they reach the non-linear stage their PDF deviates from the initial Gaussian shape. In the next section we will show that a simple, lognormal transformation of the Gaussian field is a good approximation to that of a non-linearly evolved Gaussian overdensity field. The lognormal model is defined by its 1-point PDF (Lahav and Suto, 2004):

$$p_{\ln}(\delta)\, d\delta = \frac{1}{\tilde{\rho}} \frac{1}{\sqrt{2\pi s^2}} \exp\left\{ -\frac{(\ln \tilde{\rho} - m)}{2s^2} \right\} d\delta \tag{5.1}$$

where:

$$\tilde{\rho} = 1 + \delta \tag{5.2}$$

$$m = -0.5 \ln(1 + \sigma^2) \tag{5.3}$$

$$s^2 = \ln(1 + \sigma^2) \tag{5.4}$$

$$\sigma^2 = \langle \delta^2 \rangle \tag{5.5}$$

The PDF of the quantity $\ln(1+\delta)$, that we can call $P_{\ln}[\ln\ (1+\delta)]$, is a Gaussian function with mean equal to m:

$$P_{\ln}[\ln(1+\delta)] = \frac{1}{\sqrt{2\pi s^2}} \exp\left\{ -\frac{[\ln\ (1+\delta) - m]^2}{2s^2} \right\} \tag{5.6}$$

The cumulative lognormal distribution is obtained by integration:

$$C_{\ln}(\delta) = \mathrm{erf}\left(\frac{\ln \tilde{\rho} - m}{s} \right) \tag{5.7}$$

where $\mathrm{erf}(x)$ is the *error function* defined as:

$$\mathrm{erf}(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt \tag{5.8}$$

Several motivations persuade to think that the **Eq. 5.1** could be a good model for the distribution of a non-linearly evolved mass density field. Hubble (1934) first noticed that the distribution of galaxy counts in two-dimensional cells on the sky could be well approximated by the **Eq. 5.1**, and the same possibility was also discussed by Peebles (1980). Various arguments have been provided to justify the lognormal PDF for an evolved Gaussian Field (Coles and Jones, 1991). However, the best proof is provided by N-body simulations: they show that the PDF of the mass density field is indeed well described by a lognormal model (Kayo et al., 2001). Chiang et al. (2013) used this model to represent the matter/galaxy densities, while Taruya et al. (2002) and Hilbert et al. (2011) investigated the extent to which the lognormal model describes the statistics of weak-lensing fields. Because of the many possible applications in astrophysical and cosmological context, lognormal models have been widely used to produce mock catalogues of evolved mass and galaxy density fields. Numerical codes to generate lognormal random fields are already available. One example is the C++ code of Xavier et al. (2016) that generates correlated lognormal field defined on the sphere[1]. However, we prefer to develop our own code to guarantee a better control over the procedure and to customize the mock catalogue generation to match the characteristics of the 2MPZ catalogue. We stress that what the code generates is the mass density field. The galaxy density field will also be lognormal only in the limits of linear galaxy bias, which is what we assume here.

The lognormal model is one of the simplest ways of defining a fully self-consistent random field which always has $\rho > 0$ and, most importantly, is one of the few non-Gaussian random fields for which interesting proprieties can be computed analytically. It has no more free parameters than the Gaussian field from which it is derived and in many cases analytic results are obtained more easily for this model than the Gaussian

[1]See `http://www.astro.iag.usp.br/~flask/`

(see Coles and Jones (1991) and references therein). These reasons and the further consideration that the accuracy of the lognormal model increases when a smoothing is applied (and the projection on the 2D sphere can be regarded as a form of smoothing) led us to adopt the lognormal model in generating mock catalogues that could well reproduce the 2MPZ catalogue (see **Chapter 4**). Finally, we notice that one can transform a density field characterized by a given 1-point PDF to another one obeying a different PDF without changing its 2-point correlation properties. This suggests that one can take a realization of a Gaussian density field with a well defined power spectrum and obtain a lognormal field with the same spectrum. This is the main idea behind the making of the lognormal mock catalogues that we describe in the next section.

5.2 Lognormal maps and their angular power spectrum

We now want to generate a lognormal field on a sphere starting from the corresponding Gaussian field, and thus verify that its angular power spectrum is consistent with the expected one. We will proceed as follows.

1. Generate a Gaussian field with a known input angular power spectrum (following the same procedure described in **Section 3.4.1**);

2. Measure the full-sky angular power spectrum of that map as a consistency check;

3. Apply the lognormal transformation:

$$\delta_{\ln} = \exp(\delta + m) - 1 \tag{5.9}$$

   with m the `rms` of the Gaussian field δ. $\delta_{\ln}$ is the lognormal field.

4. Compute the angular power spectrum of the lognormal field and compare it with the Gaussian one.

To generate the input power spectrum (step 1) we use `CLASS` with a ΛCDM model with $\Omega_b = 0.04$, $\Omega_{CDM} = 0.21$, $\Omega_\Lambda = 0.75$, $T_{cmb} = 2.7255$, $h = 0.73$, $A_s = 2.21 \cdot 10^{-10}$. Similarly to what we have done in **Section 3.4.1**, we generate a Gaussian realization of an overdensity field with angular spectrum in the form of an `HEALPix` map with $n_{side} = 64$. We then verify that its mean is zero and compute its `rms` (**Eq. 5.5**). For each pixel of the map, we use **Eq. 5.9** in order to obtain a lognormal map. The entire procedure is then repeated to generate 10 different realizations. The resulting PDFs are shown in **Fig. 5.1**. In the same figure, for comparison, we also show the expected lognormal PDF. This sanity check confirms that the generated field is effectively lognormal.

In a second test we compare the angular spectra of the Gaussian and the lognormal fields to verify that they are equal, as expected. In this test we consider three spectra. The theoretical one, generated by `CLASS`, that we used as input. The full-sky one, that we have measured as the mean of the 1000 Gaussian realizations. And the lognormal one, measured from the mean of the 1000 lognormal transformations. To check whether these spectra are in agreement, we compute their differences and compared them with the `rms` scatter among realizations. The results are shown in **Fig. 5.2**. In the top panel the input angular spectrum (black curve) is plotted with the mean (Gaussian) fullsky spectrum (blue dotted) and the mean lognormal spectrum (green dot-dashed). The red dashed line is plotted for reference and represents the result of the same sanity check shown in **Fig. 3.5**. The power spectrum of the lognormal field closely matches that of the Gaussian field, as expected. Significant discrepancies are detected below the pixel scale and reflects the map resolution. The three lower panels provide a more quantitative estimate of this agreement. In each of them the differences are very small (about 0.005% between the Gaussian and the lognormal field) and well within the `rms` of the mean (the shaded area).

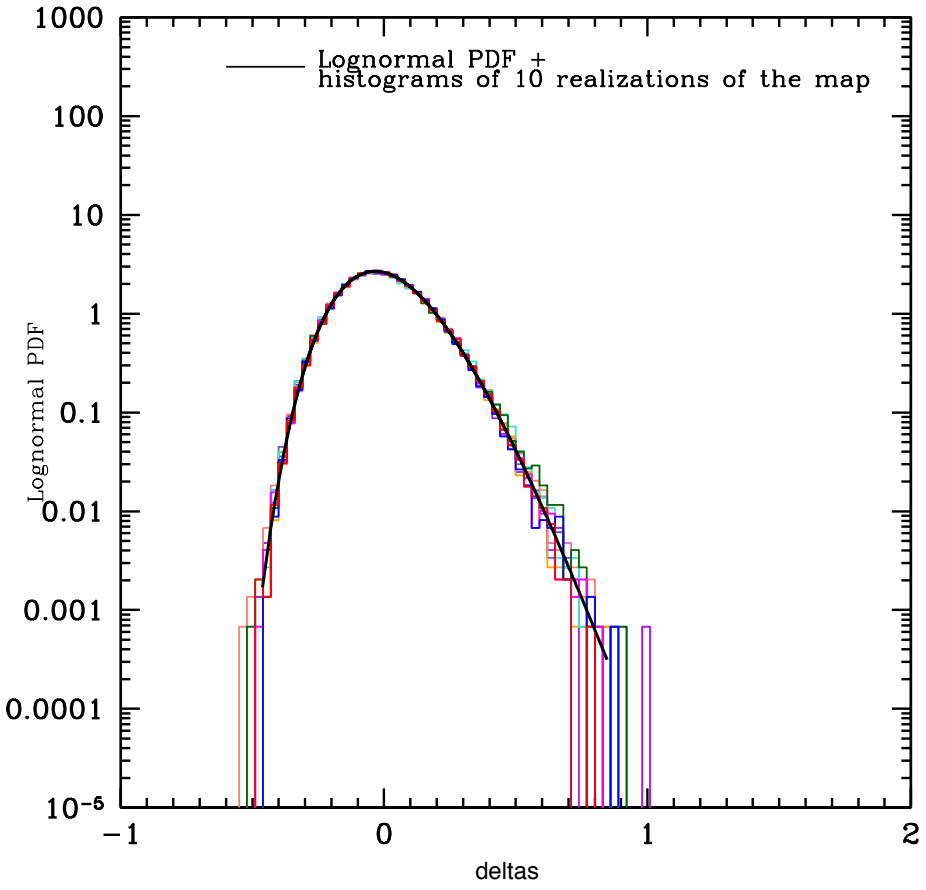

Figure 5.1: Individual PDFs of the lognormal density field in 10 different realizations (histograms) vs. the expected probability distribution function expressed by **Eq. 5.1** (solid black curve). On the Y axis we show the frequency of each $\delta_{\ln}$ value.

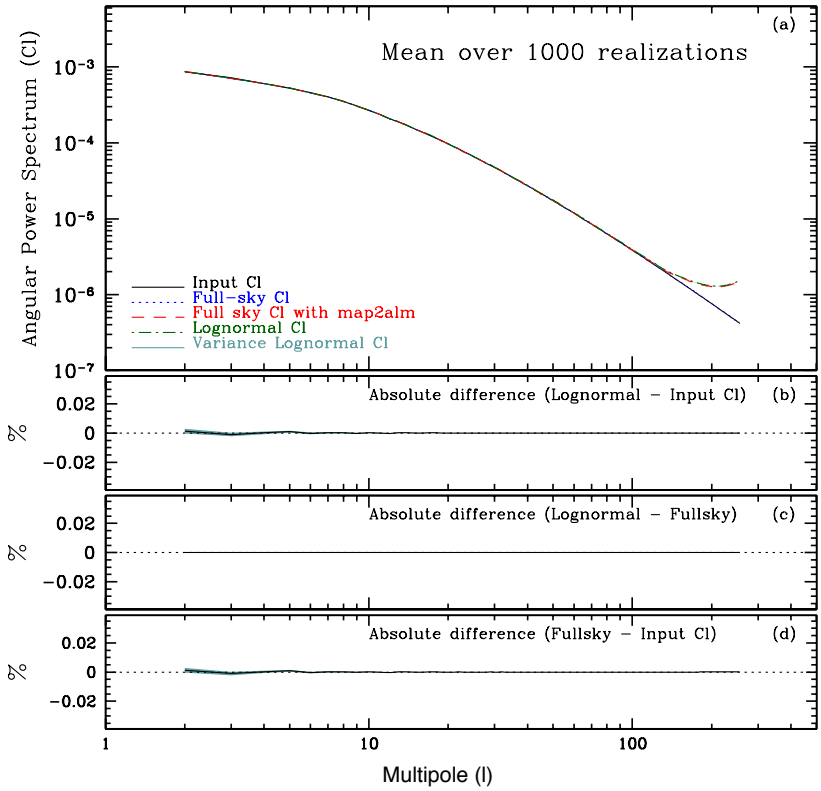

Figure 5.2: Top panel. Input angular spectrum from `CLASS` (black curve), mean spectrum among the 1000 Gaussian realizations (blue dotted), mean among the corresponding lognormal transformations (green dot-dashed). The mean is over 1000 realizations. Red dot-dashed curve shows the same sanity check as in **Fig. 3.5**. Lower panels: % difference between the angular spectra. From top to bottom: i) lognormal vs. input, ii) lognormal vs. Gaussian. iii) Gaussian vs. input. The shaded area represents the `rms` scatter of the mean.

5.3 From the lognormal maps to mock galaxy catalogues

Our main goal is to generate a mock catalogue mimicking a photometric-redshift galaxy catalogue (namely the 2MPZ) to measure their angular power spectra in different redshift shells. So we aim at generating different lognormal mock catalogues of objects in different spherical shells with the same geometrical mask of the 2MPZ catalogue and the same photometric distribution.

5.3.1 Angular part of the catalogue

Our goal is now to generate a discrete distribution of objects in a spherical shell that traces an underlying 2D lognormal mass density field. The inputs here are a lognormal density field in the form of a `HEALPix` map with a given angular spectrum that we have generated in the previous section and the total number of objects in the shell, $\mathcal{N}$. The mean object density in the map is set equal to the mean mass density:

$$\bar{\rho} = \frac{\mathcal{N}}{N_{\text{pixels}}} \tag{5.10}$$

Since the lognormal maps specify, for each pixel, the overdensity of the tracers $\delta_{\ln}[i]$, where i indicates the ith pixel, then the expected mean number of objects in the pixel is:

$$\langle N[i] \rangle = (1 + \delta_{\ln}[i]) \, \frac{\mathcal{N}}{N_{\text{pixels}}} \tag{5.11}$$

Assuming that tracers are a Poisson random process, we set the number of objects in the pixel $N[i]$ equal to a Poisson deviate with mean $\langle N[i] \rangle$. With this procedure we determine the number of objects in each pixel for each individual lognormal realization. Each object is placed at a random angular position within its pixel. **Fig. 5.3** shows the angular map of the normalized objects' number density in one of the lognormal realizations

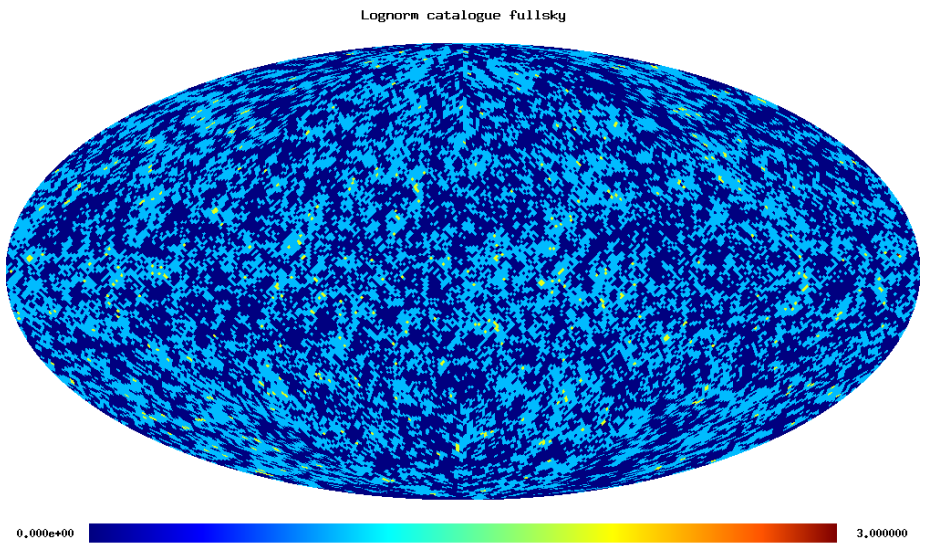

Figure 5.3: 2D map of the normalized object's number density in one of the 2MPZ lognormal mock. The color code indicates the amplitude of the normalized density.

of the 2MPZ catalogue. In **Fig. 5.4** (top panel) we compare the angular spectrum of the lognormal catalogue of discrete objects (blue solid) with that of the parent lognormal density map (red solid) already shown in **Fig. 5.3**. The bottom panel shows the % difference. The input angular spectrum from `CLASS` is also shown for reference (black, solid). We did not detect any systematic discrepancy. Random differences increase with ℓ, reflecting the shot-noise in the discrete objects catalogue.

To mimic the effect of the Galactic absorption, Zodiacal light and all other effects that induce a non-uniform completeness across the sky, we introduce a binary angular mask that removes all objects within the masked areas. In **Fig. 5.5** we superimpose the mask of the 2MPZ catalogue (see **Chapter 4**) to the lognormal catalogue shown in see **Fig. 5.3**. The effect of the mask is to correlate angular multipoles and distribute power among them. **Fig. 5.6** (top panel) is the analog of **Fig. 5.4** and shows the angular power spectrum of the lognormal masked catalogue (blue solid) with that of the parent lognormal density map (red solid). The bottom panel shows

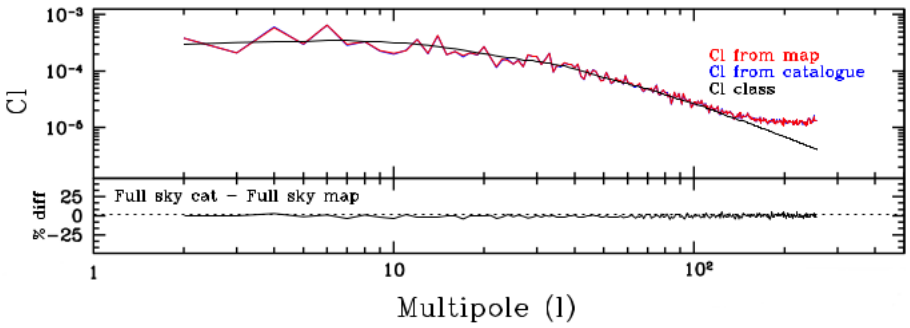

Figure 5.4: Comparison between the spectrum computed from the lognormal catalogue and the one computed directly from the Poisson sampled map (see text for details). The black solid line represents the input angular power spectrum computed with `CLASS`; the red solid line is the spectrum computed from the overdensity lognormal map while the blue solid line is the spectrum computed from the catalogue. The bottom panel shows the percentage difference between the two spectra.

the % difference. The effect of the 2MPZ angular mask is below 15% and mainly seen at small multipoles, where the effect of the mask is larger and removes power on the corresponding angular scales. In the next section we shall investigate how well this effect is modeled by convolving a theoretical power spectrum with the mixing matrix (**Eq. 3.34**).

5.3.2 Multiple redshift shells

The angular catalogue obtained from a reference angular power spectrum using the lognormal model contains the angular positions of a given number of objects with a known angular spectrum but does not specify their radial position. Since we are interested in producing a mock catalogue of objects in a redshift shell, the procedure described so far would be sufficient to mimic an ideal catalogue of objects with well measured distances within a given distance range. However, as we have seen, redshift as distance indicators are neither accurate, because of redshift distortions, nor precise, because of error measurements. Let us ignore redshift distortions that can

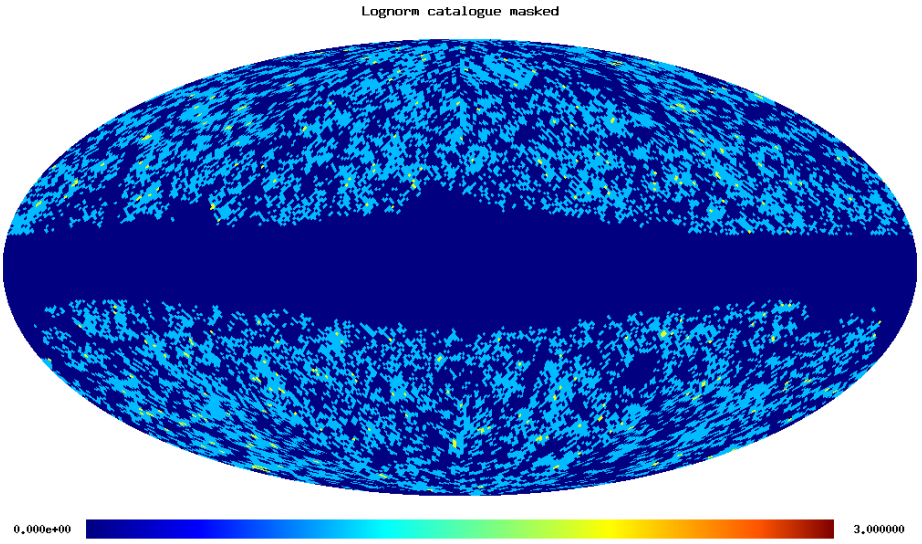

Figure 5.5: 2D map of the normalized object's number density in one of the 2MPZ lognormal masked mocks. The color code indicates the amplitude of the normalized density.

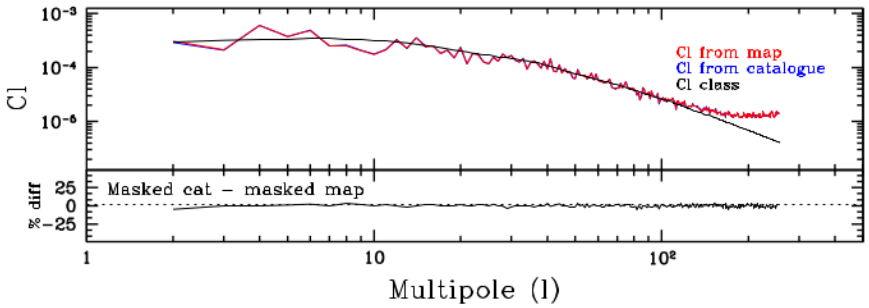

Figure 5.6: Comparison between the spectrum computed from the lognormal catalogue (blue curve) and the one computed directly from the map in the case of the presence of an angular mask (red curve). The black solid line represents the input angular power spectrum computed with `CLASS`. The bottom panel shows the percent difference between the blue and the red curves.

be accounted for in the modeling phase and let us focus on random errors. They depend on the method used to estimate redshifts. Photometric redshifts are much less precise than the spectroscopic ones.The strategy is to assign both spectroscopic and photometric redshifts to mocks galaxies ignoring uncertainties on the former and only account for random errors in the latter. The end product will be a simulated photo-z catalogue of objects for which we know their spectroscopic redshift (or radial position) too.

Let us focus on how to generate a mock photo-z catalogue. If $P(z_{\rm spec})$ is the probability of a spectroscopic redshift $z_{\rm spec}$, then the corresponding probability for $z_{\rm phot}$ is:

$$P(z_{\rm phot}) = \int P(z_{\rm phot}|z_{\rm spec})P(z_{\rm spec})dz_{\rm spec} \tag{5.12}$$

where $P(z_{\rm phot}|z_{\rm spec})$ is the conditional of $z_{\rm phot}$ given $z_{\rm spec}$.

Our goal is to generate a mock photo-z catalogue mimicking the observed $P(z_{\rm phot})$ (which we assume to be proportional to a target galaxy redshift distribution $dN/dz_{\rm phot}$) in which, for each object, we know the true redshift $z_{\rm spec}$ (and therefore the corresponding $P(z_{\rm spec})$). From real datasets we can obtain both $P(z_{\rm phot})$ and $P(z_{\rm spec})$. The latter can be either inferred from the spectroscopic subsample used to calibrate the photometric redshift or from **Eq. 5.12** using techniques like the Richardson-Lucy method (e.g. Szapudi and Pan (2004)) once the conditional probability function, which can also be obtained from the training sample, is specified. Here we use a different approach described below.

1. We generate a volume-limited sample of object in $z_{\rm spec}$ space. A volume-limited catalogue is a sample of objects with constant number density within a given (spectroscopic) redshift interval. The angular positions of the objects are randomly selected on the 4π sphere.

2. We perturb the radial position of each object by random sampling the observed $P(z_{\rm phot}|z_{\rm spec})$. In our mocks we shall neglect the presence

of outlyiers in the observed $z_{\rm phot}$ vs. $z_{\rm spec}$ relation. They correspond to catastrophic errors in the estimate of $z_{\rm phot}$ and constitute a small fraction of 2MPZ galaxies (Bilicki et al., 2014). Once outlyiers are removed the $P(z_{\rm phot}|z_{\rm spec})$ is well approximated by a Gaussian with width independent from $z_{\rm spec}$. We have adopted this approximation.

3. The resulting sample is characterized by a photo-z distribution, $P'(z_{\rm phot})$, that in general does not correspond to the observed one. We enforce the matching by Monte Carlo re-sampling the objects distribution to the target $P(z_{\rm phot})$. This can only be done if $P'(z_{\rm phot}) > P(z_{\rm phot})$ everywhere in the $z_{\rm phot}$ range of interest. To satisfy this condition we require a high density of objects in the parent volume-limited catalogue.

The final result is an all-sky catalogue of objects with random angular position, $z_{\rm phot}$, $z_{\rm spec}$. Their photometric redshift distribution $dN/dz_{\rm phot}$ is built to match the observed one.

To illustrate the procedure to generate $z_{\rm phot}$ and the impact of $z_{\rm phot}$ errors we show, in **Fig. 5.7**, the ideal case of a $z_{\rm spec}$ catalogue with uniform $P(z_{\rm spec})$. The corresponding redshift distribution is represented by the black histogram. Let us divide the catalogue in different subsamples in four different $z_{\rm spec}$ ranges. $z_{\rm phot}$ errors displace (in $z_{\rm phot}$ space) objects from their original z-bin into the neighboring ones. The colored lines quantify the effect. They show the corresponding $z_{\rm phot}$ histograms whose tails 'leak' into the nearby z-bins. As a consequence a sample selected in a $z_{\rm phot}$ bin, will include objects whose true $z_{\rm spec}$ do not belong to the $z_{\rm phot}$ bin and, therefore, characterized by different clustering properties.

5.3.3 2MPZ mocks

We applied the procedure described in **Sections 5.3.1** and **5.3.2** to obtain the 2MPZ mocks. Our goal is to measure 2MPZ galaxy clustering (see **Sec. 4.2** for more details) at different redshifts. It can be qualitatively appreci-

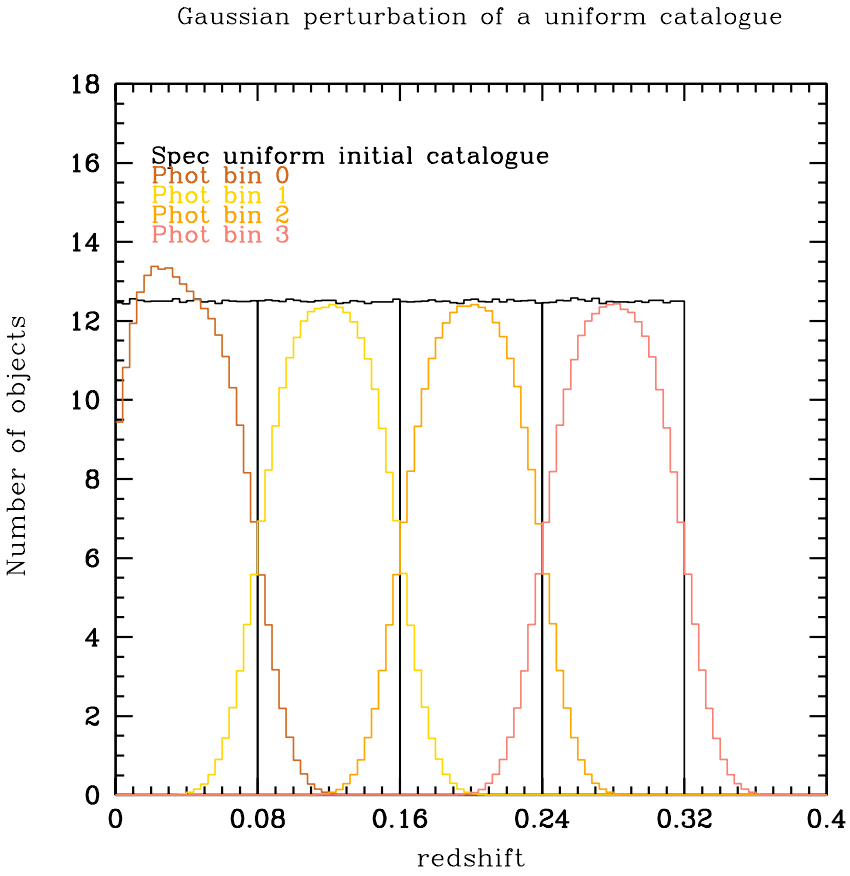

Figure 5.7: An example of the effect of a Gaussian $z_{\rm phot}$ error with $\sigma_{\rm phot} = 0.015$ on a uniform spectroscopic catalogue. A spectroscopic catalogue with $P(z_{\rm spec}) =$ constant, divided in four $z_{\rm spec}$ bins (black solid histogram). Histograms with different shades of yellow indicates the corresponding $z_{\rm phot}$ distributions once $z_{\rm phot}$ errors have been included.

ated from **Fig. 5.7** that, in order to minimize the impact of photo-z errors we need to consider z-intervals significantly larger than the redshift error. On the other hand, redshift bins should be small enough to minimize the impact of evolutionary effects. As a compromise we decided to consider these three redshift bins:

$$\text{first redshift bin } [0 : 0.08] \tag{5.13}$$

$$\text{second redshift bin } [0.08 : 0.16] \tag{5.14}$$

$$\text{third redshift bin } [0.16 : 0.24] \tag{5.15}$$

This choice is justified by the fact that about the half of the total objects belonging to the 2MPZ catalogue are below $z_{\rm phot} = 0.08$. The remaining two redshift bins contain $\sim 1/3$ and $\sim 1/6$ of the total objects. We further assume that $P(z_{\rm phot}|z_{\rm spec})$ is a Gaussian function with variance equal to:

$$\sigma_{\rm phot} = 0.015 \tag{5.16}$$

in agreement with Bilicki et al. (2014) (see **Sec. 4.2**). As a consequence the width of each z-bin is larger than $2\sigma_{\rm phot}$, which limits the amplitude of the leakage effect.

The end product of this procedure is a mock catalogue of objects with a $dN/dz_{\rm phot}$ distribution shown in **Fig. 5.8** (blue, solid histogram) that matches, by construction, the target dN/dz of the 2MPZ catalogue (red solid), i.e. from having imposed that the number of mock objects matches that of 2MPZ galaxies in each z-bin. When we generate the mocks this constraint will hold on average, that is to say we include the Poisson noise to determine the number of object per bin in each 2MPZ mock realization. The green solid histogram shows the corresponding $dN/dz_{\rm spec}$. As expected, the two do not match, causing the angular correlation properties of the objects in a $z_{\rm phot}$ shell to differ from those of the objects in the corresponding $z_{\rm spec}$ interval.

Photometric redshift errors affect the estimate of the angular spectrum. To correct for their influence, we need to model the true redshift distribu-

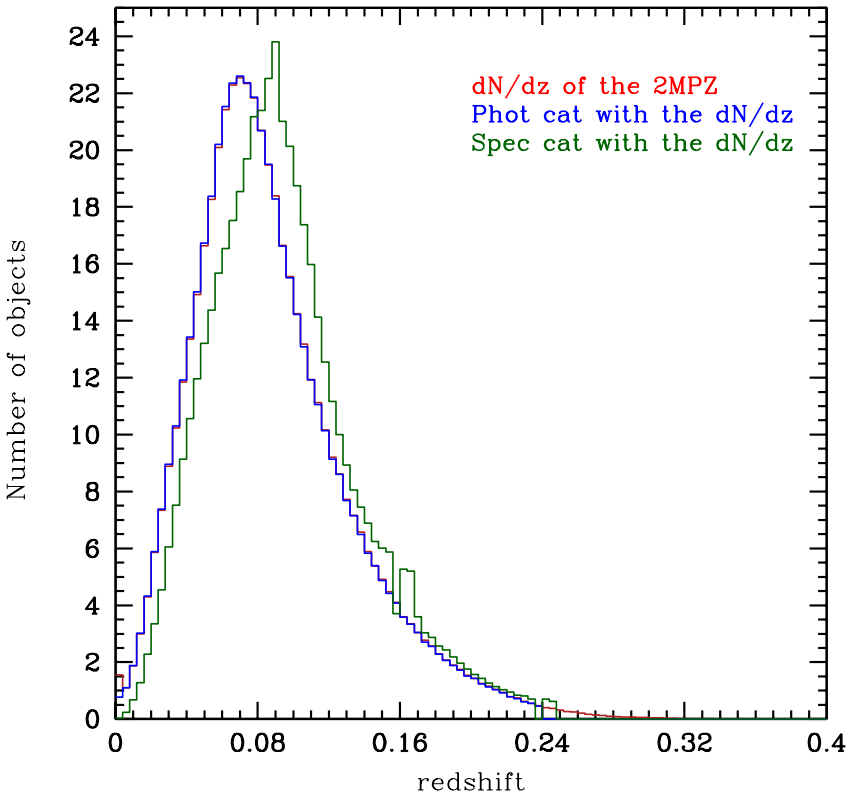

Figure 5.8: The final result. The blue solid line represents the final photometric catalogue, which is overlapped to the target dN/dz (red solid line). The green solid line represents the spectroscopic final catalogue. All objects have been drawn from a parent, volume-limited catalogue in spectroscopic-z space (see text for details).

tion $dN/dz_{\rm spec}$ of galaxies selected according to their z-phot. Instead of using the mocks, we rely on the procedure proposed by Sheth and Rossi (2010). The idea is to estimate the true redshift distribution, $dN/dz_{\rm spec}$, from the observed $dN/dz_{\rm phot}$, i.e. to evaluate:

$$P(z_{\rm spec}) = \int P(z_{\rm spec}|z_{\rm phot})P(z_{\rm phot})dz_{\rm phot} \tag{5.17}$$

This requires assuming a conditional probability $P(z_{\rm spec}|z_{\rm phot})$ that can be obtained from the data. Assuming that $P(z_{\rm spec}|z_{\rm phot})$ is Gaussian, which is approximately true, we can easily estimate the quantity on the lhs *convolving* the two measurables on the rhs.

Fig. 5.9 shows the $dN/dz_{\rm spec}$ distributions of 2MPZ galaxies selected in the three photo-z bins considered in our analysis. The figure clearly illustrate the spread introduced by z-phot errors.

To quantify their impact on our analysis, we show in **Fig. 5.10** the angular power spectrum of objects at their photo-z positions (red curve) vs. that of objects at their true position (black curve). Photometric errors systematically decrease the amplitude of the measured spectrum on all angular scales. This is not surprising, since these errors induce decoherence in galaxy clustering along the radial direction. We shall model this effect when comparing model with data in **Chapter 7**.

We use the mocks to quantify the impact of the geometry mask and to check how well the convolution of the model by the mixing matrix matches observations. To do this we proceed as in **Sec. 3.4.2**, namely convolving the fullsky spectrum computed from the fullsky lognormal map, (showed in **Fig. 5.4** with a red line) with the mixing matrix **Eq. 3.34**. The result is shown in **Fig. 5.11** for the mean over 1000 realizations of the lognormal map - fullsky and masked. The z-bin considered here is $0.16 < z_{\rm spec} < 0.24$. Here the solid blue line is the mean over 1000 realizations of the spectra computed from fullsky lognormal maps, the solid green line is the mean of the spectra computed from the masked lognormal

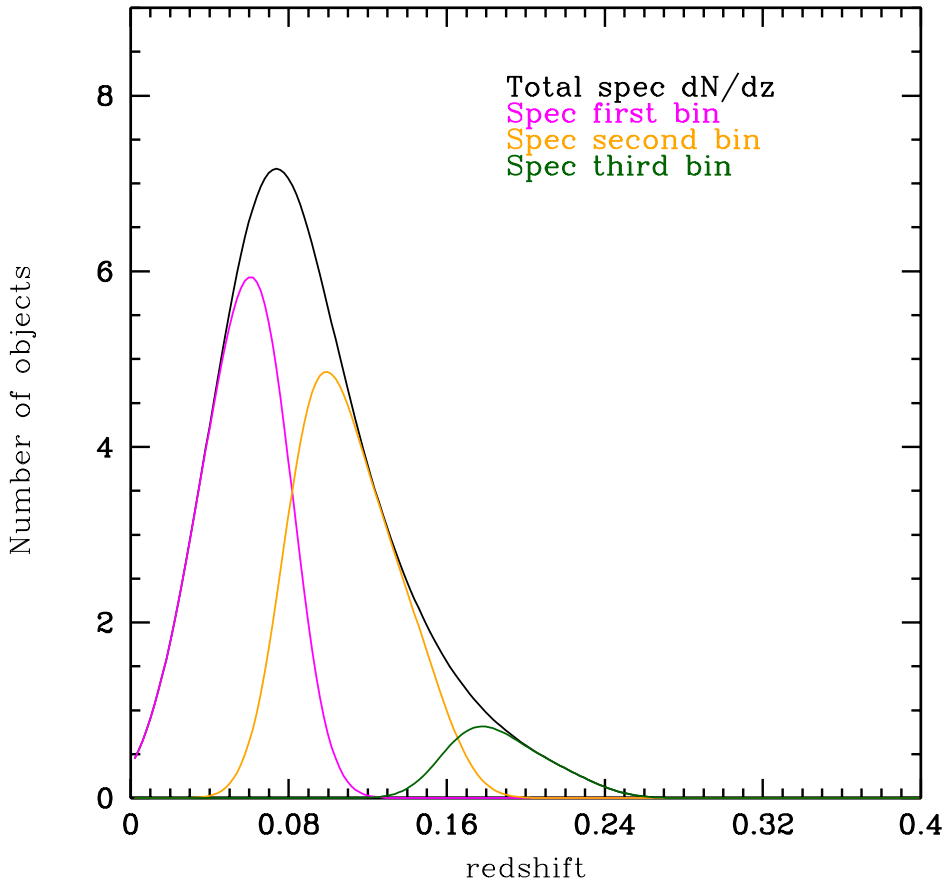

Figure 5.9: The full spectroscopic redshift distribution of the 2MPZ catalogue (black solid) compared with the spectroscopic distributions obtained for the three z-bins under study after having performed the convolution. The magenta solid line corresponds to the first photometric bin ($0 < z_{\rm phot} < 0.08$), the orange line to the second photometric bin ($0.08 < z_{\rm phot} < 0.16$) and the green line to the third photometric bin ($0.16 < z_{\rm phot} < 0.24$).

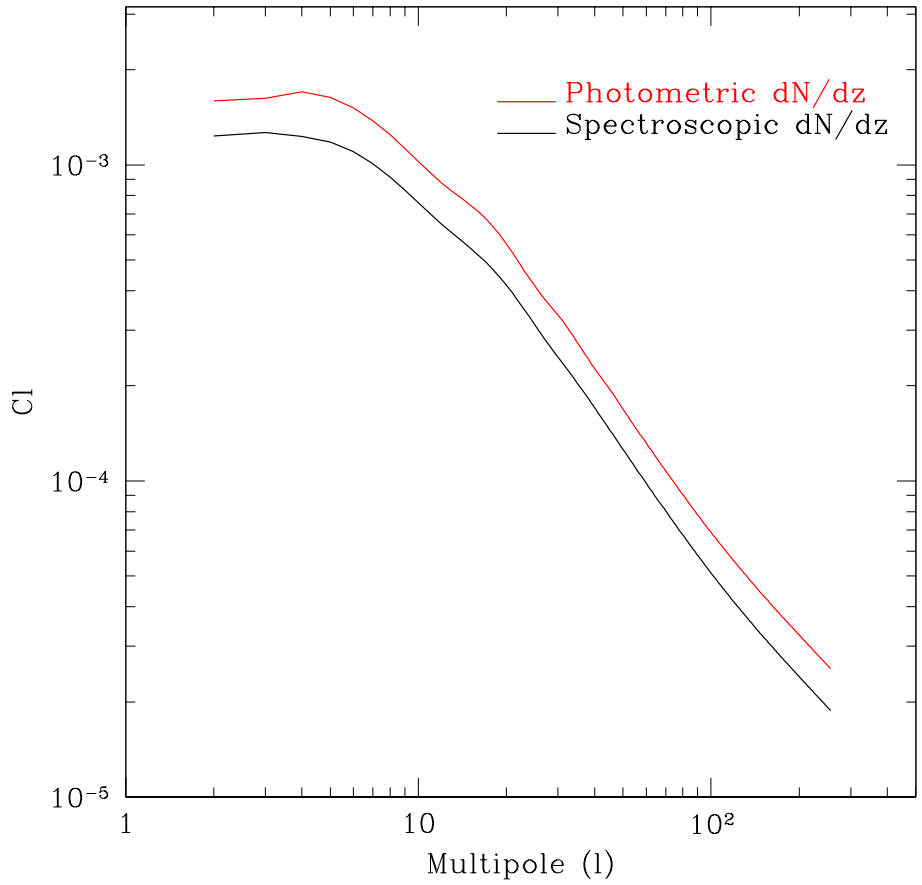

Figure 5.10: Two model angular spectra obtained from `CLASS` for the second redshift bin ($0.08 < z_{\mathrm{phot}} < 0.16$) in the case of the photometric redshift distribution (red solid) or the spectroscopic counterpart (black solid).

maps while the red solid line is the mean of the convoluted spectra. The bottom panel shows the % difference between the masked and convolved case. The convolution with the mixing matrix is effective, in the sense that theoretical predictions convolved with this matrix provides good fit to the measured angular spectrum.

5.3.4 Covariance matrix from the mocks

The second major goal of the mock catalogue, beyond the investigation of systematic effects, is that of estimating the random error and their covariance or, more specifically, to estimate the covariance matrix of the estimated angular power spectrum.

The latter is defined as follows:

$$\mathrm{Cov}(\ell,\ell') = \frac{1}{N_{\mathrm{mocks}}} \sum_{i,j=1}^{N_{\mathrm{mocks}}} \left(C_\ell^i - \bar{C}_\ell^i\right)\left(C_{\ell'}^j - \bar{C}_{\ell'}^j\right) \tag{5.18}$$

where N_{mocks} is the total number of mock catalogues, while $\bar{C}_\ell^i$ is the mean angular spectrum over the mock catalogues:

$$\bar{C}_\ell^i = \frac{1}{N} \sum_{i=1}^{N_{\mathrm{mocks}}} C_\ell^i \tag{5.19}$$

The covariance matrix allows one to quantify the correlation between the different multipoles of the spectrum, which is often expressed in terms of the **correlation coefficients**:

$$\rho_{\ell,\ell'} = \frac{\mathrm{Cov}(\ell,\ell')}{\sqrt{\mathrm{Cov}(\ell,\ell)\mathrm{Cov}(\ell',\ell')}} \tag{5.20}$$

The covariance matrix is a necessary input to both the χ^2 and MCMC likelihood analysis that we shall perform in the next chapter.

For the full-sky case, and as long as linear theory holds, the multipoles are uncorrelated, hence the covariance matrix is diagonal. The presence of a geometry mask, instead, induces correlation among the multipoles, so that in general the off-diagonal elements of the covariance matrix differ

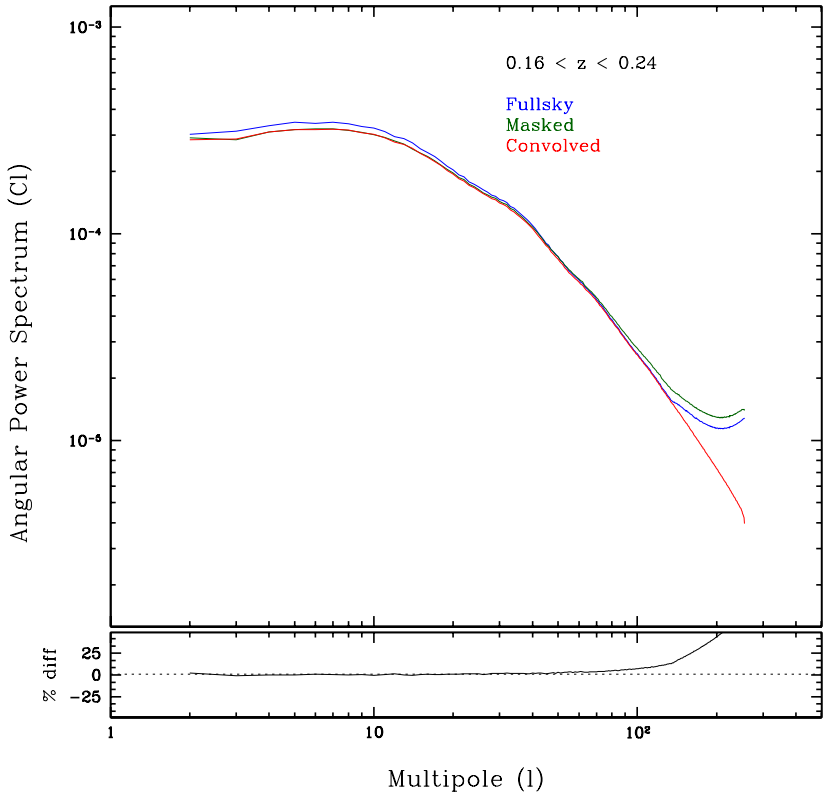

Figure 5.11: Comparison between the mean of the spectra computed from the masked lognormal map and the convolution with the mixing matrix **Eq. 3.34** for the third bin ($0.16 < z < 0.24$). The solid blue line is the mean over 1000 realizations of the spectra computed from full-sky lognormal maps, the solid green line is the mean of the spectra computed from the masked lognormal maps while the red solid line is the mean of the convolution spectra. The bottom panel shows the percent difference between the masked and convolved case.

from zero. In our case, however, we have some reason to suspect that the off-diagonal elements should be small:

1. We generate mocks using a linear Gaussian field. This means that off-diagonal terms are only induced by the geometry mask, and the effect of nonlinear evolution is ignored. However, we expect that this second contribution should be small since, as we have already pointed out, the projection effects of an angular map, smooth out nonlinear features.

2. The effect of the mask is larger on large angular scales. However, these scales will be excluded from the analysis since they are most prone to systematic errors induced by imperfect corrections for Galactic extinction and Zodiacal light emission.

Before checking if the above prejudice is correct, let us elaborate on the estimate of the covariance matrix a bit further. The precision in estimating the elements of the matrix is determined by the number of independent mocks. As a rule of thumb, a matrix with $N \times N$ elements would require at least $N \times N$ mocks. Since we have generated 144 mocks this means that we can afford only 12 entries. For this reason we bin our measured spectrum in 12 linearly spaced intervals ranging from $\ell = 0$ to $\ell = 144$. **Fig. 5.12**, **5.13** and **5.14** show the normalized covariance matrices of the angular spectra in the three redshift bins considered. The amplitude of the coefficient is color coded according to correlation amplitude. Green and blue represents large vs small correlation. It is quite clear that the matrix is close to diagonal. The off-diagonal elements have amplitudes 3 to 10 times smaller than the diagonal ones.

This confirms our prejudice and justifies our choice to ignore in the forthcoming χ^2 and MCMC analyses the off-diagonal term (although our intention is to perform in the future a MCMC analysis with the full covariance matrix).

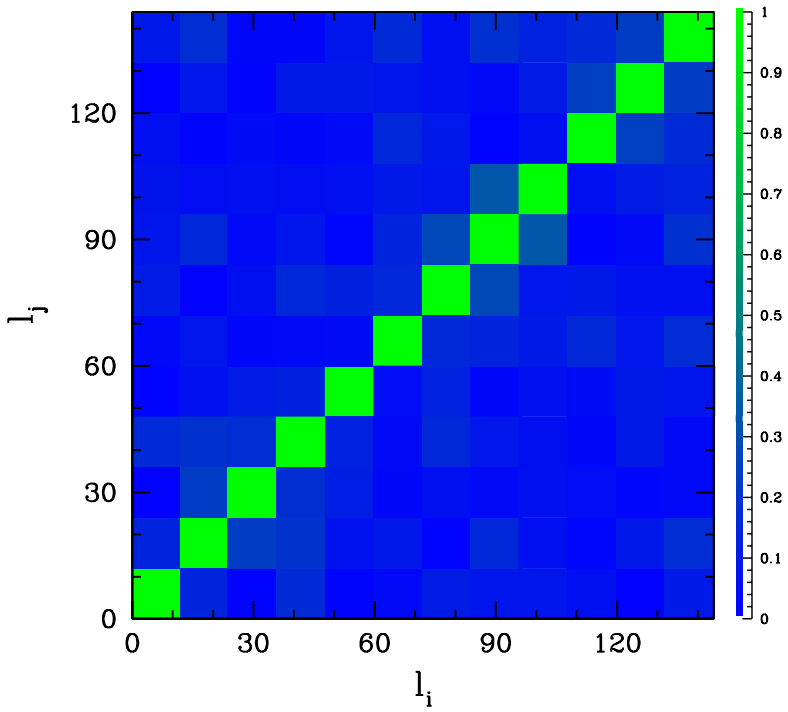

Figure 5.12: The correlation matrix computed from 144 mocks 2MPZ-like for the redshift bin corresponding to $0 < z < 0.08$.

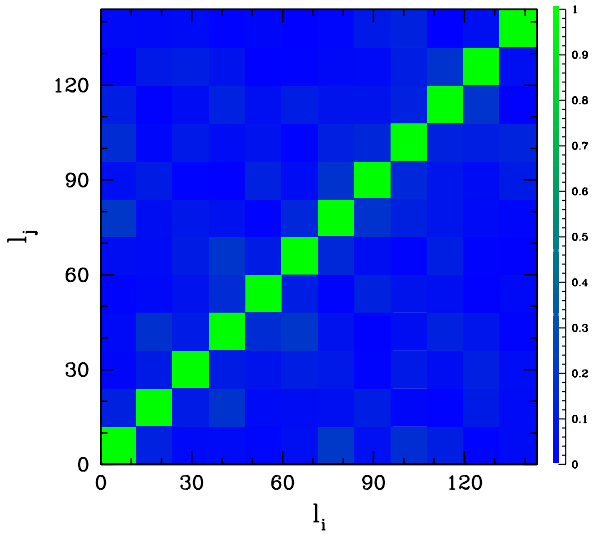

Figure 5.13: The correlation matrix computed from 144 mocks 2MPZ-like for the redshift bin corresponding to $0.08 < z < 0.16$.

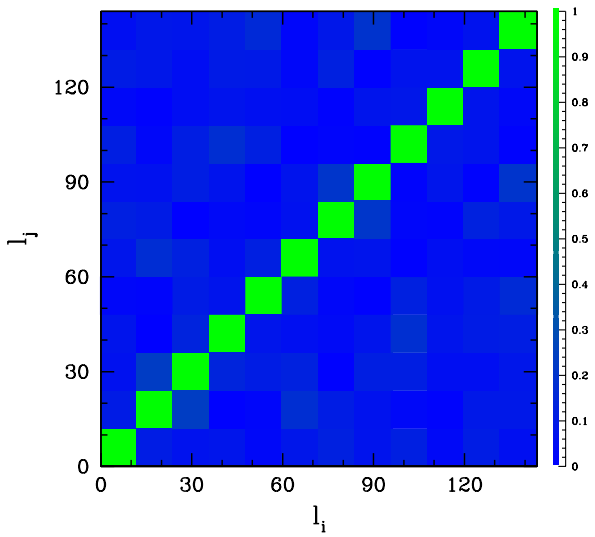

Figure 5.14: The correlation matrix computed from 144 mocks 2MPZ-like for the redshift bin corresponding to $0.16 < z < 0.24$.

Chapter 6

Statistical analysis of the mock catalogues

The goal of this Section is to introduce, calibrate and validate two methods to estimate cosmological parameters from angular spectra that we shall apply to the real dataset in **Chapter 7**. The first one is the standard χ^2 statistics. The second one is the *Monte Carlo Markov Chain* sampling of the likelihood function. In this Chapter we apply them to the set of mock catalogs (both full sky and masked) presented in **Chapter 5**. The goal is to estimate cosmological parameters by comparing the measured angular spectrum to model predictions. We shall focus on the estimate of three parameters: the dark matter density parameter, Ω_{CDM}, the spectral amplitude A_s and the baryon density Ω_b. We shall perform two different analysis. Given the sizable errors expected from a purely angular clustering analysis, we won't vary all three parameters simultaneously. Instead we will vary only two or one of them, keeping the others fixed at their true value. We shall perform our analysis in the same redshift bins used for the real analysis and will restrict ourselves, in some tests, to a single redshift bin.

6.1 Statistical tools

6.1.1 The χ^2 analysis

The standard χ^2 statistics allows us to determine the model $y_{\rm model}$ that best fits a set of data $y(x_i)$. The model, that depends on a number (M) of free parameters $(a_0...a_{M-1})$, is evaluated at the positions x_i and compared to the measurement to form the χ^2 statistics that, in the case of independent measurements, is (Press et al., 1992):

$$\chi^2 \equiv \sum_{i=0}^{N-1} \left(\frac{y_i - y(x_i|a_0...a_{M-1})}{\sigma_i} \right)^2 \tag{6.1}$$

where σ_i is measurement error (assuming that the model carries no uncertainties), N is the number of independent measurements and $N - M$ is the number of degrees of freedom (d.o.f.). In our analysis the measured quantity is the angular power spectrum evaluated in a set of ℓ-bins of size Δ_ℓ, $C_{[\ell]}$, and its random error σ_ℓ that we estimate from the scatter among the mocks. In the following we shall omit the square parentheses and use the ℓ symbol both for the binned and unbinned quantities. In our case the χ^2 statistics is then:

$$\chi^2 = \sum_{\ell-\rm bins} \left(\frac{C_\ell^{\rm model}(\Omega_{CDM}) - C_\ell^{\rm experiment}}{\sigma_\ell^2} \right)^2 \tag{6.2}$$

We use this χ^2 statistics for different purposes. 1) Assuming negligible systematic errors we use **Eq. 6.2** to estimate parameters from the spectrum measured in the dataset. In this case the quantity σ_ℓ is computed from the scatter among the mocks. 2) To detect possible systematic errors. In this case we use the mean spectrum obtained by averaging among the mocks and the uncertainties are the errors on the mean.

Eq. 6.2 assumes no covariance among the errors. The covariance matrix we built from the mock justifies this hypothesis.

The best fitting parameters are found at the minimum of the χ^2 function of **Eq. 6.2**. To estimate the error on these parameter, we assume that

the data follow a Gaussian statistics and use the difference from its minimum value, $\Delta\chi^2$, to estimate the confidence level. The relation depends on the number of free parameters in the analysis (see e.g. (Press et al., 1992)). For the case of 1 and 2 parameters the region containing the 68.27% probability (i.e. 1σ) is defined by:

$$\Delta\chi^2 = 1 \quad \text{and} \quad \Delta\chi^2 = 2.3 \tag{6.3}$$

respectively. The χ^2 method has some drawbacks the most serious of which is computational requirement. **Eq. 6.2** needs to be evaluated in correspondence of different sets of free parameters $(a_0...a_{M-1})$ which implies calling `CLASS` to evaluate the angular spectrum and, in case of mask, to perform a convolution. In so doing we need to regularly and densely sample the parameter space. With no *a priori* information on the expected parameter values, and in the case of multi-dimensional analysis, this search becomes computationally challenging. In such cases a smarter searching strategy, like the MCMC one described in the next section, is then required.

6.1.2 `MontePython`

An alternative powerful way to infer cosmological information from large scale clustering is represented by the *Monte Carlo Markov Chains* or MCMC (Metropolis et al. (1953), Kirkpatrick et al. (1983) among others). We will not go into details, but the huge advantage of MCMC is that it efficiently sample the multidimensional parameter space by preferentially visiting the regions in which the likelihood is higher. The ability of effectively sample the likelihood relies on two MCMC properties.

The first one is that the distribution function $\pi(\vec{x})$, proportional to the probability of visiting a point $\vec{x}$ of the parameter space, can be sampled not via unrelated, independent points, but rather by a **Markov chain**. This latter is a sequence of points $\vec{x}_0$, $\vec{x}_1$, $\vec{x}_2$...where each point $\vec{x}_i$ is chosen from a distribution that depends only on the value of the immediately preceding point $\vec{x}_{i-1}$. In other words, the memory of the chain extends only to

one previous point, and it's completely defined by a transition probability function of two variables $p(\vec{x}_i|\vec{x}_{i-1})$.

In addition to this, if $p(\vec{x}_i|\vec{x}_{i-1})$ is chosen to satisfy the *detailed balance equation*:

$$\pi(\vec{x}_1)p(\vec{x}_2|\vec{x}_1) = \pi(\vec{x}_2)p(\vec{x}_1|\vec{x}_2) \tag{6.4}$$

then it can be showed that the Markov chain will sample $\pi(\vec{x})$ ergotically, i.e. it will eventually visit every point $\vec{x}$ in proportion to $\pi(\vec{x})$. The ergodicity guarantees that the equilibrium distribution is rapidly approached from *any* starting point $\vec{x}_0$[1]. It is thus necessary to find a transition probability $p(\vec{x}_i|\vec{x}_{i-1})$ that satisfies the **Eq. 6.4**. To do this, we use the popular Hastings algorithm (Metropolis et al. (1953) and Hastings (1970)) in which one chooses as a proposal distribution $q(\vec{x}_2|\vec{x}_1)$, for example, a multivariate normal distribution centered on $\vec{x}_1$. In order to generate a step starting at $\vec{x}_1$, one have to generate a candidate point $\vec{x}_{2c}$ by drawing from the proposal distribution. An *acceptance probability* $\alpha(\vec{x}_1, \vec{x}_{2c})$ is then computed as:

$$\alpha(\vec{x}_1, \vec{x}_{2c}) = \min\left(1, \frac{\pi(\vec{x}_{2c})q(\vec{x}_1|\vec{x}_{2c})}{\pi(\vec{x}_1)q(\vec{x}_{2c}|\vec{x}_1)}\right) \tag{6.5}$$

According to this probability, the candidate point $\vec{x}_{2c}$ is accepted or rejected (in which case the point is left unchanged, $\vec{x}_2 = \vec{x}_1$). The net result of this process is a transition probability:

$$p(\vec{x}_2|\vec{x}_1) = q(\vec{x}_2|\vec{x}_1)\alpha(\vec{x}_1, \vec{x}_2) \qquad \vec{x}_2 \neq \vec{x}_1 \tag{6.6}$$

that can be proved to satisfy the **Eq. 6.4**.

To perform MCMC we use the publicly available `MontePython` code (Audren et al., 2013), interfaced with `CLASS`. The structure of the code is

[1]Of course, if we start from a very unlikely point, successor points will themselves be quite unlikely until we rejoin a more probable part of the distribution. There is thus a need of *burn-in* an MCMC chain by stepping to, and discarding, a certain number of points $\vec{x}_i$.

simple. A *parameter file* allows one to specify the parameters involved in the study, their role (for example *derived*, *nuisance* etc.), the user guess for their best fitting value and eventually the range in which the parameter can vary. These parameters are passed directly to `CLASS`. Measurements are stored in a *data* folder. Finally, an `_init_.py` file initializes the overall procedure, importing the data, the model, and computing the effective likelihood. At this point the MCMC starts sampling the parameter space through the Metropolis-Hastings algorithm. The number of steps in the chain can be set by the user. The output consists of the full chain and the χ^2 value corresponding to the parameters that have been varied. After some chains are produced, the `info` command allows one to decide upon their convergence, to output the best fit parameters, their credible intervals, joint and marginalized posterior probabilities as well as covariance matrices of the relevant free parameters.

One of the major advantage of using `MontePython` is that several likelihoods has been already written, each of them associated to a particular past experiment. Unfortunately, no such examples was available for the angular power spectrum. Therefore we had to write our own piece of program, which consists of a python code that reads in the data, call `CLASS` to generate model spectra, perform convolution with the mixing matrix (see **Eq. 3.47**) and multiplication by the pixel window function (**Eq. 3.63**) and, finally, computes the likelihood $\mathcal{L} \propto e^{-\chi^2}$, where χ^2 is the one provided in **Eq. 6.2** (or its version with more than one parameter).

6.2 Application to a single spectrum from a full-sky mock catalogue map

As a first application of the cosmological analysis, we estimate the parameters and their errors from the angular spectra measured over the full sky. As this is a sanity check intended to validate the statistical tools of our analyses, we shall consider the mean measured spectrum averaged over

10000 realization of the ΛCDM model used in **Chapter 3**, with $\Omega_b = 0.04$, $\Omega_{CDM} = 0.21$, $\Omega_\Lambda = 0.75$, $T_{cmb} = 2.7255$, $h = 0.73$, $A_s = 2.46 \cdot 10^{-9}$ (i.e. the blue dotted curve in **Fig. 3.5**).

6.2.1 χ^2 analysis with one free parameter (Ω_{CDM})

The results of the analysis are shown in **Fig. 6.1**. The black curve shows the χ^2 (**Eq. 6.2**) evaluated in 21 linearly spaces ℓ-bins of size $\Delta_\ell = 6.4$ from $\ell_{\rm min} = 1$ to $\ell_{\rm max} = 144$. We have divided the range $\ell = [\ell_{\rm min}, \ell_{\rm sup}]$ into 40 bins, with $\ell_{\rm min} = 0$ and $\ell_{\rm sup} = 4n_{side} = 256$ but performed the χ^2 analysis only in the first 21 out to $\ell_{\rm max} = 144$ corresponding to the angular scale of the pixel. $\ell_{\rm max}$ was kept fixed in all χ^2 analyses whereas we varied $\ell_{\rm min}$ to check the robustness of our results and the size of random error when very large angular scales are excluded from the analysis. Which, as we have anticipated, is what we do in the real data analysis to minimize the chance of potential biases.

The analysis recovers the correct Ω_{CDM} value irrespective of the $\ell_{\rm min}$. Increasing $\ell_{\rm min}$ has the effect of increasing the depth of the curve and of increasing its width, hence reducing the size of the random errors. This seemingly strange behavior (the smaller is the ℓ-range considered, the better is the fit) is due to the fact that cutting the low ℓ-values removes the part of the spectrum dominated by cosmic variance, so that the only contribution to the total error is shot noise.

Note that the sampling of Ω_{CDM} ($\Delta_{\Omega_{CDM}} = 0.01$) is larger than the 1σ random error estimated from the MCMC analysis presented in the next section. This means that this χ^2 analysis is not sensitive enough to estimate random errors, but still useful to detect systematic errors larger than the random ones if they were present.

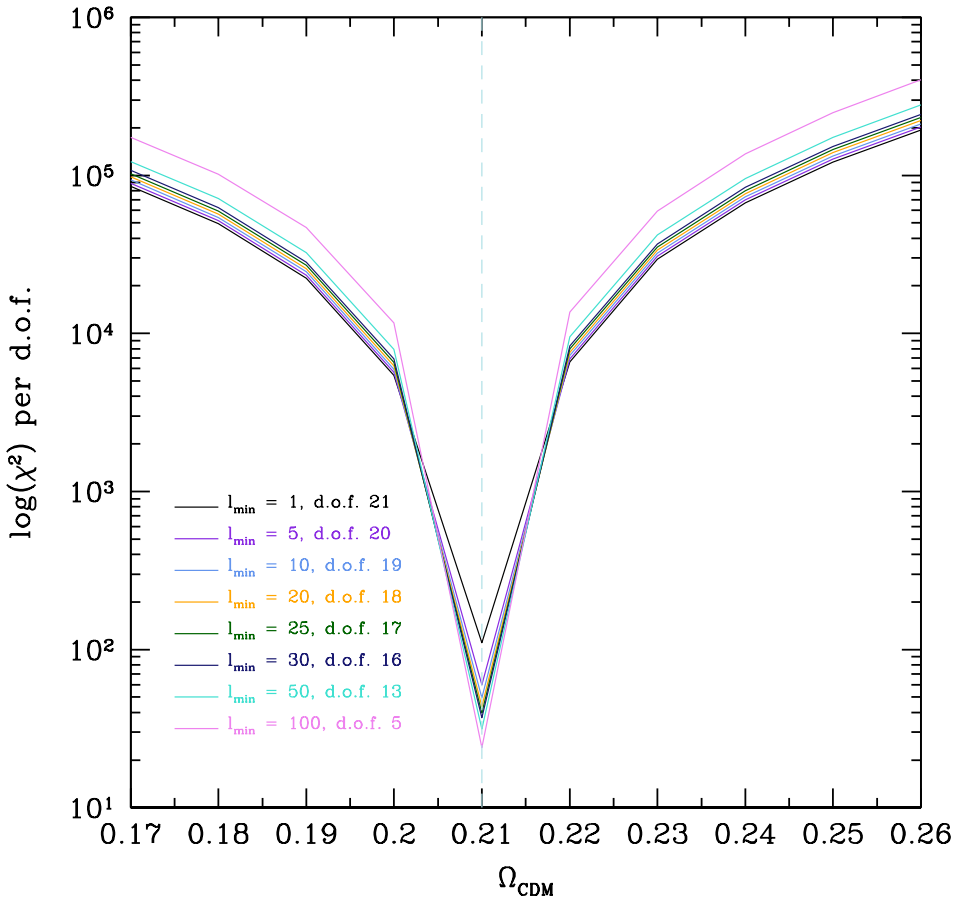

Figure 6.1: χ^2 curve for the Ω_{CDM} parameters computed using increasingly smaller fractions of the measured angular spectrum. All other cosmological parameters were set equal to their true values. All the minima coincide with the expected value (vertical dashed line).

6.2.2 `MontePython` analysis with one free parameter

The same analysis has been repeated by maximizing the likelihood rather than minimizing the χ^2 using the `MontePython` code (MP) to efficiently sample the likelihood function. In this case we consider only one spectrum measured in 20 bins from $\ell = 12$ to $\ell = 144$ (roughly corresponding to the blue curve in **Fig. 6.1**). The choice of $\ell_{\rm min} = 12$ is not arbitrary. It corresponds to the minimum ℓ-value below which the effects of a 2MPZ-like geometry mask are significant (see **Sec. 3.4.2**) and other potential large scale observational biases may be present. MP requires an input Ω_{CDM} value, that we set equal to 0.3. The prior range is set to $[0, 0.7]$. We then run 30 chains of 10000 steps each. The result is shown in **Fig. 6.2** in which we plot the posterior probability for Ω_{CDM}. The best fit value and its 1σ error turns out to be:

$$\Omega_{CDM,\,\text{best fit}} = 0.209 \pm 0.004 \tag{6.7}$$

as indicated in the plot. As in the χ^2 analysis, MP recovers the expected Ω_{CDM} value, excluding systematic errors larger than 2%.

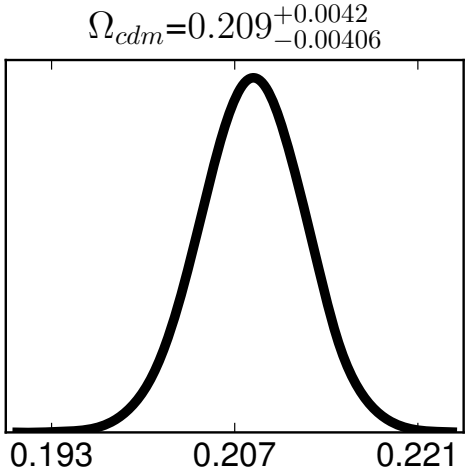

Figure 6.2: The posterior probability for Ω_{CDM} obtained with `MontePython` in the full-sky case (see text for details).

6.2.3 `MontePython` analysis with two free parameters

To further validate the MP procedure we have repeated the MCMC analysis letting two paramters: i) Ω_{CDM} and ii) A_s free to vary instead of one. We considered the same spectral range and binning as in the 1 parameter analysis. The guess values are 0.3 for Ω_{CDM} and 3×10^{-9} for A_s. The prior range is set to $[0, 0.7]$ for the former and $[0.1, 4] \times 10^{-9}$ for the latter. We then run 30 chains of 10000 steps each. **Fig. 6.3** shows the marginalized probability (bottom right and upper panels) for Ω_{CDM} and A_s as well as their joint probability (bottom left panel).

For the marginalized probability the best fit values and their 1σ uncertainties for the two parameters are:

$$\Omega_{CDM,\,\text{best fit}} = 0.23 \pm 0.05 \tag{6.8}$$

$$A_{s,\,\text{best fit}} = (2.0 \pm 0.8) \times 10^{-9} \tag{6.9}$$

In both cases we recover the true values. However, neither probability is symmetric around the best fit value. They're both significantly skewed to larger values. This reflect the degeneracy between the spectral amplitude and the dark matter density parameter that is clearly illustrated by the contours of the joint probability. A degeneracy that, as we shall see, will prominently feature also the real data analysis.

6.3 Application to a single spectrum from a masked map

We repeat the analysis performed in the previous section including the effect of the geometry mask. Model predictions now include the effect of the mask, obtained as described in **Sec. 3.4.2**.

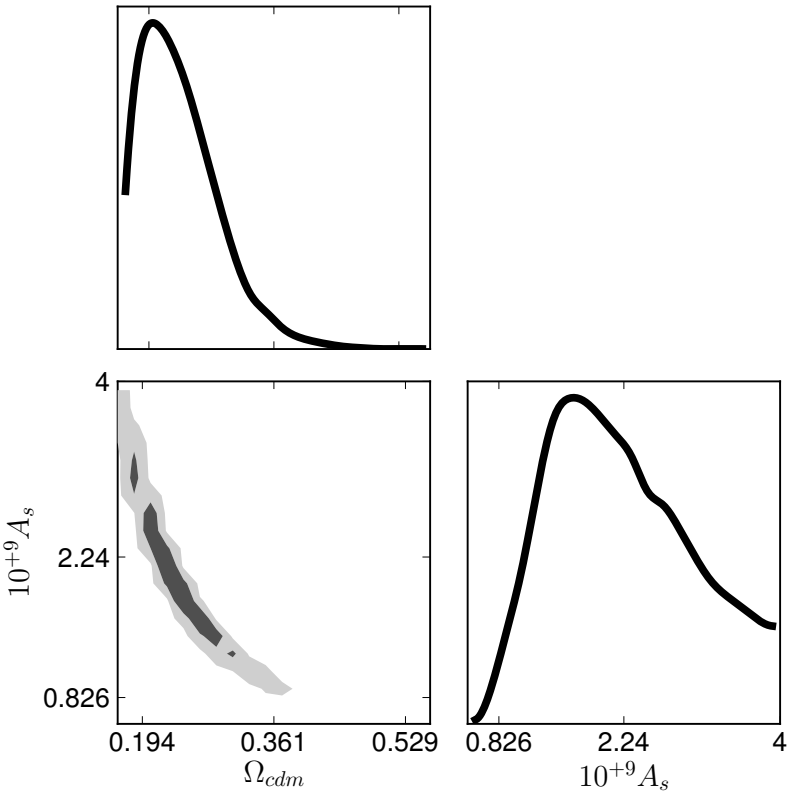

Figure 6.3: Bottom right panel and upper panel: the marginalized posterior probability obtained with `MontePython` in the full-sky case with Ω_{CDM} and A_s as free parameters. Bottom left panel: the two dimensional posterior distribution. See text for details.

6.3.1 χ^2 analysis with one free parameter (Ω_{CDM})

Fig. 6.4 shows the results of our χ^2 analysis on a grid of 10 values of Ω_{CDM}, from 0.17 to 0.26 with step 0.01. Again the χ^2 values have been normalized to the degrees of freedom, that depend on the $\ell_{\rm min}$ value adopted. The value of $\ell_{\rm max}$ has been again set to 144, so we have considered up to 22 ℓ-bins.

As in the previous case, the Ω_{CDM} binning is slightly larger than the random error. Therefore, the lack of a offset from the expected value indicates that systematic errors are not larger than the random ones. This is valid for all cases explored except when $\ell_{\rm min} = 100$. In this case the measured value is significantly larger than the expected one. This systematic error arises from the fact that the geometry mask shifts power from large to small scales. Since with $\ell_{\rm min} = 100$ we only sample the latter and since the amplitude of the spectrum is equal to its true value, the best fit is found for higher Ω_{CDM} values. This result indicates that one cannot restrict the analysis on small angular scales assuming that they're less affected by masking. However, we point out that even in the extreme case of $\ell > 100$, the effect is rather minor ($\sim 5\%$).

6.3.2 `MontePython` analysis with one free parameter

Fig. 6.5 shows the posterior probability as the result of the analysis on the masked spectrum with `MontePython`. As in the previous analysis, the measured spectrum is a mean over 10000 realizations which, however, now consist of masked maps. The expected Ω_{CDM} value, prior ranges and number of chains are the same as in **Sec. 6.2.2**. Also in this case we consider the angular spectrum in 20 multipole bins that span the range $\ell = [12, 144]$. The best fit value is equal to:

$$\Omega_{CDM,\,\text{best fit}} = 0.207 \pm 0.004 \tag{6.10}$$

Also in this case we succeed in recovering the expected value within the 1σ random error. The size of the errorbar is similar to that we have obtained

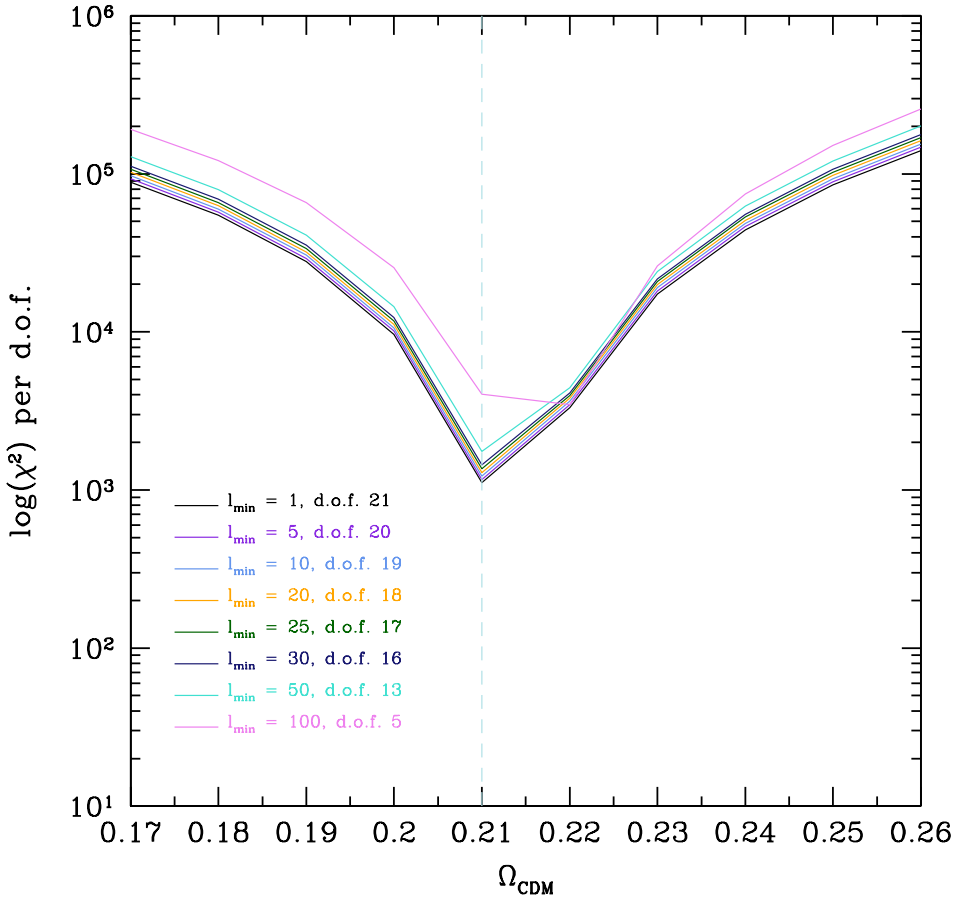

Figure 6.4: Same as **Fig. 6.1** but for the case of partial sky coverage.

in the full sky case. This is not surprising since in both analyses we ignored multipoles $\ell < 12$, where cosmic variance dominates the error budget.

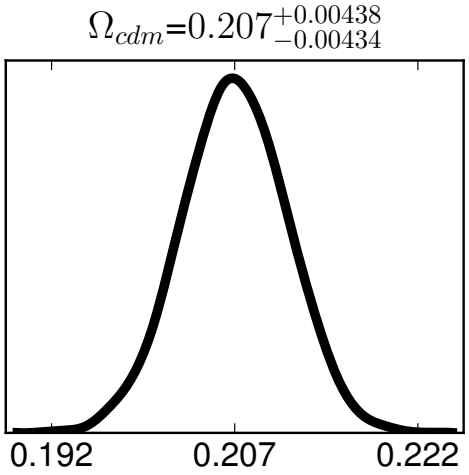

Figure 6.5: Same as **Fig.6.2** but for the case of partial sky coverage.

6.3.3 `MontePython` analysis with two free parameters

Similarly to **Sec. 6.2**, we now consider two free parameters, Ω_{CDM} and A_s. The results, shown in **Fig. 6.6**, are quite similar to those of the full-sky case. The best fitting values and their random errors are:

$$\Omega_{CDM,\text{ best fit}} = 0.26 \pm 0.04 \tag{6.11}$$

$$A_{s,\text{ best fit}} = (1.9 \pm 0.6) \times 10^{-9} \tag{6.12}$$

We note that in this case the mask has introduced a systematic error of the same order of the random one. As a result, and because of the parameter degeneracy, the overestimate in the Ω_{CDM} value is compensated by an underestimate of the spectral amplitude A_s.

The fact that random errors are similar to those obtained in the full-sky case confirms that when we restrict the analysis to large ($\ell > 12$) multipoles we largely remove the effect of cosmic variance. Random errors

are dominated by the shot-nose component which is the same as in the full-sky case.

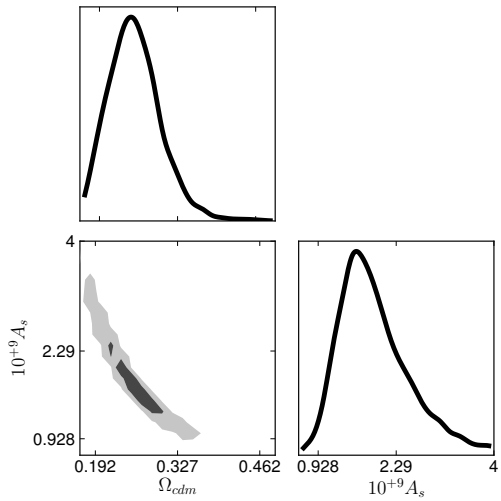

Figure 6.6: Same as **Fig. 6.3** but for the case of partial sky coverage.

6.4 Realistic case, angular spectra in 3 redshift bins

The previous tests allows us to validate the two methods used for the statistical analysis. We know want to apply one of them (namely, the χ^2 analysis) to the 2MPZ mocks described in **Chapter 4**. These mocks include the effect of the geometry mask. However, the effect of photo-z error is ignored here and will only be included in the next Chapter. The analysis will be performed using the angular power spectra estimated in the three z-bins.

The main goal of the analysis is to evaluate the systematical errors in the estimate of the cosmological parameters and highlight parameters' degeneracies. We focus on three cosmological parameters: the dark matter and the baryon number densities Ω_{CDM} and Ω_b and the spectral amplitude A_s. The statistical error on the mean of the measurements is evalu-

ated as the `rms` scatter among the mocks divided by the square root of the number of mocks. The systematic errors on the parameters will be quantified by the offset between the best fit Ω_b, Ω_{CDM} and A_s values from the theoretical ones.

We perform different χ^2 analyses. First, we estimate Ω_{CDM} keeping fixed Ω_b and A_s. Then, we vary only Ω_{CDM} and Ω_b, keeping fixed the spectral amplitude. Finally, we fix Ω_b to its input value and let vary Ω_{CDM} and A_s. These analyses are the same we perform on the real data in **Sec. 7.2**. Best fit values are searched for on a grid in which Ω_{CDM} ranges from 0.15 to 0.27 with step 0.01, Ω_b ranges from 0.01 to 0.08 with step 0.005 and A_s ranges in $[2.00, 3.99] \times 10^{-9}$ with step 0.01×10^{-9}.

We look for a best fit in the multipole range $\ell = [12, 100]$. The motivations for this choice are the same as in the previous sections $\ell = 12$ is the minimum value below which the geometry mask effects are significant, while $\ell = 100$ corresponds to an angular scale safely larger than the pixel size. As anticipated, since the covariance matrix computed from the mocks is almost diagonal (see **Sec. 5.3.4**), we ignore covariance among the ℓ-modes.

We stress the fact that since mock catalogs were generated assuming a linear power spectrum, the model predictions with which we compare the estimated spectra were also obtained (from `CLASS`) using linear theory.

6.4.1 χ^2 analysis with one free parameter (Ω_{CDM})

First redshift bin, $0 < z < 0.08$

The χ^2 statistics shown in **Fig. 6.7** is the mean of the mock catalogues (solid curve). The 1σ scatter, obtained in correspondence of $\chi^2 - \chi^2_{\rm min} \equiv \Delta\chi^2 = \pm 1$ is shown with a horizontal dotted red line. The minimum is found at $\Omega_{CDM} = 0.22$ with an offset of $\Delta_{\Omega_{CDM}} = 0.01$ from the expected value ($\Omega_M = 0.21$). The 1σ random error turns out to be smaller than the resolution of the Ω_{CDM} grid. Therefore, we can conclude that the systematic

error is small in amplitude but slightly larger than the random one.

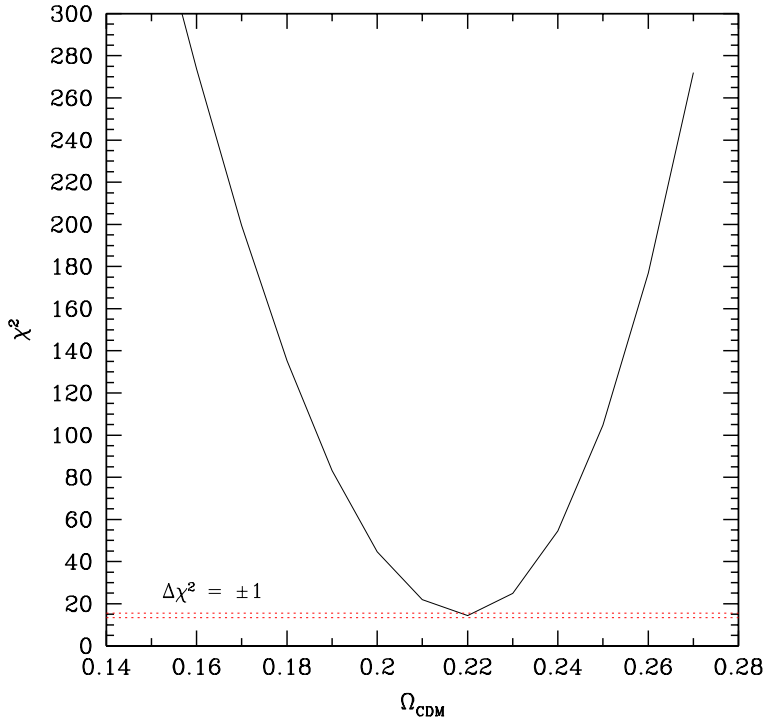

Figure 6.7: Solid line: the mean χ^2 obtained from the mocks fixing all cosmological parameters to their true values but Ω_{CDM}. Red dotted line: the 1σ scatter $\Delta\chi^2 = \pm 1$.

Second redshift bin, $0.08 < z < 0.16$

The results from the second z-bin are shown in **Fig. 6.8**, which is analogous to **Fig. 6.7**. The plot shows the mean χ^2 of the mocks as a function of Ω_{CDM}, the only free parameter in the fit. Again, we detect a systematic offset of $\Delta_{\Omega_{CDM}} = 0.01$ since the minimum is found at $\Omega_{CDM} = 0.22$ instead of $\Omega_{CDM} = 0.21$, and the random error is again smaller than the resolution of the grid.

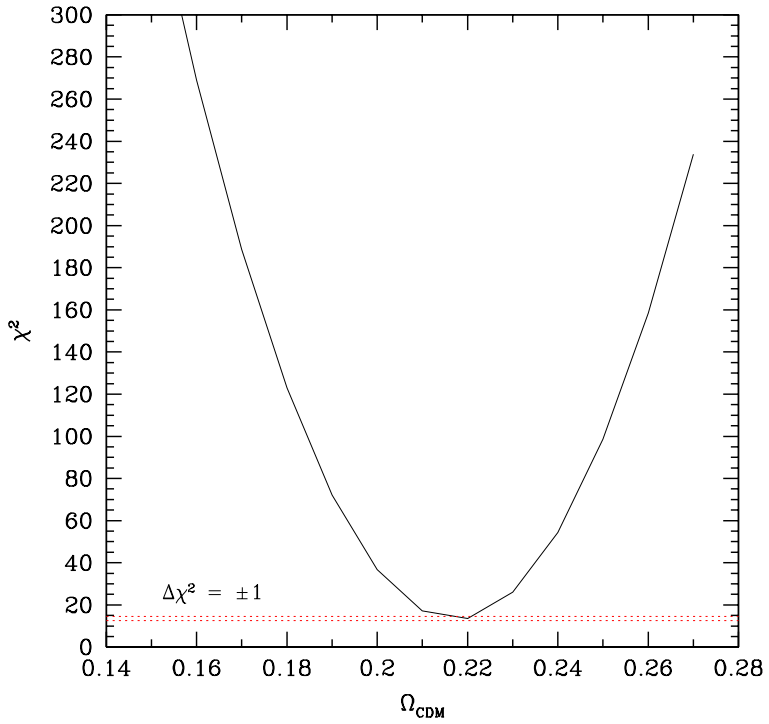

Figure 6.8: Same as **Fig. 6.7** but for the second redshift bin.

Third redshift bin, $0.16 < z < 0.24$

The results for the third bin are shown in **Fig. 6.9**. Here the best fit is found for $\Omega_{CDM} = 0.22$ and the random error, that here matches the size of the Ω_{CDM}-bin, is comparable to the systematic offset.

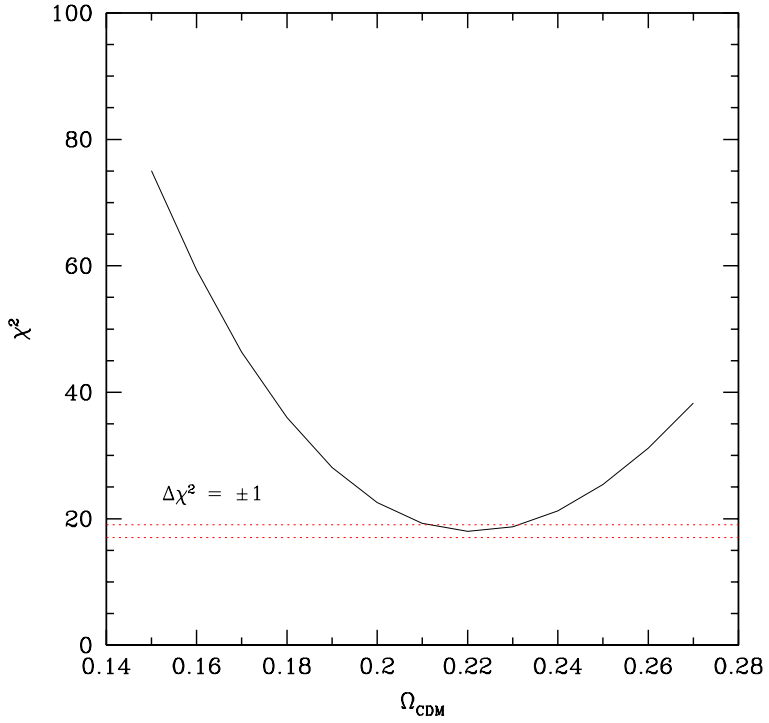

Figure 6.9: Same as **Fig. 6.7** but for the third redshift bin.

6.4.2 χ^2 analysis with two free parameters (Ω_{CDM} and Ω_b)

First redshift bin, $0 < z < 0.08$

Fig. 6.10 shows the result of the χ^2 analysis when varying both Ω_{CDM} and Ω_b. The blue, magenta and black solid lines represent the 1σ, the 2σ and the 3σ error contours. They are not centered around the expected values (black dot in figure, corresponding to $\Omega_{CDM} = 0.21$, $\Omega_b = 0.04$). In fact

the best fitting values ($\Omega_{CDM} = 0.24$, $\Omega_b = 0.06$) are systematically larger than the reference one and the offset is more than 3σ.

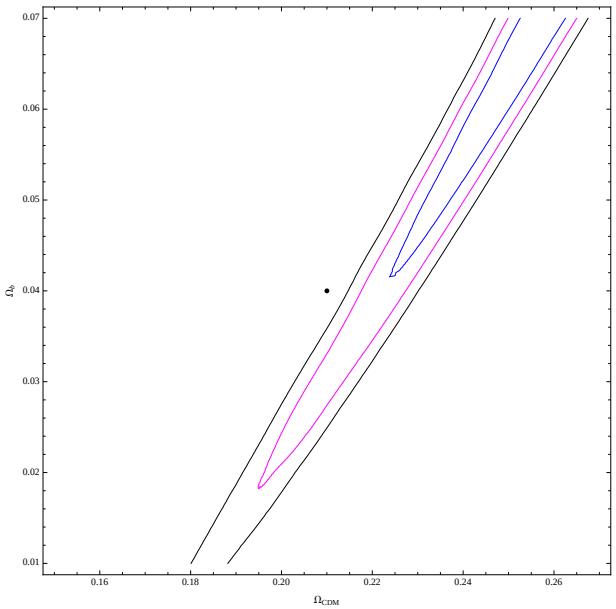

Figure 6.10: Solid curves: 1σ, 2σ and 3σ contours (blue, magenta and black colors) corresponding to values of $\Delta\chi^2 = 2.30,\ 6.18$ and 11.8 above the minimum value obtained by averaging over the χ^2 of all the mocks. The black dot shows the expected value ($\Omega_{CDM} = 0.21$, $\Omega_b = 0.04$).

The contours also highlights the almost linear relation between Ω_{CDM} and Ω_b, which explains why most analyses compare the baryon fraction Ω_b/Ω_{CDM} to Ω_{CDM}. In this test we prefer to stick to the Ω_{CDM} vs. Ω_b representation. The best fitting values are $\Omega_{CDM} = 0.25$ and $\Omega_b = 0.065$. They overestimate the true value by $\Delta_{\Omega_{CDM}} = 0.04$ and $\Delta_{\Omega_b} = 0.025$, respectively.

The offset between the true and the best fit values quantifies the systematic errors induced by the mask. As seen, this latter produces a loss of power in the angular power spectrum that, in turns, induces a systematic offset in the Ω_{CDM} and Ω_b values. In the 1-parameter case this causes Ω_{CDM} to be overestimated. The effect is amplified when adding one degree of freedom and considering two parameters that are positively

correlated.

The loss of power induced at large angular separation by the mask is evident in **Fig. 6.11** in which we compare the input angular spectrum (black solid curve) with the mean over the mocks (blue solid line). The effect is as large as 50% at large angular separations (bottom panel of the same figure).

Second redshift bin, $0.08 < z < 0.16$

In **Fig. 6.12** is shown the result of the χ^2 analysis on the two parameters Ω_{CDM} and Ω_b for the second bin. The best fit values are $\Omega_{CDM} = 0.22$ and $\Omega_b = 0.04$. The blue solid curve represents the 1σ probability contour, the magenta one the 2σ contour while the black one the 3σ contour. The black dot shows the position of the expected values ($\Omega_{CDM} = 0.21$, $\Omega_b = 0.04$). In the second redshift bin we do not detect any systematic error. Probably this is because the effect of the geometry mask is now less severe, in the sense that in a nearby redshift shells the physical scales that are affected by the mask are smaller than in a far away shell. This means that in the distant shell more cosmological information can be extracted from the ℓ-range unaffected by the mask. **Fig. 6.13**, in which we show the measured vs. the input spectrum, confirms this interpretation.

Third redshift bin, $0.16 < z < 0.24$

Fig. 6.14 is the analogous of **Figs. 6.12** and **6.10**. As seen in the one parameter estimate, the random errors are significantly larger than the other z-bins, as testified by area within the contours in **Fig. 6.14**.

The lack of significant systematic errors is due to the minor impact of the mask and, consequently, the reduced loss of power is, again, quantified by **Fig. 6.15**, that, in analogy with **Figs. 6.11** and **6.13**, compare the input power spectrum with the mean over the mocks.

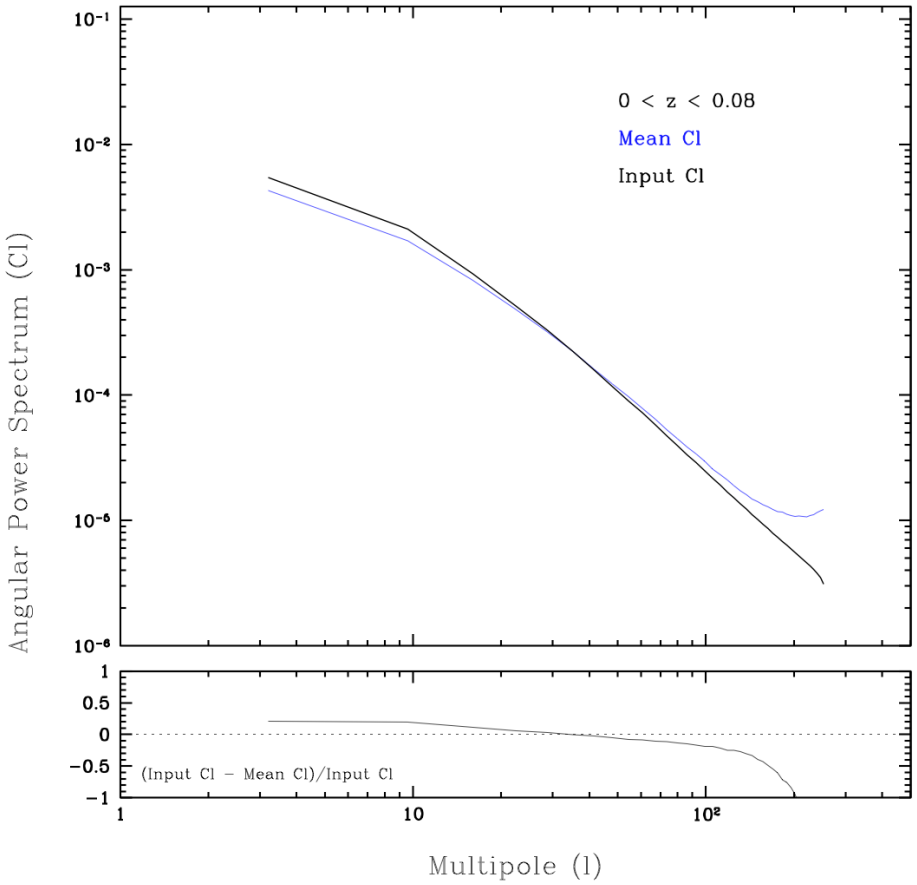

Figure 6.11: Upper panel: the mean of the angular power spectrum computed from the 144 2MPZ-like mocks (blue solid line) compared with the input convolved spectrum obtained from `CLASS` (black solid line). Bottom panel: the relative difference between the input spectrum and the mean over the angular spectra of the mocks.

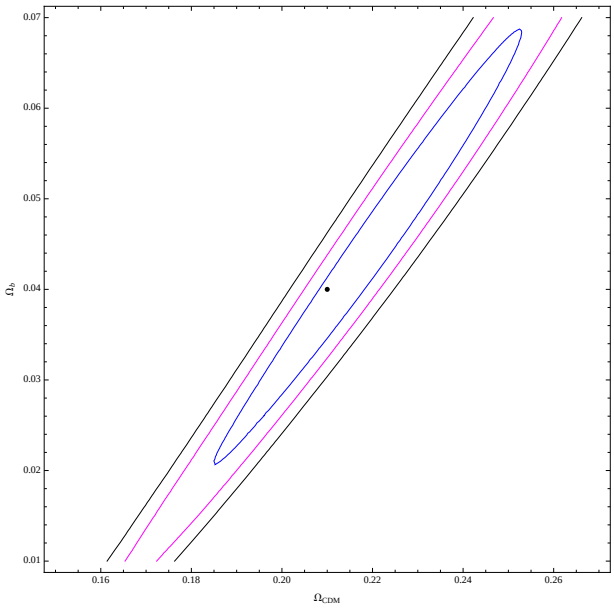

Figure 6.12: Same as **Fig. 6.10** but for the second redshift bin.

6.4.3 χ^2 analysis with two free parameters (Ω_{CDM} and A_s)

First redshift bin, $0 < z < 0.08$

Finally, we searched for the best fitting Ω_{CDM} and A_s values when all other parameters were fixed. The results are shown in **Fig. 6.16**, which, analogously to **Fig. 6.10**, shows the 1σ, 2σ and 3σ probability contours together with the expected values ($\Omega_{CDM} = 0.21$ and $A_s = 2.46 \times 10^{-9}$, black dot). Here the best fitting values, $\Omega_{CDM} = 0.27$ and $A_s = 1.7 \times 10^{-9}$, are systematically offset from the true ones. The origin of that shift, that confirms the overestimate in the Ω_{CDM} value seen in the previous section, is the effect of the geometry mask, as we have already discussed.

We notice that A_s and Ω_{CDM} are not independent. In fact they are significantly anti-correlated, a feature that we shall detect in the analysis of the real data.

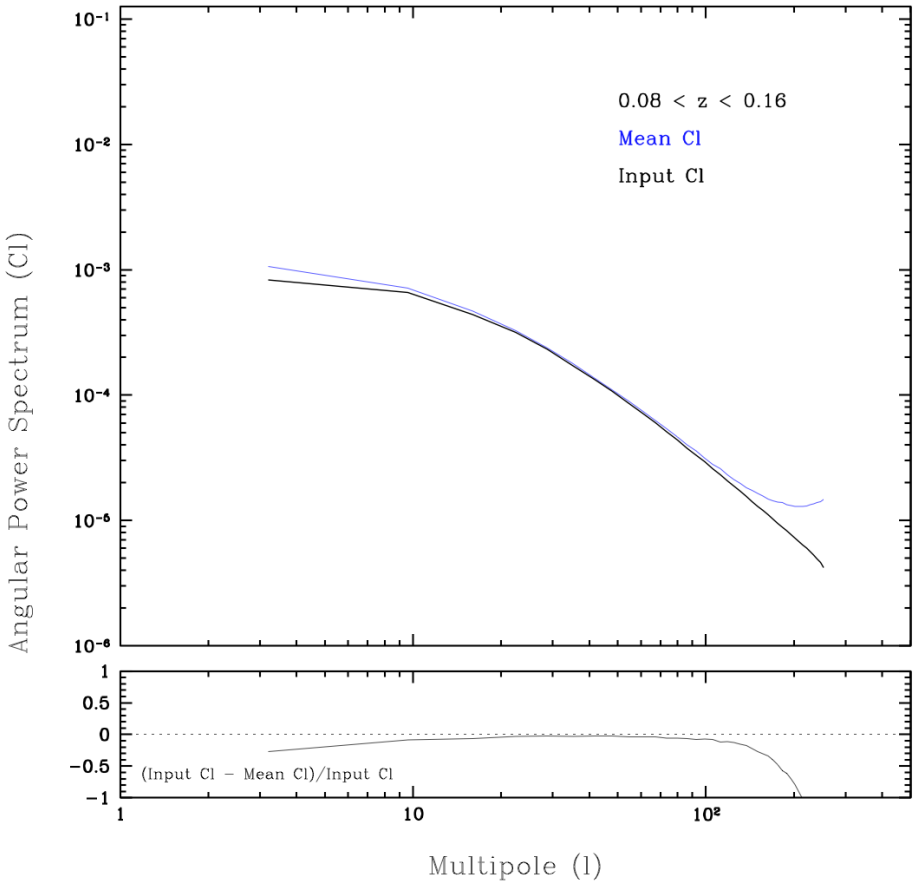

Figure 6.13: Same as **Fig. 6.11** but for the second redshift bin.

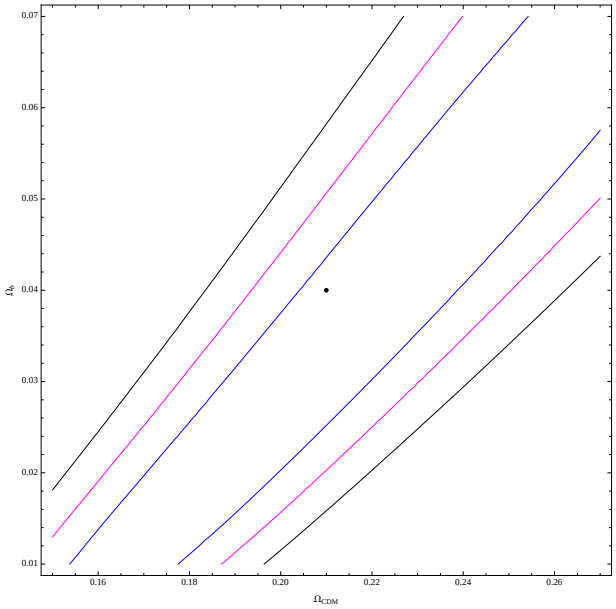

Figure 6.14: Same as **Fig. 6.10** but for the third redshift bin.

Second redshift bin, $0.08 < z < 0.16$

Fig. 6.17 is the analogous of **Fig. 6.16**, showing the results of the χ^2 obtained fixing Ω_b and varying Ω_{CDM} and A_s. The degeneracy between A_s and Ω_{CDM} is now more evident thanks to the fact that in this case the probability contours are centered around their true values ($\Omega_{CDM} = 0.21$ and $A_s = 2.46 \times 10^{-9}$, black dot). As for the Ω_b vs. Ω_{CDM} case, we find no systematic offset, the best fit values ($\Omega_M = 0.23$ and $A_s = 2.23 \times 10^{-9}$) are within the 1σ probability contour.

Third redshift bin, $0.16 < z < 0.24$

Fig. 6.18 shows the χ^2 analysis for the third bin when varying both Ω_{CDM} and A_s. This plot confirms the anti-correlation between the two parameters, and the fact that random errors are significantly larger in the third bin. Again, no systematic error is detected in this case.

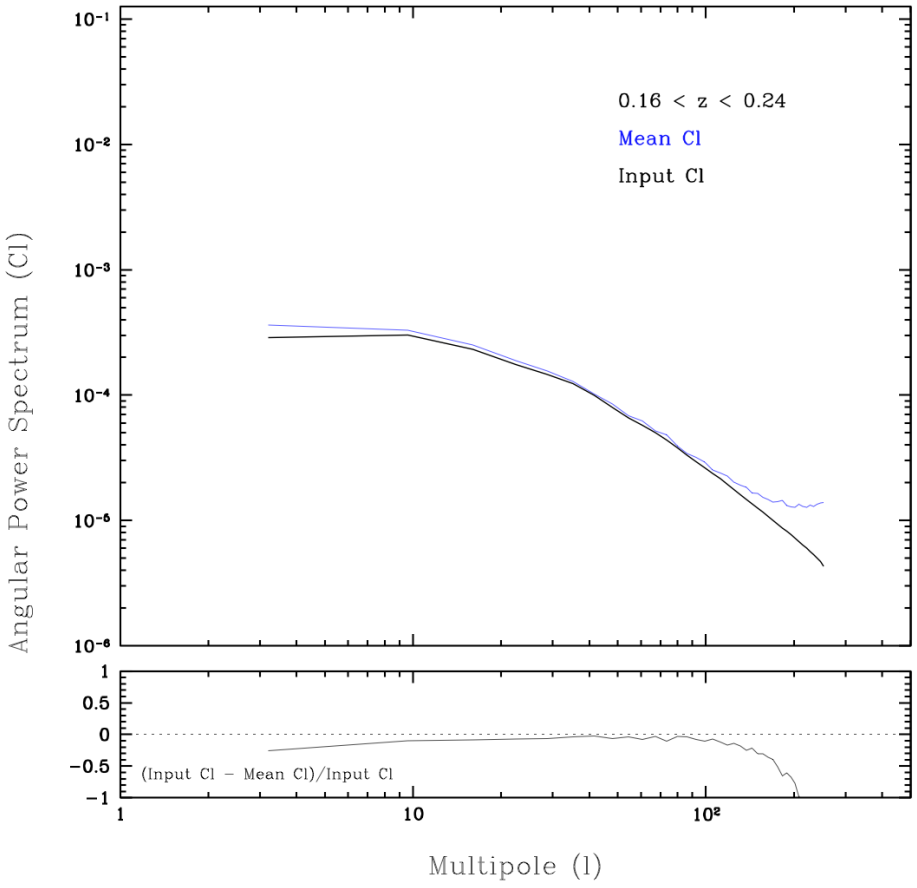

Figure 6.15: Same as **Fig. 6.11** but for the third redshift bin.

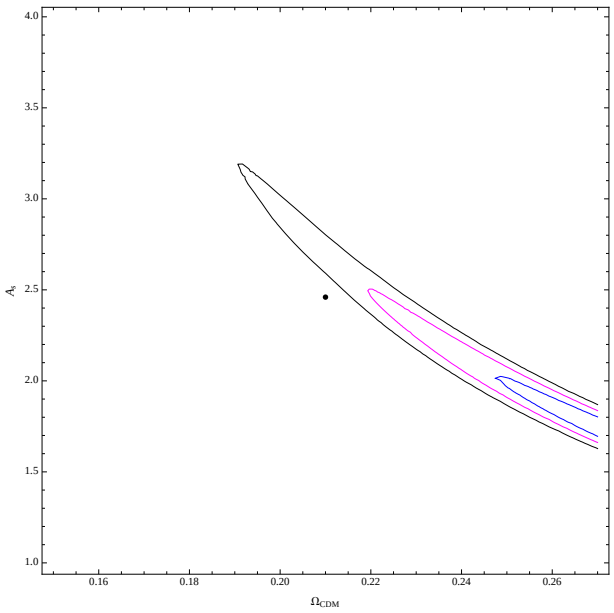

Figure 6.16: Same as **Fig. 6.10** but referring to the case of the spectral amplitude (A_s) and dark matter density (Ω_{CDM}) free parameters.

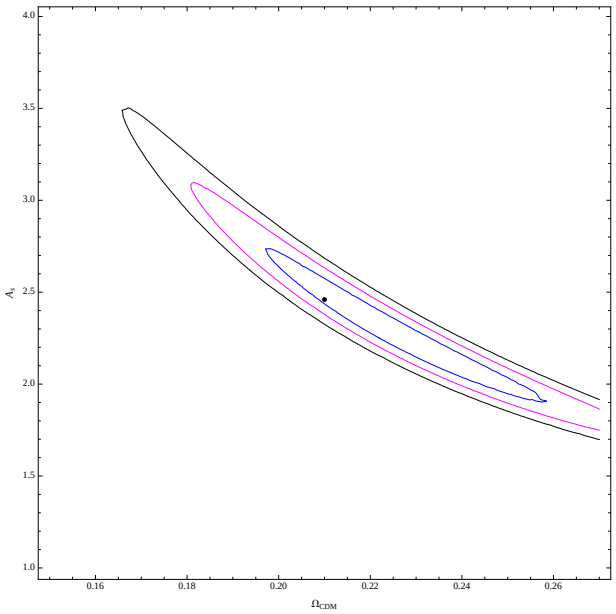

Figure 6.17: Same as **Fig. 6.16** but for the second redshift bin.

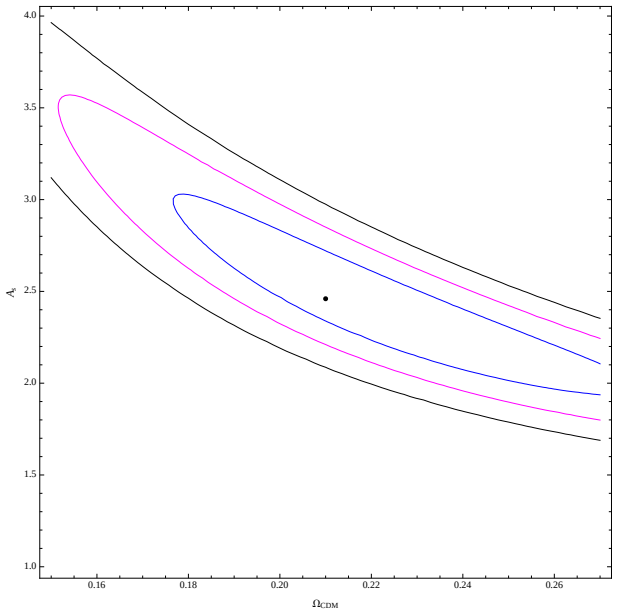

Figure 6.18: Same as **Fig. 6.16** but for the third redshift bin.

Chapter 7

2MPZ clustering analysis

All the machinery developed so far can now be applied to the real 2MPZ catalogue described in **Chapter 4**. In the following sections, we will first provide an estimate of the angular power spectrum of 2MPZ galaxies in the three redshift bins specified by **Eq. 5.15**. We then perform a `MontePython` analysis to extract cosmological parameters from the measured spectra. We shall focus here on four parameters, A_s, Ω_{CDM}, the total matter density Ω_M and the baryon fraction $f_b \equiv \Omega_b/\Omega_M$. First, we will estimate A_s and Ω_{CDM}, keeping fixed the baryon mass density and all other parameters to their Planck values (Planck Collaboration et al., 2014). Then, we will estimate the total mass density Ω_M and f_b fixing A_s to its best fit value. We will then investigate the dependence of the angular power spectrum on the luminosity of the objects by applying different magnitude cuts to the sample. We conclude with a discussion of the results.

7.1 The angular power spectrum of 2MPZ galaxies estimate

In **Fig. 7.1** we show the measured angular power spectrum in the three redshift bins chosen for our analysis: $0 < z < 0.08$, $0.08 < z < 0.16$ and $0.16 < z < 0.24$. As discussed in **Sec. 5.3.3**, the size of the bin is

a compromise that allows us to have large statistical sample with little evolutionary effects and limited impact of photo-z errors. All sample have the same geometry mask shown in **Fig. 3.8**. The three panels of **Fig. 7.1** show the angular spectra measured in the three redshift shells.

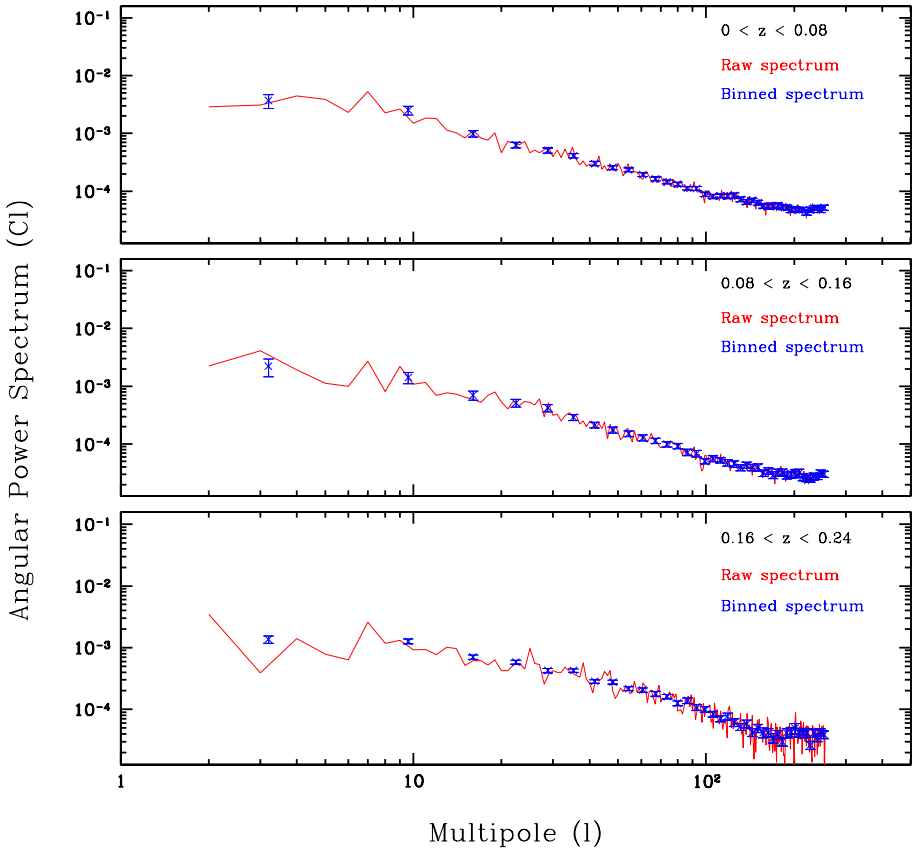

Figure 7.1: The angular power spectra measured from the 2MPZ catalogue (Bilicki et al., 2014) with our code in three redshift bins. Each panel shows the unbinned angular spectrum (red curve) whereas the binned one is represented by the asterisks together with the 1σ errorbars.

For each redshift shell, we measured the angular power spectrum in

the multipole range $[0, 256]$. The value of $\ell_{\rm max} = 256$ has been set by the angular resolution of the `HEALPix` map, $n_{side} = 64$, which corresponds to an angular scale of 0.0039°. In **Fig. 7.1** these spectra are shown with a continuous red line. The same figure also shows the binned angular spectra (blue asterisks), for which we have considered 40 linearly spaced bins from $\ell = 0$ to $\ell = 256$ (hence with ℓbin-size $\Delta_\ell = 6.4$). The errorbars are also shown, which have been obtained from the `rms` scatter of the mocks normalized to the amplitude of the 2MPZ spectrum.

For the comparison with the theoretical models we use the binned spectrum. This choice is dictated by our intention to compare the results presented in this thesis with the future MCMC analysis with the full covariance matrix. In fact, since the block covariance matrix turned out to be close to diagonal, here we shall present only the results of a MCMC analysis in which all diagonal terms have been ignored.

Not all the bins will be considered in the `MontePython` analysis. To minimize the impact of the geometry mask and of possible residuals observational biases, we decided to exclude multipoles below $\ell = 12$. In addition, we exclude all modes with $\ell > 100$ since it is beyond this multipole that the signature of the shot noise error becomes evident. Only for the third bin, where the shot-noise level become important even before $\ell = 100$, we decided to fix $\ell_{\rm max} = 70$. The choice of $\ell_{\rm min}$ and $\ell_{\rm max}$ is somewhat arbitrary but definitely conservative. We have checked the robustness of our results to the choice of the ℓ-range and found that changing $\ell_{\rm min}$ from 20 to 12 and ranging $\ell_{\rm max}$ from 70 to 200 has little impact on the results.

When compared to the mock spectra, the real ones have significantly more power on small scales. This is not surprising since mocks spectra were obtained from linear theory whereas in the real world nonlinear effects amplify the growth of density fluctuations on small scales. The effect is well illustrated by **Fig. 7.2** in which the binned 2MPZ angular spectrum

measured in the second z-bin (black starred symbols with the errorbars) is superimposed to a nonlinear ΛCDM model spectrum with cosmological parameters $\Omega_{CDM} = 0.19$ and $A_s = 4.71 \times 10^{-9}$, representing the best fit to the data (blue curve). The red curve shows the linear theory predictions. Both the linear and nonlinear models were obtained from `CLASS`. Both spectra include the effect of the mask, angular resolution on small scales (**Eq. 3.63**) and linear redshift space distortions on large scales. The amplitude of the linear spectrum is well below the nonlinear model, as expected, and the measurement for $\ell > 20$.

Back to **Fig. 7.1**, we only show the errobars for the binned spectra to avoid overcrowding the plot. These errors represent the diagonal element of the covariance matrix obtained from the mocks and normalized to match the amplitude of the measured spectra (i.e. multiplied by the ratio between the real and the mock spectra). The offset in the amplitude of the real and mock 2MPZ spectra is not independent on the angular scale, i.e. the bias is not exactly linear. To account for this, we normalize the amplitude of the errors to the ratio of the two spectra in each ℓ-bin. We therefore obtain an ℓ-dependent error normalization that we will also use to rescale the amplitude of the full covariance matrix.

Fig. 7.3 shows the measured binned spectra in the three bins and their errors. Amplitudes and shapes are quite similar with, however, some remarkable difference. The amplitude of the angular spectrum in the first bin is larger than that in the second bin. This is not unexpected. In fact, it mainly reflects the effect of photo-z errors, which depresses the spectral amplitude, and is larger for the second bin. This effect is not counterbalanced by the evolution of galaxy bias, which is expected to be small in this z-interval. The amplitude of the spectrum in the third bin matches that in the first bin. This behavior seems anomalous. The effect of photo-z errors should be similar to that in the second bin, so such a large amplitude would require a dramatic increase in galaxy bias. We postpone a more

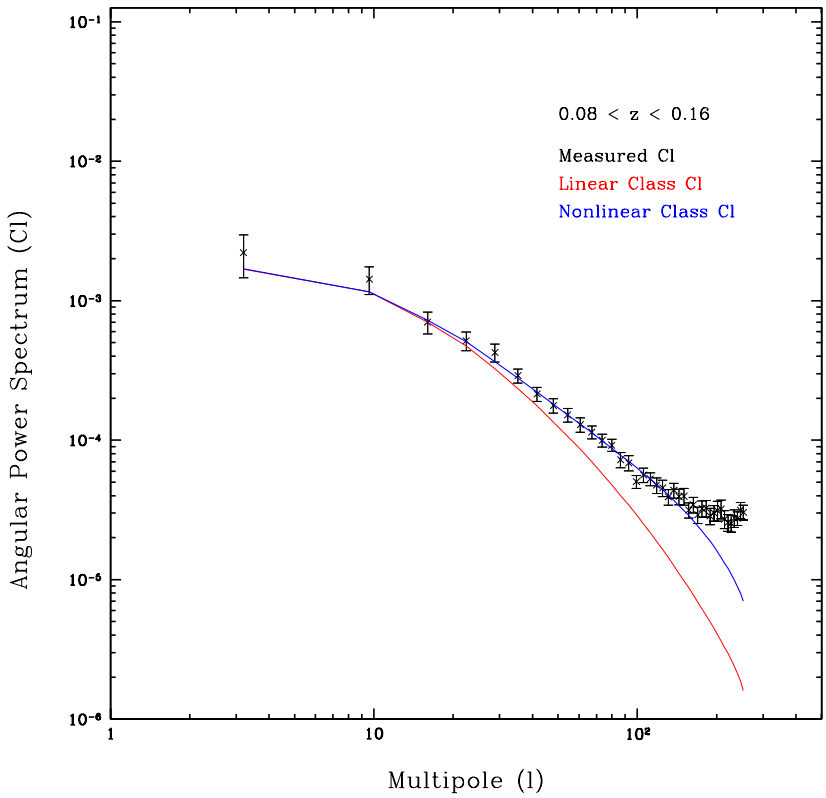

Figure 7.2: The binned angular power spectra measured from the 2MPZ catalogue in the second bin $0.08 < z < 0.16$ (black starred symbols with the errorbars) compared with two models obtained with `CLASS`: a linear model (red solid) and a non-linear one (blue solid) to which we have also added the redshift space distortions correction.

detailed and quantitative discussion on this issue to the next section.

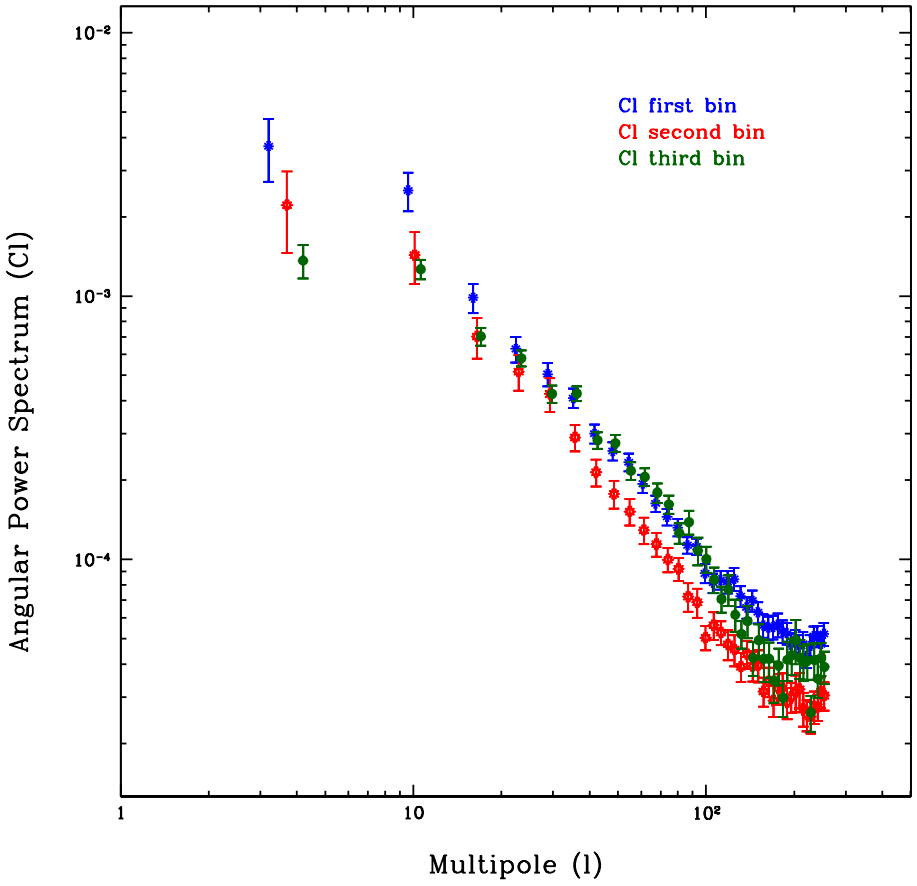

Figure 7.3: The binned angular power spectra measured from the 2MPZ catalogue in the three redshift bins and their 1σ errors. To avoid overcrowding the plot, for the second and the third redshift bin the ℓ values have been shifted by a value 0.5 and 1 respectively.

Another remarkable difference between the spectra is the size of the error bars, that are larger in the third bin. This is due to the comparatively smaller number of objects at high redshift in the galaxy catalogue.

7.2 Ω_{CDM} and A_s from the MCMC analysis

In this section we repeat the first cosmological analysis performed on the mock spectra using, however, real data. Here we focus on 2 parameters, the dark matter density and the spectral amplitude, keeping all the others fixed to the Planck cosmology values (Planck Collaboration et al., 2014). The main goal here is to check whether the measured values are in the right ballpark. Systematic deviations would reveal possible issues with the observational data and their treatment since the presence of obvious systematics in both the estimator and the χ^2 analysis have been excluded by the tests with the mocks. We are also interested in assessing the amplitude of the random errors to evaluate the precision that one can achieve by the clustering analyses using other photo-z catalogues that will become available in the future.

In the MCMC analysis we compare our measurements with the theoretical prediction of the angular power spectrum of the matter in a flat ΛCDM model obtained from `CLASS` for the case of a nonlinear evolution of density perturbation simulated using `Halofit` (Smith et al., 2003). Since the angular power spectrum is the projection of the 3D one over the celestial sphere, we need to specify the range of redshift considered in the analysis and the statistical weight attached to each redshift. In the framework of `CLASS` modeling, this is done by specifying the redshift range and the redshift distribution of the mass tracers considered. For our analysis we then need to specify the 2MPZ galaxy dN/dz within each one of the three redshift bins that we have considered. In particular, we will use the spectroscopic redshift distributions corresponding to the three photometric redshift bins discussed in **Sec. 5.3.2** and shown in **Fig. 5.10**.

Moreover, since our analysis is performed in redshift space, coherent motions on large scale amplify the power at small multipoles. An effect which, in the linear regime, is known as Kaiser effect (Kaiser, 1984). This effect is fully accounted for in the `CLASS` model once the value of the mass parameter is specified.

Finally, as anticipated in **Chapter 3**, we multiply the theoretical spectrum by the pixel window function expressed by **Eq. 3.63** to mimic the effect of the pixelized map.

7.2.1 First redshift bin, $0 < z < 0.08$

Fig. 7.4 shows the result of the `MontePython` analysis on the first z-bin, from 0 to 0.08. The prior range is set to $[0, 0.7]$ for Ω_{CDM} and $[0.1, 10]$ for A_s. We run 50 chains of 10000 steps each. The figure shows the marginalized probability (bottom right and upper panels) for Ω_{CDM} and A_s as well as their joint probability (bottom left panel).

For the marginalized probability the best fit values for the two parameters are:

$$\Omega_{CDM,\,\text{best fit}} = 0.34 \tag{7.1}$$

$$A_{s,\,\text{best fit}} = 1.78 \times 10^{-9} \tag{7.2}$$

Given the skewness of the distribution, we report the minimum credible intervals for the two parameters in **Table 7.1**. The results are consistent,

Table 7.1: The minimum credible intervals for the dark matter density parameter and the primordial amplitude in the first redshift bin.

	68.26% probability	95.4% probability
Ω_{CDM}	$[0.19, 0.40]$	$[0.14, 0.50]$
$A_s \times 10^9$	$[1.13, 2.62]$	$[0.92, 4.42]$

within the 68% probability interval, with the Planck values (see **Table 1.1**). By comparing the spectral amplitude with the Planck one (2.14×10^{-9} with negligible errors compared to those of this analysis), we can infer the linear bias of 2MPZ galaxies:

$$b_{\text{first z}-\text{bin}} = \frac{A_{s,\,\text{2MPZ}}}{A_{s,\,\text{Planck}}} = [0.52, 1.05] \tag{7.3}$$

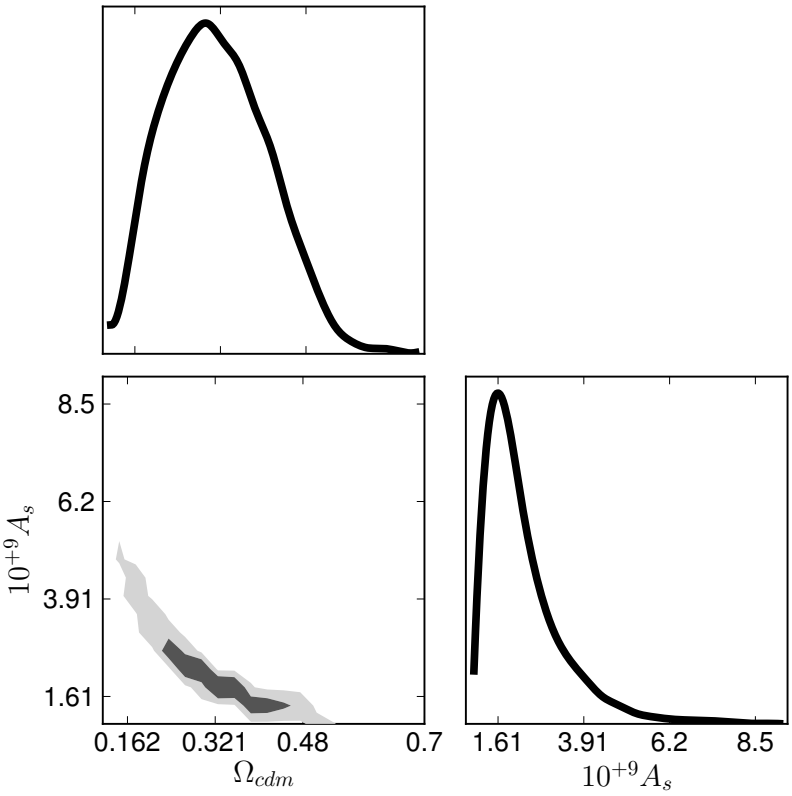

Figure 7.4: Bottom right panel and upper panel: the marginalized posterior probability obtained with `MontePython` for the first redshift bin, $0 < z < 0.08$ with Ω_{CDM} and A_s as free parameters. Bottom left panel: the two dimensional posterior distribution. See text for details.

in the 68% probability interval, i.e. consistent with the hypothesis that 2MPZ galaxies are unbiased tracers and with the bias of 2MRS galaxies ((Westover, 2007) and (Corsi, 2015)). As for Ω_{CDM}, the Planck value of 0.263 for $h = 0.67$ (fixed in our analysis) is well within the 68% probability interval.

The corresponding best fit spectrum is shown in the upper panel of **Fig. 7.5** with a black solid line, together with the measured spectrum and its errorbars only shown in the range of the analysis, namely $\ell = [12, 100]$. The bottom panel shows the percent difference between data and model. In most ℓ-bins the discrepancy is within the 1σ errorbars, confirming that the model is a good fit to the data.

7.2.2 Second redshift bin, $0.08 < z < 0.16$

We then repeat the same analysis for the spectrum measured in the second bin, i.e. for $0.08 < z < 0.16$. The result is shown in **Fig. 7.6**, where the marginalized probability (bottom right and upper panels) for Ω_{CDM} and A_s together with their joint probability (bottom left panel) are displayed. In this case the best fit values for the two parameters are:

$$\Omega_{CDM,\text{ best fit}} = 0.19 \tag{7.4}$$

$$A_{s,\text{ best fit}} = 4.71 \times 10^{-9} \tag{7.5}$$

while the minimum credible intervals are shown in **Table 7.2**.

Table 7.2: The minimum credible intervals for the dark matter density parameter and the primordial amplitude in the second redshift bin.

	68.26% probability	95.4% probability
Ω_{CDM}	$[0.14, 0.24]$	$[0.11, 0.29]$
$A_s \times 10^9$	$[3.04, 6.09]$	$[2.55, 8.71]$

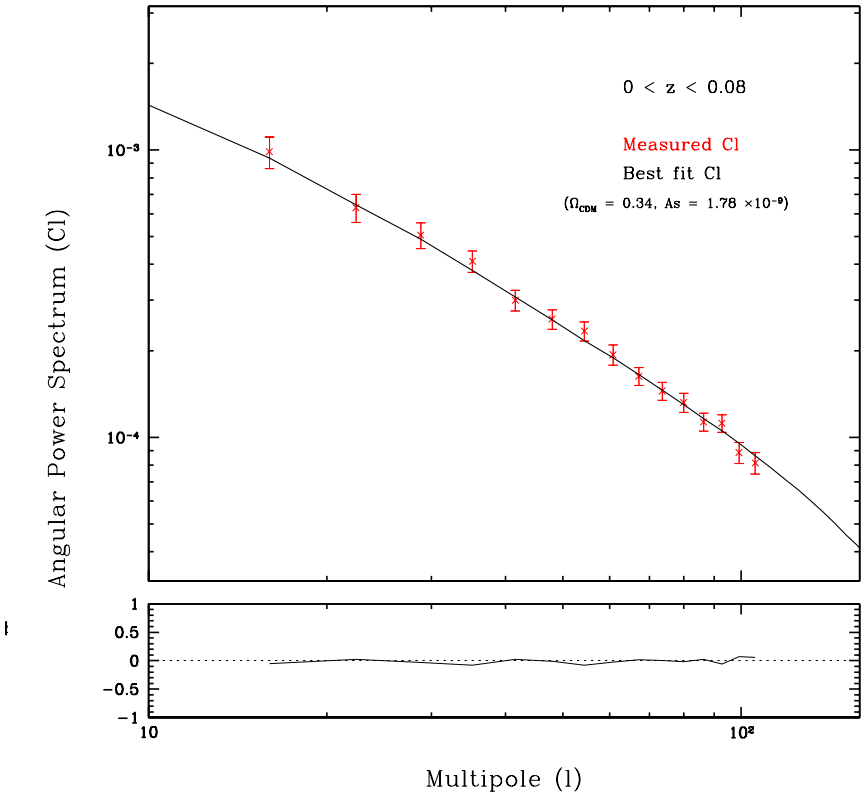

Figure 7.5: Upper panel: the angular power spectrum measured from the 2MPZ catalogue in the first bin $0 < z < 0.08$ (red asterisks and errorbars) in the range $\ell = [12, 100]$ compared with the best fit model obtained from `CLASS` with $\Omega_M = 0.34$ and $A_s = 1.78 \times 10^{-9}$ with all other parameters fixed to the Planck values. Bottom panel: the relative difference between the two spectra.

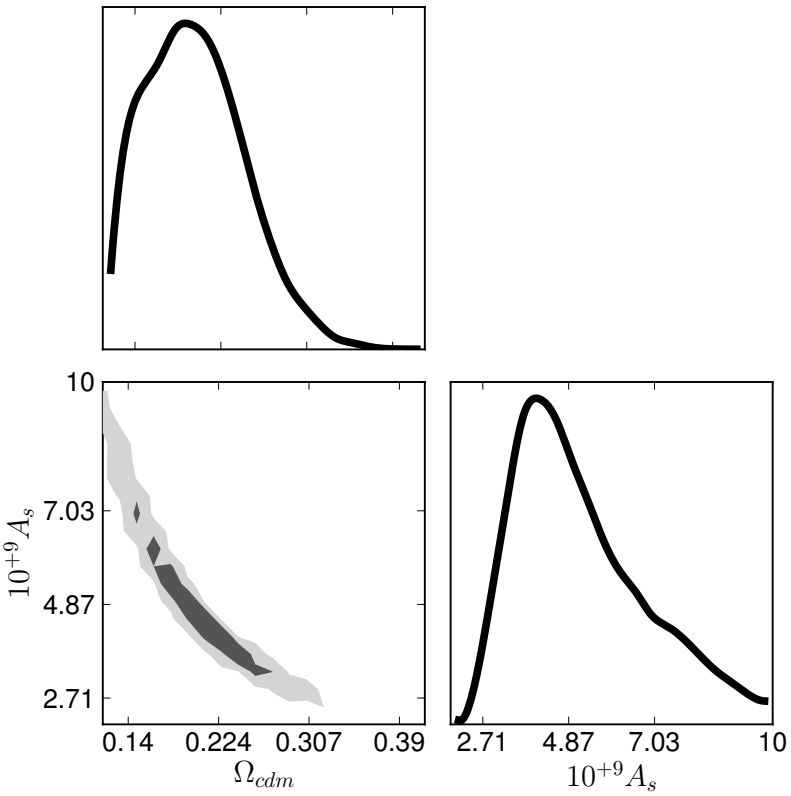

Figure 7.6: Same as **Fig. 7.4** but for the second redshift bin.

Also for this z-bin we can infer the linear bias of 2MPZ galaxies by comparing the spectral amplitude with the Planck one (2.14×10^{-9}). We obtain:

$$b_{\text{second z}-\text{bin}} = \frac{A_{s,\,\text{2MPZ}}}{A_{s,\,\text{Planck}}} = [1.42, 2.84] \tag{7.6}$$

in the 68% probability interval, i.e. consistent with the fact that, because of the flux cut, galaxies in the interval are more luminous and more biased than in the first one. The larger bias shifts upward the likelihood contours in the A_s - Ω_{CDM} plane. We notice that, apart form the offset, the contours shape and, therefore, the degeneracy between the two parameters, are almost identical in the first two bins. The similarity of the contours is reassuring since it excludes the possibility of redshift-dependent systematic errors.

Fig. 7.7, analogous to **Fig. 7.5**, shows in the upper panel the comparison between the best fit spectrum and the spectrum measured from the 2MPZ catalogue in the second bin. Also in this case the model provides a good fit to the data with relative differences always smaller than 20% in the ℓ-range considered.

7.2.3 Third redshift bin, $0.16 < z < 0.24$

The probability contours obtained from the `MontePython` analysis of the third bin, i.e. the one with $0.16 < z < 0.24$ are shown in **Fig. 7.8**. They do show the same (A_s, Ω_{CDM}) degeneracy and span a similar range in Ω_{CDM} with respect to the previous z-bins. However, the spectral amplitude is much larger than in the other bins and inconsistent with any reasonable biasing scheme and redshift-dependence. Indeed the best fit values are $\Omega_{CDM} = 0.26$ and $A_s = 6.74 \times 10^{-9}$ with minimum credible intervals shown in **Table 7.3**.

This implies:

$$b_{\text{third z}-\text{bin}} = \frac{A_{s,\,\text{2MPZ}}}{A_{s,\,\text{Planck}}} = [2.88, 3.38] \tag{7.7}$$

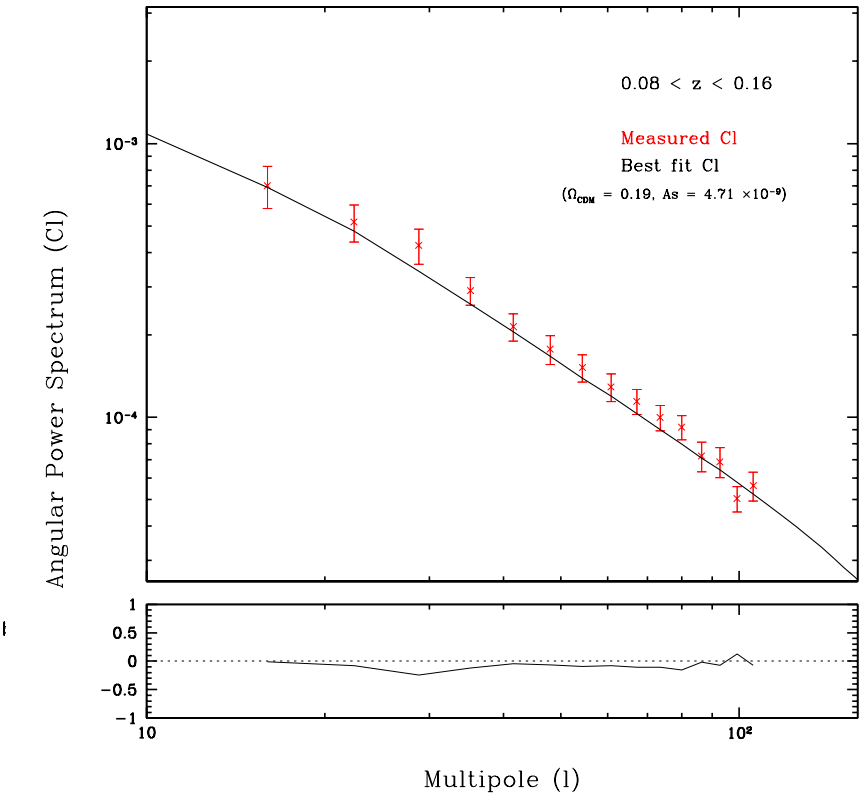

Figure 7.7: Same as **Fig. 7.5** but for the second redshift bin.

Table 7.3: The minimum credible intervals for the dark matter density parameter and the primordial amplitude in the third redshift bin.

	68.26% probability	95.4% probability
Ω_{CDM}	$[0.23, 0.28]$	$[0.21, 0.3]$
$A_s \times 10^9$	$[6.18, 7.25]$	$[5.74, 7.94]$

These bias values are very large indeed and seems to point to the presence of some systematics in the data rather than to a genuine highly biased population.

The quality of the fit is also less satisfactory with theoretical predictions that are systematically below data at small ℓs (**Fig. 7.9**). The reason for this discrepant result is not clear. One possibility, that we shall explore in **Sec. 7.4**, is some observational bias related to bright objects that preferentially populate this redshift bin.

7.3 Ω_M and baryon fraction f_b from the MCMC analysis

In this section we perform a cosmological analysis analogous to the one seen in the previous section but for the total matter density parameter, Ω_M, and the baryon fraction, $f_b \equiv \Omega_b/\Omega_M$. The value of the spectral amplitudes in the three bins are set equal to their best fit values estimated in the previous section. All other parameters are fixed to their Planck values (Planck Collaboration et al., 2014). Previous studies have used angular spectra to estimate the baryon fraction (Blake et al. (2007) and Thomas et al. (2011)). The rationale for focusing on this parameter rather than the baryon density itself is that baryon oscillations, whose absolute amplitude depend on Ω_b, are very weak in the angular spectra. To detect baryonic features one has to subtract off the continuous part of the spec-

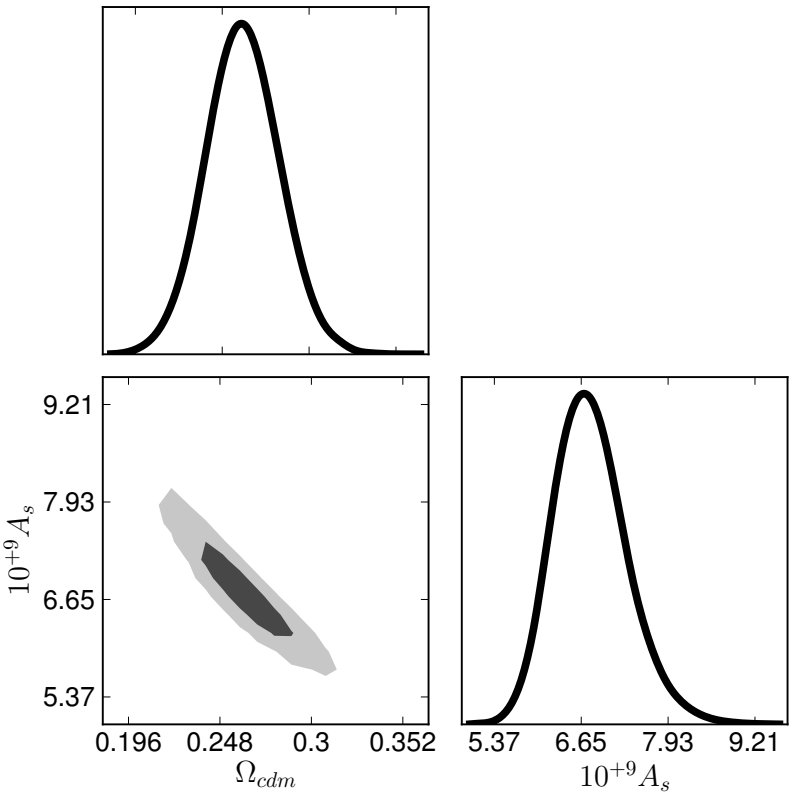

Figure 7.8: Same as **Fig. 7.4** but for the third redshift bin

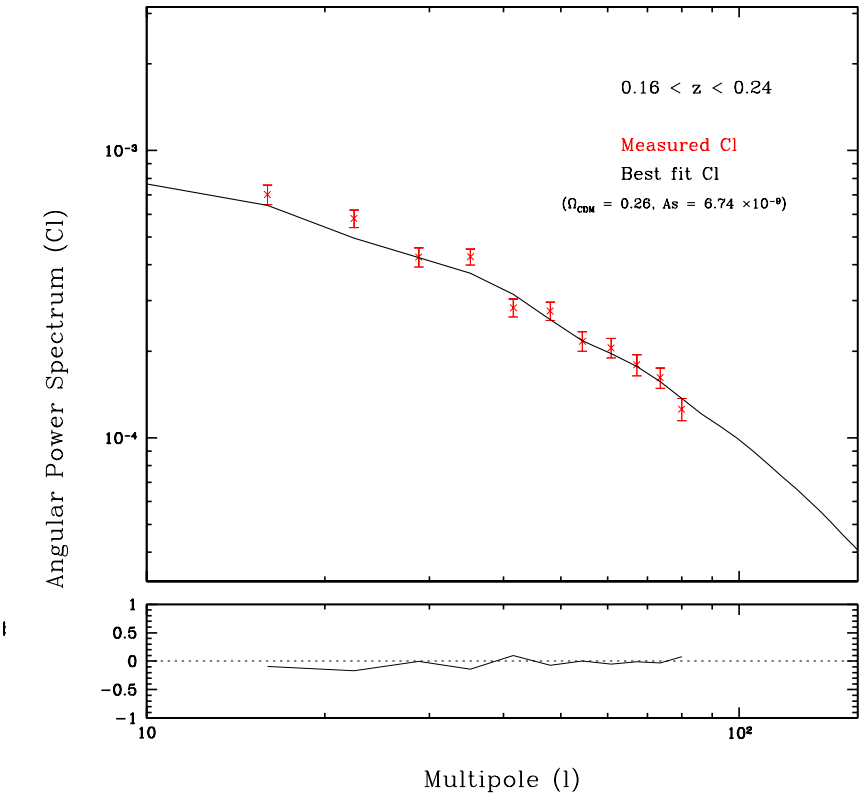

Figure 7.9: Same as **Fig. 7.5** but for the third redshift bin.

trum, that depends on Ω_M. Hence, the better sensitivity to the baryon fraction Ω_b/Ω_M.

As for the previous analysis, we compare our measurements with the theoretical power spectrum of matter in a flat ΛCDM obtained from `CLASS` using `Halofit` (Smith et al., 2003). We consider the same redshift bins as in **Sec. 7.2** and the corresponding spectroscopic redshift distributions shown in **Fig. 5.9**. Theoretical models include the effect of redshift distortions, pixel and geometry window functions.

7.3.1 First redshift bin, $0 < z < 0.08$

The result of the `MontePython` analysis for the first redshift bin, ranging from 0 to 0.08, is shown in **Fig. 7.10**. The spectral amplitude has been fixed to its best fit value for this bin, i.e. $A_s = 1.78 \times 10^{-9}$. We ran 100 chains of 10000 steps each. The figure shows the joint probability for Ω_M and f_b.

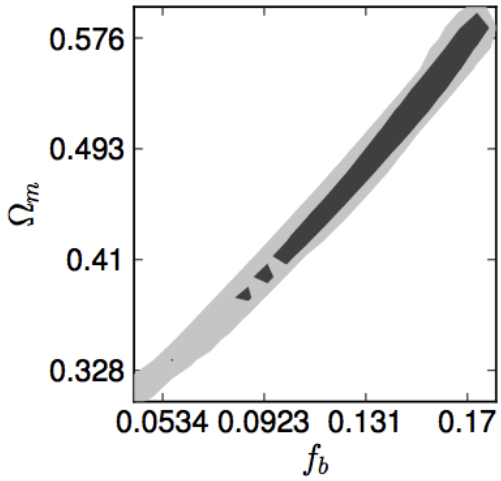

Figure 7.10: The joint probability for the total matter density parameter Ω_M and the baryon fraction f_b. The spectral amplitude has been fixed to its best fit value, $A_s = 1.78 \times 10^{-9}$.

The best fit parameters in this case are:

$$\Omega_{M,\text{ best fit}} = 0.50 \tag{7.8}$$

$$f_{b,\text{ best fit}} = 0.14 \tag{7.9}$$

The minimum credible intervals for the two parameters in **Table 7.4**.

Table 7.4: The minimum credible intervals for the total matter density parameter and the baryon fraction in the first redshift bin.

	68.26% probability	95.4% probability
Ω_M	$[0.42, 0.59]$	$[0.33, 0.59]$
f_b	$[0.10, 0.17]$	$[0.05, 0.18]$

The baryon fraction is fully consistent with the Planck value ($f_b = 0.15$ with negligible errors compared to those of this analysis) within the 68% probability interval. It is also interesting to compare our results with those of Blake et al. (2007) and Thomas et al. (2011). These analyses are based on photometric redshifts of the SDSS luminous red galaxies. Blake et al. (2007) considered a sample of about 600000 objects whereas Thomas et al. (2011) used an updated data release. The redshift range considered ($0.4 < z < 0.7$) does not overlap with that of our more local sample. However, the f_b value is not expected to change with the redshift. Indeed, their baryon fraction estimates (0.16 ± 0.036 and 0.173 ± 0.046 for Blake et al. (2007) and Thomas et al. (2011), respectively; 1-σ errors quoted) is entirely consistent with our result. Note that the values quoted were obtained by combining the results of different redshift slices.

7.3.2 Second redshift bin, $0.08 < z < 0.16$

The result of the same analysis for the second redshift bin, from 0.08 to 0.16, is shown in **Fig. 7.11**, where the joint probability for Ω_M and f_b is displayed. The best fit of the spectral amplitude for this bin is $A_s = 4.71 \times 10^{-9}$. Again, we ran 100 chains of 10000 steps each.

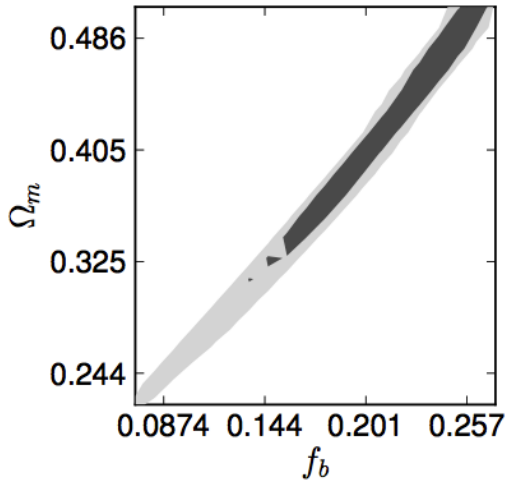

Figure 7.11: Same as **Fig. 7.10** but for the second redshift bin, $0.08 < z < 0.16$. The spectral amplitude has been fixed to its best ft value for this bin, i.e. $A_s = 4.71 \times 10^{-9}$.

The best fit parameters in this case are:

$$\Omega_{M,\text{ best fit}} = 0.41 \tag{7.10}$$

$$f_{b,\text{ best fit}} = 0.20 \tag{7.11}$$

with the minimum credible intervals shown in **Table 7.5**. We notice that

Table 7.5: The minimum credible intervals for the total matter density parameter and the baryon fraction in the second redshift bin.

	68.26% probability	95.4% probability
Ω_M	$[0.36, 0.51]$	$[0.25, 0.52]$
f_b	$[0.16, 0.26]$	$[0.09, 0.27]$

in this second bin the value of f_b is larger but still consistent, within the 68% probability interval, with that of the first redshift bin. The uncertainty interval is larger too, although the relative error is also similar to that of the first bin. Overall, the estimates of the baryon fraction in the two bins are consistent with each other.

7.3.3 Third redshift bin, $0.16 < z < 0.24$

Fig. 7.12 is the same as **Fig. 7.10** and **Fig. 7.11** but for the third redshift bin, from 0.16 to 0.24. The best fit of the spectral amplitude for this bin is $A_s = 6.74 \times 10^{-9}$. Again, we ran 100 chains of 10000 steps each.

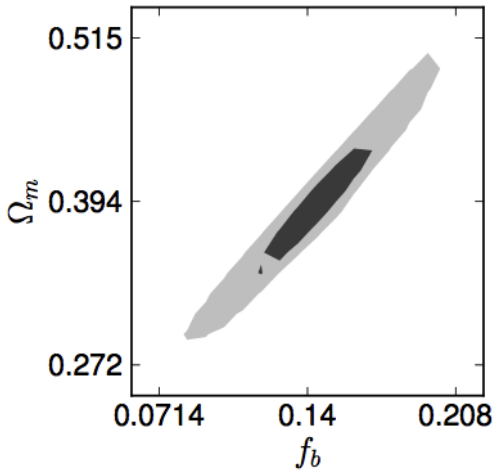

Figure 7.12: Same as **Fig. 7.10** but for the second redshift bin, $0.16 < z < 0.24$. The spectral amplitude has been fixed to its best ft value for this bin, i.e. $A_s = 6.74 \times 10^{-9}$.

The best fit parameters in this case are:

$$\Omega_{M,\text{ best fit}} = 0.38 \tag{7.12}$$

$$f_{b,\text{ best fit}} = 0.14 \tag{7.13}$$

with the minimum credible intervals shown in **Table 7.5**.

Also in the third redshift bin the estimated baryon fraction is consistent with the results of the first two bins. This is quite remarkable, since we took the large spectral amplitude measured in the third bin as a potential clue of systematic error. This result shows that our baryon fraction estimate is robust to this systematic effect, assuming that indeed there is one.

Table 7.6: The minimum credible intervals for the total matter density parameter and the baryon fraction in the third redshift bin.

	68.26% probability	95.4% probability
Ω_M	$[0.35, 0.43]$	$[0.30, 0.49]$
f_b	$[0.12, 0.17]$	$[0.09, 0.19]$

The consistency among our f_b estimates in the three bins implies that our results are also consistent with those of Thomas et al. (2011) and Blake et al. (2007) and Planck Collaboration et al. (2014).

7.4 Luminosity dependence

The large spectra amplitude measured in the third bin that we have measured in **Sec. 7.2** can have different origins, one of which is the presence of some unidentified systematic errors in the data themselves. In **Sec. 7.3**, however, no anomalous behavior has been detected in the third bin, i.e. the values of Ω_M and f_b measured in the three redshift bins are consistent with each other. Therefore, before looking for systematic errors in the data treatment, we need to consider the hypothesis that the large A_s value in the third bin is a genuine rather than an anomalous feature.

The 2MPZ galaxies have been selected by their flux. As a result, objects in the third bins are typically brighter than in the first one. Since luminous objects are more biased, we want to check whether the large A_s simply reflects the effect of luminosity bias.

Corsi (2015) has determined the luminosity dependence of the linear bias of 2MRS galaxies, a sample that significantly overlaps with that of 2MPZ galaxies in the first bin. The result of that analysis was that the bias increases by about 40% over 1 decade in luminosity. In terms of A_s this would imply a factor of 2 increases from faint to bright objects in the first bin.

In order to investigate the issue further and check whether the spectral amplitude in the third redshift bin is indeed anomalous or it is in line with expectations, we have applied different flux cuts and measured the angular power spectra of the corresponding subsamples. Starting from the $m_K = 13.9$ magnitude cut of the whole 2MPZ sample, applied to guaranteed a nominal 95% completeness (Bilicki et al., 2014), we considered 7 cuts up to $m_K = 11.75$. Extracting subsamples with progressively brighter flux cuts increases the completeness, at the price of selecting a progressively smaller number of increasingly biased tracers. **Fig. 7.13** shows the result of this test in which the magnitude cut was increased from $m_k = 13.9$ to $m_k = 11.75$. The figure shows the angular spectra computed for different magnitude limits (panels in rows) in the full redshift range (first column) and in the three redshift bins (second, third and forth column). Note that in this case the logarithmic scale is only used for the y-axis. The black dots in each panel indicates the reference spectrum computed for the full redshift range and $m_{\rm K} \leq 13.9$. Over-imposed to it, the colored symbols with errorbars represent the angular spectrum of the individual subsample. The size of the errorbars indicate the Gaussian errors. We use the Gaussian approximation since our mock catalogues, which are not designed to reproduce the clustering properties as a function of luminosity, cannot be used to estimate the errors directly.

Looking at the plots in the first column, we clearly see that the amplitude of the spectrum increases with the luminosity and that the effect is almost independent from ℓ, especially at large angular scales. This is to be expected since more luminous objects preferentially form in the more massive halos and, therefore, are more biased with respect to mass. The systematic up-shift of the spectrum with the luminosity is therefore the typical, expected, signature of a luminosity dependence linear bias.

Let us now focus on the individual z-bins (the remaining columns) and, more specifically, on the $m_K = 13.9$ cut for the third bin. To account for a luminosity dependence bias, a comparison with the angular spectrum of

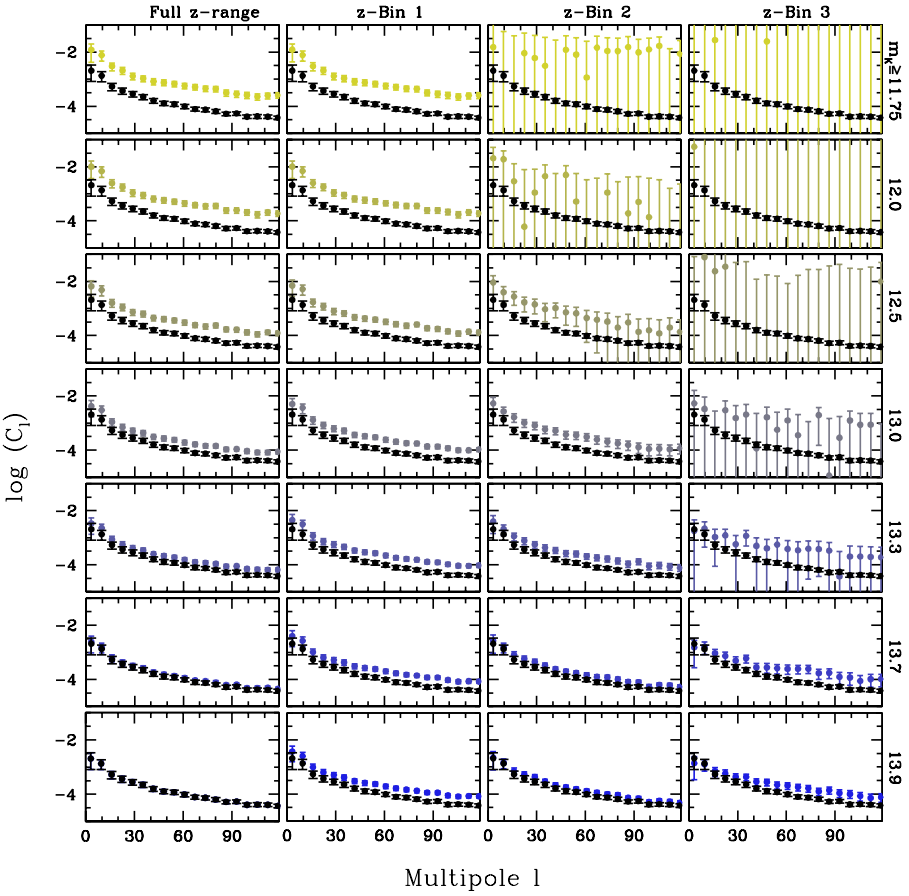

Figure 7.13: The angular power spectra measured from different subsamples of the 2MPZ catalogue with different magnitudes cuts. Each column show the spectra computed in different redshift ranges, each raw the spectra obtained with different magnitude cuts. See text for details.

a different redshift bins should be performed at the same luminosity cut. For this we need to derive the corresponding flux cuts for the other two bins.

We first convert apparent to absolute magnitudes:

$$M_{K_s} = m_{K_s} - 25 - 5\log(d_L) - k(z) - e(z) \tag{7.14}$$

where d_L is the luminosity distance. Since the sample is local, we use the approximation $d_L = c\ z/100$ and express distances in Mpc$/h$. Following Corsi (2015), we use the k-correction of Kochanek et al. (2001) and Branchini et al. (2012):

$$k(z) = -6\log(1+z) \tag{7.15}$$

and the evolutionary correction $e(z) = 3.04z$. We estimate **Eq. 7.14** at the mean galaxy redshift within each z-bin: $\langle z \rangle = 0.056, 0.109, 0.187$ respectively. The corresponding magnitude cuts turned out to be:

$$m_K = 11.198,\ 12.67,\ 13.9 \tag{7.16}$$

Since we did not perform these exact magnitude cuts, we considered the cases $m_K = 11.75$ for the first bin, $m_K = 12.5$ for the second and, obviously, $m_K = 13.9$ for the third.

We are now in the position of comparing the spectral amplitudes. To do that we simply consider the value of the angular spectrum at $\ell = 28$. This multipole is large enough to exclude mask effect and small enough to minimize the effect of nonlinearities. The ratio of the corresponding spectral amplitudes for the first, second and third bin (A_1, A_2 and A_3, respectively) are:

$$\frac{A_2}{A_1} = 1.3 \pm 0.3 \quad \frac{A_3}{A_1} = 0.35 \pm 0.06 \quad \frac{A_3}{A_2} = 0.28 \pm 0.07 \tag{7.17}$$

Errors have been obtained by propagating 1-σ spectral uncertainties. These ratios are of the order of unity and there is no evidence for an excess power in the spectrum of the third bin. On the contrary, its amplitude is slightly smaller than at lower redshifts. In other words, we find

no evidence for anomalous power in the third bin once the comparison is performed at a similar luminosity cut.

Back to the first bin, let us consider the spectral amplitudes corresponding to the fainter and brighter magnitude cuts. We find:

$$\frac{A_{1,\,11.75}}{A_{1,\,13.9}} = 2.5 \pm 0.4 \tag{7.18}$$

Considering that the corresponding luminosity range (1 decade) is similar to that analyzed by Corsi (2015), this ratio is fully consistent with a factor of 2 increases of bias with luminosity. Also in this case, we find no anomalous behavior of the angular spectrum as a function of luminosity.

We therefore conclude that the large A_s value obtained in **Sec. 7.2** is consistent with the expected luminosity dependent bias and does not hint at any luminosity-dependence systematic error in the data selection.

Chapter 8

Conclusions

In the recent years, the Large Scale Structure of the Universe, i.e. the spatial distribution of matter on cosmological scales, has proved to be one of the most powerful probes to investigate the proprieties of the Universe, and in particular to unveil its dark components: dark matter and dark energy. As a consequence, an increasing number of galaxy surveys has mapped the three dimensional distribution of galaxies. More ambitious surveys are planned, the most ambitious of which, the Euclid survey, will probe about the 2% of the observable Universe. The goal is to estimate cosmological parameters at the % accuracy, an important pre-requisite if one wants to discriminate between different cosmological models.

Beside the spectroscopic redshift surveys, where the redshift is measured through spectroscopy, photometric redshift surveys offer the possibility to probe larger volumes of the Universe in which the spatial distribution of the matter is more densely sampled with less precision but with better statistics, using objects with a much wider range of physical properties, like their intrinsic luminosity.

Because of the rather poor resolution along the radial direction, an effective way of studying the clustering properties and their evolution over cosmic time in photometric surveys is that of considering the angular correlation properties of the objects in different redshift shells to sample the

time axis in a discrete and yet not a too discontinuous way.

In this thesis we have applied this tomographic approach to estimate the angular power spectrum of galaxies in the recently released 2MPZ catalog. This dataset contains about 935000 galaxies distributed over 80% of the sky.

The goal was twofold. First of all, to assess the possibility of using this dataset for cosmological investigations and, second of all, use it to obtain an independent estimate of cosmological parameters in the local Universe. The hope is that this study will provide the expertise and the guidelines to perform similar analyses to future wide photometric-redshift catalogs, containing many more objects and probing much deeper regions, like the one that will be obtained by the photometric survey of the Euclid satellite.

With these general goals as guidelines, we have addressed the specific issues detailed below.

- We have developed our own code to estimate the angular spectrum of a distribution of discrete objects over some fraction of the 2D-sphere. The code, written in C++ language, implements the Peebles (1973) unbiased estimator that automatically corrects for incomplete sky coverage. The code interpolates a discrete distribution over a 2D pixelized map of surface density of objects and then computes its angular spectrum. To do so, the code implements a number of `HEALPix` routines, which makes the code itself highly portable.

  The same code is also used to compute the mixing matrix that quantifies the multipole mixing induced by the partial sky coverage. In this step we have used the approximated expression for the mixing matrix, proposed by Peebles (1973), which is fully consistent with the angular spectrum estimator.

  The code was thouroughly tested. Firstly, we have cross-compared outputs with an analogous code developed by Sreekumar Balan (private communication) for the Weak Lensing Science working Group in the Euclid Consortium. Then, we applied the code to distribution

of objects with a well known power spectrum to look for systematic errors and found none on scales larger than the angular resolution corresponding to the pixel size. This is a rather demanding tests in case of partial sky coverage since, in that case, we also test the goodness of the estimated mixing matrix and that of the convolution of the reference spectrum.

- Having excluded systematic errors in the estimator by means of somewhat ideal tests, we turned to the problem of assessing the presence and the amplitude of both systematic and random errors of the angular spectrum estimated for a more realistic galaxy catalogue matching the characteristics of 2MPZ.

  We did this by means of simplified mock galaxy catalogs in which the realistic touch consisted in mimicking a sample with the same dN/dz and sky coverage as the real one.

  To compromise between the competing needs of generating many independent samples containing about one million of objects each and that of limiting the computational time to generate them, we decided to rely on the lognormal model to generate samples of objects with a realistic 1-point PDF and a well known angular spectrum.

  The code has been tested to check that the lognormal transformation does not modify the 2-point statistics of the starting Gaussian field. This delicate step has been calibrated in collaboration with Henrique Xavier, author of the FLASK toolkit (Xavier et al., 2016), a set of codes that generates similar lognormal mocks targeted at weak lensing studies.

  In the end we generated 144 independent lognormal mocks each one consisting of three subsamples of objects in three redshift slices ranging from $z = 0$ to $z = 0.24$, with a dN/dz consistent, within Poisson errors, with the 2MPZ one. Each object of the catalogue is characterized by two angular coordinates, its cosmological redshift and its

photometric redshift.

These mocks were used to compute a 12×12 covariance matrix of the binned angular spectrum. The covariance matrix turns out to be close to diagonal, for which reason we have ignored the off-diagonal terms.

The mock analysis also allowed us to characterize and quantify the systematic errors induced by the geometry mask. The effect of the mask is to removing power from large scale. The effect is not perfectly accounted for by convolving the model with the estimated mixing matrix. The mismatch is more evident for small multipoles. For this reason we decided to exclude the multipole range $\ell < 12$ from the analysis.

- After having quantified systematic and random errors, we have performed two different analyses to estimate some cosmological parameters from the measured angular spectrum.

  The first one is a standard χ^2 statistics. The second one is based on the likelihood statistics that we efficiently sample with a MCMC technique using the `MontePython` (MP) code. As anticipated, in both analyses we only consider the diagonal terms of the covariance matrix. We stress that MP did not come with a dedicated module for the analysis of angular clustering. For this reason we had to write a new module that we have used for the analysis.

  To calibrate the two statistical analyses and assess the presence of systematic errors we applied both of them to the mock catalogs and compared results with theoretical expectations from the `CLASS` code.

  In our analysis we considered the multipole range $\ell = [12, \ell_{\rm max}]$, where the maximum value, $\ell_{\rm max}$, was set to match the shot noise level.

  In the analyses we focused first on the dark matter density Ω_{CDM} parameter alone (i.e. keeping all other cosmological parameters fixed at

their true values), detecting no systematic effect. Then we considered parameter pairs. We chose (Ω_{CDM}, A_s) and (Ω_M, Ω_b). In both cases the analysis highlighted a significant, expected, degeneracy among the parameters in all redshift bins explored. Small systematic errors were detected in the first bin only. They probably reflect the effect of the geometry mask. No systematic effects were detected in the other bins. Random errors are larger in the outer bins.

- In the real data analysis we focused on measuring Ω_{CDM} and A_s, and Ω_M and f_b. Since the analysis of the mocks showed no differences between the MCMC and the χ^2 analysis, we relied on the first to analyze the real data.

  In the modeling phase we add two additional effects respect to the mock analyses. Firstly, we multiply the theoretical spectrum by the pixel window function to mimic the effect of the pixelized map. Secondly, in order to quantify the impact of photo-z errors, we compute the *true* redshift distribution, $dN/dz_{\rm spec}$, convolving the $dN/dz_{\rm phot}$ with a Gaussian $P(z_{\rm spec}|z_{\rm phot})$ following the idea of Sheth and Rossi (2010). Using the correct redshift distribution is important since photometric errors artificially reduce the spectral amplitude on scale corresponding to the size of the error.

- Regarding the analysis of Ω_{CDM} and A_s, it reveals the same type of parameter degeneracy seen in the mocks, with a significant anti-correlation between the two parameters. The results of the first two z-bins are quite similar, and are consistent with the Planck values within the 68% probability intervals. By comparing the spectral amplitude with the Planck one, we also inferred the linear bias of 2MPZ galaxies in the different redshift bins. For the first bin this value is $[0.52, 1.05]$ in the 68% probability interval, that is consistent with the bias estimated for 2MRS galaxies ((Westover, 2007) and (Corsi, 2015)). The value of the dark matter density parameter, Ω_{CDM}, which

lies in the 68% probability interval $[0.19, 0.40]$, is also consistent with the Planck value (0.263).

For the second bin the linear bias is $[1.42, 2.84]$ in the 68% probability interval. This larger bias, consistent with the fact that in this z-bin galaxies are more luminous and more biased than in the first z-bin, shifts upward the likelihood contours, decreasing the best fit value for Ω_{CDM}. However, since the degeneracy between Ω_{CDM} and A_s found in this bin is almost identical as the one found in the previous z-bin, we exclude the presence of redshift-dependent systematic errors.

The results of the third bin reveals much larger, possibly anomalously large, spectral amplitude. The Ω_{CDM} value, however, is consistent with that of the two other bins. A dedicated analysis has been performed to clarify the origin of this large amplitude as we will see in the following.

- The results of the analysis of the total matter density parameter, Ω_M and the baryon fraction f_b for the three redshift bins are fully consistent (in the 68% probability interval) with those of similar studies, like the ones by Thomas et al. (2011) and Blake et al. (2007), that have used angular spectra to estimate baryon fraction. Our results are also consistent with the one from Planck Collaboration et al. (2014). In this analysis the value of the spectral amplitudes in the three redshift bins have been set equal to their best fit values estimated in the previous (Ω_{CDM}, A_s) analysis.

  The fact that also in the third redshift bin the estimated f_b is consistent with the results coming from the first two redshift bins indicates that our estimate is robust to the possible systematic error that, as the large spectral amplitude measured suggests, could affect the third redshift bin.

- To look for possible clues on the origin of the large spectra ampli-

tude measured in the third bin, we first consider the hypothesis that this result is a genuine rather than an anomalous feature. The 2MPZ galaxies, in fact, have been selected by their flux, meaning that objects in the third z-bin are typically brighter than those in the first one. Since luminous objects are more biased, it is possible that the large A_s simply reflects the effect of a luminosity dependence linear bias. To investigate this issue, we first applied different flux cuts on the 2MPZ catalogue and measured the angular power spectra of the corresponding subsamples. Then we convert apparent to absolute magnitudes following Corsi (2015), in order to compare the angular power spectrum of the different redshift bins at the same *luminosity* cut.

The comparison between the spectral amplitudes in the three redshift bins shows no evidence for an excess power in the spectrum of the third bin. We therefore conclude that the large A_s value obtained in the (Ω_{CDM}, A_s) analysis is consistent with the expected luminosity dependent bias and does not hint at any luminosity-dependence systematic error in the data selection.

In conclusion, this thesis was a pilot study to test the possibility of exploiting angular clustering in photometric-redshift surveys to complement that of spectroscopic surveys.

We point out that this is not the first study of this type. Thomas et al. (2011) and Blake et al. (2007) performed a similar analysis using the photometric SDSS surveys of luminous galaxies. However, their studies are different from ours over several aspects. First of all, their samples are deeper and do not overlap with ours. Also, they have a smaller sky coverage and, therefore, a higher density of tracers. Furthermore, their analysis is more simplistic in the sense that they do not use mock catalogues to estimate errors but adopts Gaussian errors instead.

The results of the 2MPZ analysis are encouraging for developing this type of analysis further. Indeed, the work presented in this thesis can be

expanded along several directions.

First of all, more realistic mock catalogs that take into account photo-z errors and correlation among the redshift bins, including some biasing scheme and mimicking the characteristic of other surveys, could be produced at a relatively low computational cost, allowing one to compute precise covariance matrices for 2-point angular statistics.

A more systematic use of the MCMC analysis is also needed in order to explore the parameter space more thoroughly and find which parameter combination is better constrained with angular clustering analyses at different redshifts. All this should be folded in the with the characteristics of some specific next generation photometric surveys. With this respect, it is worth stressing the the machinery developed here is already being used to generate simplified mock catalogs mimicking the characteristics of the Euclid photo-z survey.

Finally, regarding the 2MPZ analysis for the third z-bin, a parallel effort is ongoing in collaboration with Maciej Bilicki (Leiden Observatory) and Andres Balaguera-Antolinez (Roma Tre University) to pinpoint the origin of the results for the third redshift bin that we have outlined above.

Acknowledgments

My first thanks goes to my supervisor Professor Enzo Branchini, for his careful and patient guidance during this work.

Special thanks to Dott. Andres Balaguera Antolinez, for guiding me in my first steps in the world of C++ language, and for having answered to all of my questions and doubts during these years.

I thank Maciej Bilicki for having provided me the 2MPZ catalogue used in this thesis, and for his kind hospitality at Leiden University.

Thanks to Professors Lauro Moscardini and Luigi Guzzo for their comments on this work.

I would also like to thank all the colleagues who have shared with me worries and laughs and lunches and coffees during these years. Thanks to Federica, Gaia, Margherita, Alessandro, Matteo, Andrea, Erica, Marco, Danilo, Matteo, Francesco, Giorgio, Lorenzo, Gabriele, Eleonora, Paolo. Thanks to Martina, my wise and kind older sister.

Thanks to Ilaria, for having shared with me your passion for outreaching, which has now become my passion. You are the model who inspires me.

Thanks to my family, because not even once in these years I felt the distance among us.

And finally, thanks to you, for putting up with me so sweetly.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Allende Prieto, C., An, D., Anderson, K. S. J., Anderson, S. F., Annis, J., Bahcall, N. A., and et al. (2009). The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182:543–558.

Ahn, C. P., Alexandroff, R., Allende Prieto, C., Anderson, S. F., Anderton, T., Andrews, B. H., Aubourg, É., Bailey, S., Balbinot, E., Barnes, R., and et al. (2012). The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *ApJS*, 203:21.

Angulo, R. E., Baugh, C. M., Frenk, C. S., and Lacey, C. G. (2008). The detectability of baryonic acoustic oscillations in future galaxy surveys. *MNRAS*, 383:755–776.

Audren, B., Lesgourgues, J., Benabed, K., and Prunet, S. (2013). Conservative constraints on early cosmology with MONTE PYTHON. *JCAP*, 2:001.

Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., and Steward, L. (2014). Two Micron All Sky Survey Photometric Redshift Catalog: A Comprehensive Three-dimensional Census of the Whole Sky. *ApJS*, 210:9.

Blake, C., Collister, A., Bridle, S., and Lahav, O. (2007). Cosmological baryonic and matter densities from 600000 SDSS luminous red galaxies with photometric redshifts. *MNRAS*, 374:1527–1548.

Blake, C., Davis, T., Poole, G. B., Parkinson, D., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., Drinkwater, M. J., Forster, K., Gilbank, D., Gladders, M., Glazebrook, K., Jelliffe, B., Jurek, R. J., Li, I.-H., Madore, B., Martin, D. C., Pimbblet, K., Pracy, M., Sharp, R., Wisnioski, E., Woods, D., Wyder, T. K., and Yee, H. K. C. (2011). The WiggleZ Dark Energy Survey: testing the cosmological model with baryon acoustic oscillations at z= 0.6. *MNRAS*, 415:2892–2909.

Blake, C., Ferreira, P. G., and Borrill, J. (2004). The angular power spectrum of NVSS radio galaxies. *MNRAS*, 351:923–934.

Blas, D., Lesgourgues, J., and Tram, T. (2011). The Cosmic Linear Anisotropy Solving System (CLASS). Part II: Approximation schemes. *JCAP*, 7:34.

Branchini, E., Davis, M., and Nusser, A. (2012). The linear velocity field of 2MASS Redshift Survey, K_s= 11.75 galaxies: constraints on β and bulk flow from the luminosity function. *MNRAS*, 424:472–481.

Budavári, T., Connolly, A. J., Szalay, A. S., Szapudi, I., Csabai, I., Scranton, R., Bahcall, N. A., Brinkmann, J., Eisenstein, D. J., Frieman, J. A., Fukugita, M., Gunn, J. E., Johnston, D., Kent, S., Loveday, J. N., Lupton, R. H., Tegmark, M., Thakar, A. R., Yanny, B., York, D. G., and Zehavi, I. (2003). Angular Clustering with Photometric Redshifts in the Sloan Digital Sky Survey: Bimodality in the Clustering Properties of Galaxies. *ApJ*, 595:59–70.

Cavuoti, S., Brescia, M., Longo, G., and Mercurio, A. (2012). Photometric redshifts with the quasi Newton algorithm (MLPQNA) Results in the PHAT1 contest. *A&A*, 546:A13.

Chiang, C.-T., Wullstein, P., Jeong, D., Komatsu, E., Blanc, G. A., Ciardullo, R., Drory, N., Fabricius, M., Finkelstein, S., Gebhardt, K., Gronwall, C., Hagen, A., Hill, G. J., Jee, I., Jogee, S., Landriau, M., Mentuch Cooper, E., Schneider, D. P., and Tuttle, S. (2013). Galaxy redshift surveys with sparse sampling. *JCAP*, 12:030.

Cole, S., Percival, W. J., Peacock, J. A., Norberg, P., Baugh, C. M., Frenk, C. S., Baldry, I., Bland-Hawthorn, J., Bridges, T., Cannon, R., Colless, M., Collins, C., Couch, W., Cross, N. J. G., Dalton, G., Eke, V. R., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Glazebrook, K., Jackson, C., Jenkins, A., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Peterson, B. A., Sutherland, W., and Taylor, K. (2005). The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *MNRAS*, 362:505–534.

Coles, P. and Jones, B. (1991). A lognormal model for the cosmological mass distribution. *MNRAS*, 248:1–13.

Coles, P. and Lucchin, F. (1995). *Cosmology - The Origin and Evolution of Cosmic Structure*. Wiley.

Colless, M., Dalton, G., Maddox, S., Sutherland, W., Norberg, P., Cole, S., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Deeley, K., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Madgwick, D., Peacock, J. A., Peterson, B. A., Price, I., Seaborne, M., and Taylor, K. (2001). The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328:1039–1063.

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

Taylor, K. (2003). The 2dF Galaxy Redshift Survey: Final Data Release. *ArXiv Astrophysics e-prints*.

Collister, A. A. and Lahav, O. (2004). ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP*, 116:345–351.

Cooray, A. and Sheth, R. (2002). Halo models of large-scale structure. *Phys. Rep*, 372:1–129.

Corsi, M. (2015). *Estimates of galaxy bias and cosmological parameters from galaxy redshift surveys*. PhD thesis, Università degli Studi Roma Tre.

Crocce, M., Cabré, A., and Gaztañaga, E. (2011). Modelling the angular correlation function and its full covariance in photometric galaxy surveys. *MNRAS*, 414:329–349.

Dawson, K. S., Schlegel, D. J., Ahn, C. P., Anderson, S. F., Aubourg, É., Bailey, S., Barkhouser, R. H., Bautista, J. E., Beifiori, A., Berlind, A. A., Bhardwaj, V., Bizyaev, D., Blake, C. H., Blanton, M. R., Blomqvist, M., Bolton, A. S., Borde, A., Bovy, J., Brandt, W. N., Brewington, H., Brinkmann, J., Brown, P. J., Brownstein, J. R., Bundy, K., Busca, N. G., Carithers, W., Carnero, A. R., Carr, M. A., Chen, Y., Comparat, J., Connolly, N., Cope, F., Croft, R. A. C., Cuesta, A. J., da Costa, L. N., Davenport, J. R. A., Delubac, T., de Putter, R., Dhital, S., Ealet, A., Ebelke, G. L., Eisenstein, D. J., Escoffier, S., Fan, X., Filiz Ak, N., Finley, H., Font-Ribera, A., Génova-Santos, R., Gunn, J. E., Guo, H., Haggard, D., Hall, P. B., Hamilton, J.-C., Harris, B., Harris, D. W., Ho, S., Hogg, D. W., Holder, D., Honscheid, K., Huehnerhoff, J., Jordan, B., Jordan, W. P., Kauffmann, G., Kazin, E. A., Kirkby, D., Klaene, M. A., Kneib, J.-P., Le Goff, J.-M., Lee, K.-G., Long, D. C., Loomis, C. P., Lundgren, B., Lupton, R. H., Maia, M. A. G., Makler, M., Malanushenko, E., Malanushenko, V., Mandelbaum, R., Manera, M., Maraston, C., Margala, D., Masters, K. L., McBride, C. K., McDonald, P., McGreer, I. D., McMahon, R. G., Mena, O., Miralda-Escudé, J., Montero-Dorta, A. D., Montesano, F., Muna, D.,

Myers, A. D., Naugle, T., Nichol, R. C., Noterdaeme, P., Nuza, S. E., Olmstead, M. D., Oravetz, A., Oravetz, D. J., Owen, R., Padmanabhan, N., Palanque-Delabrouille, N., Pan, K., Parejko, J. K., Pâris, I., Percival, W. J., Pérez-Fournon, I., Pérez-Ràfols, I., Petitjean, P., Pfaffenberger, R., Pforr, J., Pieri, M. M., Prada, F., Price-Whelan, A. M., Raddick, M. J., Rebolo, R., Rich, J., Richards, G. T., Rockosi, C. M., Roe, N. A., Ross, A. J., Ross, N. P., Rossi, G., Rubiño-Martin, J. A., Samushia, L., Sánchez, A. G., Sayres, C., Schmidt, S. J., Schneider, D. P., Scóccola, C. G., Seo, H.-J., Shelden, A., Sheldon, E., Shen, Y., Shu, Y., Slosar, A., Smee, S. A., Snedden, S. A., Stauffer, F., Steele, O., Strauss, M. A., Streblyanska, A., Suzuki, N., Swanson, M. E. C., Tal, T., Tanaka, M., Thomas, D., Tinker, J. L., Tojeiro, R., Tremonti, C. A., Vargas Magaña, M., Verde, L., Viel, M., Wake, D. A., Watson, M., Weaver, B. A., Weinberg, D. H., Weiner, B. J., West, A. A., White, M., Wood-Vasey, W. M., Yeche, C., Zehavi, I., Zhao, G.-B., and Zheng, Z. (2013). The Baryon Oscillation Spectroscopic Survey of SDSS-III. *AJ*, 145:10.

Di Dio, E., Montanari, F., Lesgourgues, J., and Durrer, R. (2013). The classgal code for relativistic cosmological large-scale structure. *JCAP*, 11:044.

Dodelson, S. (2003). *Modern Cosmology*. Academic Press.

Doran, M. (2005). CMBEASY: an object oriented code for the cosmic microwave background. *JCAP*, 10:011.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, 7(1):1–26.

Einstein, A. (1916). Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354:769–822.

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J.,

Connolly, A., Csabai, I., Doi, M., Fukugita, M., Frieman, J. A., Glazebrook, K., Gunn, J. E., Hendry, J. S., Hennessy, G., Ivezić, Z., Kent, S., Knapp, G. R., Lin, H., Loh, Y.-S., Lupton, R. H., Margon, B., McKay, T. A., Meiksin, A., Munn, J. A., Pope, A., Richmond, M. W., Schlegel, D., Schneider, D. P., Shimasaku, K., Stoughton, C., Strauss, M. A., SubbaRao, M., Szalay, A. S., Szapudi, I., Tucker, D. L., Yanny, B., and York, D. G. (2005). Detection of the baryon acoustic peak in the large-scale correlation function of sdss luminous red galaxies. *ApJ*, 633:560–574.

Firth, A. E., Lahav, O., and Somerville, R. S. (2003). Estimating photometric redshifts with artificial neural networks. *MNRAS*, 339:1195–1202.

Fisher, K. B., Scharf, C. A., and Lahav, O. (1994). A spherical harmonic approach to redshift distortion and a measurement of Omega(0) from the 1.2-Jy IRAS Redshift Survey. *MNRAS*, 266:219.

Fry, J. N. (1986). Nonlinear statistics and biasing. *ApJL*, 308:L71–L74.

Górski, K. M., Hivon, E., Banday, A. J., Wandelt, B. D., Hansen, F. K., Reinecke, M., and Bartelmann, M. (2005). HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622:759–771.

Gunn, J. E. and Gott, III, J. R. (1972). On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution. *ApJ*, 176:1.

Guzzo, L., Scodeggio, M., Garilli, B., Granett, B. R., Fritz, A., Abbas, U., Adami, C., Arnouts, S., Bel, J., Bolzonella, M., Bottini, D., Branchini, E., Cappi, A., Coupon, J., Cucciati, O., Davidzon, I., De Lucia, G., de la Torre, S., Franzetti, P., Fumana, M., Hudelot, P., Ilbert, O., Iovino, A., Krywult, J., Le Brun, V., Le Fèvre, O., Maccagni, D., Małek, K., Marulli, F., McCracken, H. J., Paioro, L., Peacock, J. A., Polletta, M., Pollo, A., Schlagenhaufer, H., Tasca, L. A. M., Tojeiro, R., Vergani, D., Zamorani, G., Zanichelli, A., Burden, A., Di Porto, C., Marchetti, A., Marinoni, C.,

Mellier, Y., Moscardini, L., Nichol, R. C., Percival, W. J., Phleps, S., and Wolk, M. (2014). The vimos public extragalactic redshift survey (vipers). an unprecedented view of galaxies and large-scale structure at $0.5 < z < 1.2$. *A&A*, 566:A108.

Hambly, N. C., Davenhall, A. C., Irwin, M. J., and MacGillivray, H. T. (2001a). The SuperCOSMOS Sky Survey - III. Astrometry. *MNRAS*, 326:1315–1327.

Hambly, N. C., Irwin, M. J., and MacGillivray, H. T. (2001b). The SuperCOSMOS Sky Survey - II. Image detection, parametrization, classification and photometry. *MNRAS*, 326:1295–1314.

Hambly, N. C., MacGillivray, H. T., Read, M. A., Tritton, S. B., Thomson, E. B., Kelly, B. D., Morgan, D. H., Smith, R. E., Driver, S. P., Williamson, J., Parker, Q. A., Hawkins, M. R. S., Williams, P. M., and Lawrence, A. (2001c). The SuperCOSMOS Sky Survey - I. Introduction and description. *MNRAS*, 326:1279–1294.

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. *ApJL*, 374:L1–L4.

Hastings, W. K. (1970). Monte Carlo samping methods using Markov chains and their applications. *Biometrika*, 57:97–109.

Hauser, M. G. and Peebles, P. J. E. (1973). Statistical Analysis of Catalogs of Extragalactic Objects. II. the Abell Catalog of Rich Clusters. *ApJ*, 185:757–786.

Hilbert, S., Hartlap, J., and Schneider, P. (2011). Cosmic shear covariance: the log-normal approximation. *A&A*, 536:A85.

Hivon, E., Górski, K. M., Netterfield, C. B., Crill, B. P., Prunet, S., and Hansen, F. (2002). MASTER of the Cosmic Microwave Background

Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets. *ApJ*, 567:2–17.

Howlett, C., Lewis, A., Hall, A., and Challinor, A. (2012). CMB power spectrum parameter degeneracies in the era of precision cosmology. *JCAP*, 1204:027.

Hu, W. and Dodelson, S. (2002). Cosmic Microwave Background Anisotropies. *Annual Review of Astronomy and Astrophysics*, 40:171–216.

Hubble, E. (1929). A relation between distance and radial velocity among extra-galactic nebulae. *Proceedings of the National Academy of Sciences*, 15(3):168–173.

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

Huchra, J. P., Geller, M. J., Clemens, C. M., Tokarz, S. P., and Michel, A. (1996). VizieR Online Data Catalog: The CfA Redshift Catalogue, Version June 1995 (Huchra+ 1995). *VizieR Online Data Catalog*, 7193.

Huchra, J. P., Macri, L. M., Masters, K. L., Jarrett, T. H., Berlind, P., Calkins, M., Crook, A. C., Cutri, R., Erdoğdu, P., Falco, E., George, T., Hutcheson, C. M., Lahav, O., Mader, J., Mink, J. D., Martimbeau, N., Schneider, S., Skrutskie, M., Tokarz, S., and Westover, M. (2012). The 2MASS Redshift Survey - Description and Data Release. *ApJS*, 199:26.

Jarrett, T. H., Chester, T., Cutri, R., Schneider, S., Skrutskie, M., and Huchra, J. P. (2000). 2MASS Extended Source Catalog: Overview and Algorithms. *AJ*, 119:2498–2531.

Jenkins, A., Frenk, C. S., White, S. D. M., Colberg, J. M., Cole, S., Evrard, A. E., Couchman, H. M. P., and Yoshida, N. (2001). The mass function of dark matter haloes. *MNRAS*, 321:372–384.

Jeong, D. and Komatsu, E. (2006). Perturbation Theory Reloaded: Analytical Calculation of Nonlinearity in Baryonic Oscillations in the Real-Space Matter Power Spectrum. *ApJ*, 651:619–626.

Jones, D. H., Read, M. A., Saunders, W., Colless, M., Jarrett, T., Parker, Q. A., Fairall, A. P., Mauch, T., Sadler, E. M., Watson, F. G., Burton, D., Campbell, L. A., Cass, P., Croom, S. M., Dawe, J., Fiegert, K., Frankcombe, L., Hartley, M., Huchra, J., James, D., Kirby, E., Lahav, O., Lucey, J., Mamon, G. A., Moore, L., Peterson, B. A., Prior, S., Proust, D., Russell, K., Safouris, V., Wakamatsu, K.-I., Westra, E., and Williams, M. (2009). The 6df galaxy survey: final redshift release (dr3) and southern large-scale structures. *MNRAS*, 399:683–698.

Kaiser, N. (1984). On the spatial correlations of Abell clusters. *ApJL*, 284:L9–L12.

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

Kayo, I., Taruya, A., and Suto, Y. (2001). Probability distribution function of cosmological density fluctuations from a gaussian initial condition: Comparison of one-point and two-point lognormal model predictions with n-body simulations. *The Astrophysical Journal*, 561(1):22.

Kingman, J. (1992). *Poisson Processes*. Clarendon Press.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220:671–680.

Kochanek, C. S., Pahre, M. A., Falco, E. E., Huchra, J. P., Mader, J., Jarrett, T. H., Chester, T., Cutri, R., and Schneider, S. E. (2001). The K-Band Galaxy Luminosity Function. *ApJ*, 560:566–579.

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern,

M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *ApJS*, 192:18.

Kovács, A., Szapudi, I., Granett, B. R., and Frei, Z. (2013). Cross-correlation of WMAP7 and the WISE full data release. *MNRAS*, 431:L28–L32.

Lahav, O. and Suto, Y. (2004). Measuring our Universe from Galaxy Redshift Surveys. *Living Reviews in Relativity*, 7.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. ., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., and et al. (2011). Euclid Definition Study Report. *ArXiv e-prints*.

Lesgourgues, J. (2011). The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview. *ArXiv e-prints*.

Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient computation of CMB anisotropies in closed FRW models. *ApJ*, 538:473–476.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *JCP*, 21:1087–1092.

Mo, H., van den Bosch, F. C., and White, S. (2010). *Galaxy Formation and Evolution*.

Ostriker, J. P. (1986). *Galaxy Distances and Deviations from the Hubble Flow; Summary Remarks*. Springer Netherlands, Dordrecht.

Padmanabhan, N., Schlegel, D. J., Seljak, U., Makarov, A., Bahcall, N. A., Blanton, M. R., Brinkmann, J., Eisenstein, D. J., Finkbeiner, D. P., Gunn,

J. E., Hogg, D. W., Ivezić, Ž., Knapp, G. R., Loveday, J., Lupton, R. H., Nichol, R. C., Schneider, D. P., Strauss, M. A., Tegmark, M., and York, D. G. (2007). The clustering of luminous red galaxies in the Sloan Digital Sky Survey imaging data. *MNRAS*, 378:852–872.

Parkinson, D., Riemer-Sørensen, S., Blake, C., Poole, G. B., Davis, T. M., Brough, S., Colless, M., Contreras, C., Couch, W., Croom, S., Croton, D., Drinkwater, M. J., Forster, K., Gilbank, D., Gladders, M., Glazebrook, K., Jelliffe, B., Jurek, R. J., Li, I.-h., Madore, B., Martin, D. C., Pimbblet, K., Pracy, M., Sharp, R., Wisnioski, E., Woods, D., Wyder, T. K., and Yee, H. K. C. (2012). The WiggleZ Dark Energy Survey: Final data release and cosmological results. *Physical Review D*, 86(10):103518.

Peacock, J. A. (1999). *Cosmological Physics*.

Peacock, J. A. and Dodds, S. J. (1996). Non-linear evolution of cosmological power spectra. *MNRAS*, 280:L19–L26.

Peacock, J. A., Hambly, N. C., Bilicki, M., MacGillivray, H. T., Miller, L., Read, M. A., and Tritton, S. B. (2016). The SuperCOSMOS all-sky galaxy catalogue. *MNRAS*, 462:2085–2098.

Peebles, P. J. E. (1973). Statistical Analysis of Catalogs of Extragalactic Objects. I. Theory. *ApJ*, 185:413–440.

Peebles, P. J. E. (1980). *The large-scale structure of the universe*.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Project, T. S. C. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517:565–586.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014). Planck 2013 results. XVI. Cosmological parameters. *A&A*, 571:A16.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical recipes in C. The art of scientific computing*.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116:1009–1038.

Rubin, V. C., Ford, W. K. J., and . Thonnard, N. (1980). Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *ApJ*, 238:471–487.

Seljak, U. and Zaldarriaga, M. (1996). A Line-of-Sight Integration Approach to Cosmic Microwave Background Anisotropies. *ApJ*, 469:437.

Sheth, R. K. and Rossi, G. (2010). Convolution- and deconvolution-based estimates of galaxy scaling relations from photometric redshift surveys. *MNRAS*, 403:2137–2142.

Sheth, R. K. and Tormen, G. (2002). An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *MNRAS*, 329:61–75.

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam,

R., Van Dyk, S., and Wheelock, S. (2006). The Two Micron All Sky Survey (2MASS). *AJ*, 131:1163–1183.

Smith, R. E., Peacock, J. A., Jenkins, A., White, S. D. M., Frenk, C. S., Pearce, F. R., Thomas, P. A., Efstathiou, G., and Couchman, H. M. P. (2003). Stable clustering, the halo model and non-linear cosmological power spectra. *MNRAS*, 341:1311–1332.

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., de Amici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., Mather, J., Meyer, S. S., Moseley, S. H., Murdock, T., Rokke, L., Silverberg, R. F., Tenorio, L., Weiss, R., and Wilkinson, D. T. (1992). Structure in the COBE differential microwave radiometer first-year maps. *ApJL*, 396:L1–L5.

Springel, V. (2005). The cosmological simulation code GADGET-2. *MNRAS*, 364:1105–1134.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636.

Szapudi, I. and Pan, J. (2004). On Recovering the Nonlinear Bias Function from Counts-in-Cells Measurements. *ApJ*, 602:26–37.

Tagliaferri, R., Longo, G., Andreon, S., Capozziello, S., Donalek, C., and Giordano, G. (2003). Neural Networks for Photometric Redshifts Evaluation. *Lecture Notes in Computer Science*, 2859:226–234.

Taruya, A., Takada, M., Hamana, T., Kayo, I., and Futamase, T. (2002). Lognormal Property of Weak-Lensing Fields. *ApJ*, 571:638–653.

Thomas, S. A., Abdalla, F. B., and Lahav, O. (2011). The angular power spectra of photometric Sloan Digital Sky Survey luminous red galaxies. *MNRAS*, 412:1669–1685.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples. 29(2):614–614.

Vanzella, E., Cristiani, S., Fontana, A., Nonino, M., Arnouts, S., Giallongo, E., Grazian, A., Fasano, G., Popesso, P., Saracco, P., and Zaggia, S. (2004). Photometric redshifts with the Multilayer Perceptron Neural Network: Application to the HDF-S and SDSS. *A&A*, 423:761–776.

Verde, L. (2010). Statistical Methods in Cosmology. In Wolschin, G., editor, *Lecture Notes in Physics, Berlin Springer Verlag*, volume 800 of *Lecture Notes in Physics, Berlin Springer Verlag*, pages 147–177.

Verde, L. (2015). A Taste of Cosmology. *ArXiv e-prints*.

Veropalumbo, A., Marulli, F., Moscardini, L., Moresco, M., and Cimatti, A. (2016). Measuring the distance-redshift relation with the baryon acoustic oscillations of galaxy clusters. *MNRAS*, 458:1909–1920.

Way, M. J., Foster, L. V., Gazis, P. R., and Srivastava, A. N. (2009). New Approaches to Photometric Redshift Prediction Via Gaussian Process Regression in the Sloan Digital Sky Survey. *ApJ*, 706:623–636.

Westover, M. (2007). PhD Thesis, Harvard University.

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier, III, T. N., McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., Larsen, M., Cardon, J. G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L.,

and Tsai, C.-W. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *AJ*, 140:1868–1881.

Xavier, H. S., Abdalla, F. B., and Joachimi, B. (2016). Improving lognormal models for cosmological fields. *MNRAS*, 459:3693–3710.

Zel'dovich, Y. B. (1970). Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89.

Zinnecker, H. (1984). Star formation from hierarchical cloud fragmentation - A statistical theory of the log-normal Initial Mass Function. *MNRAS*, 210:43–56.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.