Scuola Dottorale in Scienze Matematiche e Fisiche

XXVII Ciclo

Estimates of galaxy bias and cosmological parameters from galaxy redshift surveys

Tutor:
Prof. Enzo Branchini

Dottoranda:
Martina Corsi

Coordinatore:
Prof. Roberto Raimondi

Contents

Introduction

General Relativity is the theoretical foundation of Modern Cosmology. Almost one century after the publication of this fundamental theory (Einstein A., 1916), the state of art of cosmology is that a concordance model consisting in an expanding Universe, approximatively flat, in which matter is around one third of the total energy content agree with observations. According to this model about 15% of the total matter content is in the form of non relativistic ordinary particles, usually referred to as *baryonic matter*, while the remaining part is in the form of a dark component made of particles not predicted by the Standard Model of particle physics. Because of its domination over the baryonic component this *dark matter* plays a crucial role in determining the evolution of the cosmic structures in the Universe. The existence of this dark component was firstly proposed by Zwicky in 1933 when analizing the motions of the galaxies in the Coma cluster. He noticed that the velocity dispersion within the cluster implied a virial mass of the cluster well above the one associated with visible galaxies. Hence the need of an additional, dominating "dark" mass component. Another indirect observational evidence of the dark matter came later (Rubin et al. 1980) from the rotation velocity profiles of stars and gas in spiral galaxies. The existence of an extra component, not interacting electromagnetically but only gravitationally, could explain the deviation of the rotation curves from the model expected by gravity theory if visible matter were responsible for the gravity field. More recently the existence of an extra component, called *dark energy*, was advocated to explain a number of additional cosmological observations. First of all 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, Permuttler et al. 1999) . Other independent evidences in favour of dark energy come from the abundance of clusters of galaxies, the Large Scale Structure and the Cosmic Microwave Background. The emerging picture is that of a Universe in which most of the matter is dark but the energy budget is dominated by yet another component, called dark energy, responsable for the accelerated expansion of the Universe. Dark energy is not the only possible theoretical explanation of cosmic acceleration. Other possibilities come if one assumes that General Relativity theory breaks down on cosmological scales and a modified theory

of gravity must be advocated instead. Despite of the remarkable progess of cosmology in the last decades, the nature of the dark matter and dark energy and, in fact, their physical reality are still under scrutinity.

The study of the *Large Scale Structure* of the Universe, namely the spatial distribution of matter on cosmological scales, coupled to the "classical" cosmological test based on distance estimates prompt the possibility of addressing these issues and, in particular, to test the validity of General Relativity. In the last decades a significant number of *galaxy surveys* have traced the spatial distribution of different types of galaxies using their redshift as distance measure. These surveys cover different sky areas and are differently dept. The 2MASS Redshift Survey (Hucra et al. 2012), for example, which we focus on in this thesis, is aimed to map the distribution of galaxies in the local Universe. Higher redshifts are spanned for example by the 2dFGRS (Colless et al. 2003) and SDSS (i.e. Abazajian et al. 2009, data release 7). These two surveys cover quite large sky patches, at intermediate redshifts: the 2dFGRS is characterized by a median redshift $z \sim 0.1$ and sky coverage 1800 deg^2, the SDSS (I - II) covers a largest sky area ($\sim$ 9000 deg^2, namely $\sim$ 23% of the sky, for the DR7). In the last few years , deeper surveys have spanned even higher redshifts such as the VIMOS Public Extragalactic Redshift Survey (VIPERS, Garilli et al. 2014), which covers a relatively small area of the sky ($\sim$ 16 and 8 square degree) but is characterized by a wide redshift coverage (extending from z =0.45 to z =0.95), WiggleZ (Parkinson D. et al. 2012, median redshift 0.6) or the BOSS survey (Ahn et al. 2014) , that's part of the third SDSS (SDSS-III) and contains more than 1 million galaxy spectra.

Assuming that galaxies trace the distribution of the matter (mostly composed by dark matter as we have seen) these surveys show that the mass is not randomly distributed in the Universe but organized in a coherent pattern of large scale structures dubbed "cosmic web." This is thought to be the result of the evolution, mostly due to the effect of gravity, of primordial perturbations on early quasi-homogeneus and isotropic density field. Density fluctuations in the dark matter component above a characteristic scale, called Jeans lenght, started growing by gravitational instability eventually forming the cosmic structure that we observe in our Universe today. The baryonic component, after decoupling itself from radiation, responds to the gravitational pull of the dark matter fluctuations and, after a short period,trace the same large scale structures as the dark matter; a picture that has been drastically confirmed by the detection of the temperature fluctuations in the Cosmic Microwave Background (COBE, Smoot et al.1992) and the spatial distribution of galaxies in the surveys (BAOs, Eisenstein et al. 2005). Structure formation theory allows to follow this evolution: a simple linear perturbative approach is valid in the limit of small perturbations, i.e. both at early epochs or on very large scales. On small scales, when density fluctuations become large, the growth can

not be described by the linear approximation and the theoretically more challenging nonlinear regime sets in. In fact, the full non linear evolution of density fluctuations can only be tested with numerical techniques: the N-body simulations. One success of theoretical cosmology is the ability to predict either analitically or numerically the statistical properties of the mass distribution and their relation with the fundamental parameters of the model, which we call *cosmological parameters*. As a consequence, the observational estimates of such statistical properties are crucial to determine the parameters and constrain the model. Hence the widespread intent and effort in building large galaxy surveys capable of tracing the three dimensional distribution of extra-galactic objects. To go from there to the undelying mass distribution one needs to assume a "bias" relation between the mass and the galaxy distribution that must be considered as an extra degree of freedom of the model.

In this thesis we investigate the possiblity of measuring the "bias relation" with present surveys and surveys of next generation. We focus on the power spectrum of the galaxy distribution, that is the Fourier transform of the 2-point correlation function and is sufficient to characterize completely the statistical properties of the matter density field if it is Gaussian distributed. We implement a code which computes the "classical" estimator of the power spectrum of the galaxy distribution proposed by Feldman, Kaiser and Peacock (Feldman et al. 1994), dubbed FKP estimator. We validate the possibility of using this estimator to all sky surveys. Not all the cosmological signals that can be detected in the galaxy power spectrum require a full sky coverage, even if they need in any case large and dense coverage in order to reduce statistical errors on the measures. As pointed out in Bilicky (2014), a coverage of the full sky can be useful for example in order to test the validity of the isotropy and homogeneity assumption at late times. In order to validate the possibility of using FKP on all sky surveys we have considered the most difficult scenario, as the one of an all sky survey and local as the 2MASS Redshift Survey, 2MRS. It is a spectroscopic redshift survey, which contains approximatively 45000 galaxies of the local Universe, at median redshift 0.03, selected in the near infrared magnitude band K_s. Some of the difficulties in computing the power spectrum of such a survey are: the small volume covered by the survey, which implies a relatively small value of the largest explorable scales, and as a consequence, the limited range of scales in which the linear perturbation theory approach can be considered valid. This issue is emphasized by the small median redshift of the survey that implies late time highly non linear evolution in a wide range of scales, too. Associated with that, high statistical errors characterize the scales of cosmological interest. Furthermore, the FKP estimator is implemented in cubic boxes and consists in the decomposition of the galaxy distribution in Fourier coefficients. The peculiar motions of the galaxies, associated with the forming cosmic structures, induce anisotropies in the spatial distribution of

these objects as it is deduced from their redshift. These distortions are along the radial direction (namely the line of sight direction). The FKP method doesn't lend naturally to the study of such distortions with an all sky local survey. Other methods more computationally expensive are expected to work better. We concentrate on the monopole of the power spectrum so the latter issue , as we'll see, is not expected to affect significantly our measures. We used mock catalogs in order to test our ability of measuring the power spectrum and its distortions due to the peculiar velocities, to the nonlinear bias relation and to the dynamical non linearities. We apply all this stuff to the 2MRS in order to measure the bias of the 2MRS galaxies and the matter density parameter. Finally, we study the possible dependence of the bias on the scale, induced by non linear effects in the galaxy power spectrum and how this dependence is constrained by future observations. Differently from the 2MRS, future redshift surveys, like *Euclid* (Laurejis R. et al., 2012), are expected to estimate the cosmological parameters at the level of per-cent and a potential scale dependent bias could affect the results obtained from the analysis of the power spectrum.

The structure of the thesis is as follows.

In Chapter 1 we present the fundamental aspects of the Standard cosmology, starting from the Friedmann equations and the linear perturbation theory solutions from which the linear power spectrum of matter can be derived. We describe how the matter power spectrum depends on the cosmological parameters and then we briefly describe the possible approaches to non linearities. We describe the galaxy power spectrum and discuss about the "galaxy bias" and distortions due to the peculiar motions of the galaxies.

In Chapter 2 we describe how the galaxy power spectrum is in practice computed starting from the spatial distribution of the galaxies. First we discuss the simple case of a constant density distribution within a cubic volume, then we consider the more general case in which both angular and radial distribution are not constant due to selection effects which characterize the real world galaxy surveys. In this context we describe the estimator proposed by Feldman , Kaiser and Peacock and how it has been practically implemented by us.

In Chapter 3 we show the test done in order to assess the performance of the code which computes the power spectrum and we show the preliminary tests done with simulations in order to set approximatively the wavenumber range in which we can push the analysis of the 2MRS. We test also the code developed in order to account for the window function of the survey and show the tests done in order to check the performance of this code with a 2MRS-like window.

In Chapter 4 we describe the 2MRS and the mock catalogs used in order to obtain an estimate of the uncertainties of the measured power spectrum and test the non linear models of bias and redshift distortions to be used in

the analysis and then, we show the results of the analysis of the 2MRS power spectrum. We study the dependence of the clustering on the luminosity of the galaxies and on the morphological type , splitting the main sample in different smaller subsamples and measuring the relative bias. Finally we show the estimates of the absolute bias parameter and the matter density parameter.

In Chapter 5 we show how we forecast the scale dependent parametrizations of the bias by means of a Fisher matrix analysis for a *Euclid*-like survey and show the results.

We conclude with a summary and discuss the future prospects.

Chapter 1

Standard Cosmology

In this Chapter we present a brief review of the basic concepts of the Background Cosmology and gravitational instability theory that togheter constitute the backbone of the Standard Cosmological Model. Then we'll focus on the power spectrum of matter, its shape and how it encodes information about fundamental cosmological parameters. Since in this thesis the main focus is on the power spectrum of galaxies and its estimates, we shall also discuss the so called "galaxy bias" relation required to link the observed galaxy spectrum to that of the matter.

1.1 Background Cosmology

One class of solutions of Einstein's equations is based on the *Cosmological Principle*. It assumes isotropy (*rotational invariance*) and homogeneity (*traslational invariance*) of the Universe. The evidence of homogeneity at extremely large scales, more than several hundreds of Mpc [1], has been source of debate up to mid-90s, but today is almost generally accepted thanks to the advent of large galaxy redshift surveys like 2dF and SDSS able to encompass scales significant larger the scale of coherence in the spatial distribution of galaxies. In other words, when smoothed on these scales the distribution of galaxies appears to be homogeneus and isotropic. These solutions imply that the Universe is filled with ideal fluids of non interacting particles characterized by an energy (or mass) density, an equation of state and a temperature, all of them constan across space, i.e. with no fluctuations. From these assumptions derives the so called Standard Cosmological Model. Under the restrictions imposed by the Cosmological Principle, the only allowed metrics is the *Friedmann-Robertson-Walker* one (FRW) that can

[1] 1 pc=3.086 $\times 10^{13}$ Km

be expanded as:

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu = -c^2 dt^2 + a^2(t)[dr^2 + S_k(r)^2(d\theta^2 + sen^2(\theta)d\phi^2)] \tag{1.1}$$

where μ and ν vary from 0 to 3, with the 0 index related to the time coordinate ($dt = dx^0$) and 1,2,3 related to the spatial coordinates. $g_{\mu\nu}$ is the *metric tensor*. In the right side of equation 1.1 c represents the speed of light, t is the cosmological *proper time*; r, θ and ϕ are *spatial comoving coordinates* that do not vary over time, whereas the physical distances do indeed vary because of the *scale factor* $a(t)$ that can either increase (expansion) or decrease (contraction) over time. Such expension/contraction is homogeneous and isotropic so that the Cosmological Principle remains valid over cosmic epochs. $S_k(r) = Rsen(r/R)$, r , $Rsenh(r/R)$ respectively in case of positive ("close" Universe) , null ("flat" Universe) or negative curvature ("open" Universe) . k refers to the *curvature parameter*, $k = k'/R^2$, where k' can be equal to 1, 0, -1 and identifies with its value the sign of the curvature, respectively for close , flat and open geometry and R is the *curvature radius*. The *scale factor* $a(t)$ has the dimensions of a lenght as opposite to the adimensional comoving coordinates. Its value at present epoch, $a(t_0)$, is convectionally set to 1. Expliciting the curvature parameter k , the FRW metric becomes:

$$ds^2 = -c^2 dt^2 + a(t)^2 \left[\frac{dx^2}{1 - kx^2} + x^2(d\theta^2 + sen^2(\theta)d\phi^2) \right] \tag{1.2}$$

where x=$S_k(r)$.

Eintein's field equations link space-time geometry and energy content of the Universe and are given by:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu} \tag{1.3}$$

where the *Einstein tensor* $G_{\mu\nu}$ describes geometry; $R_{\mu\nu}$ is the *Ricci tensor* and depends on metric and its derivative; R is the *Ricci scalar*, contraction of Ricci tensor $R = g^{\mu\nu} R_{\mu\nu}$; G is the gravitational constant and $T_{\mu\nu}$ is the *energy-momentum tensor* and describes the energy content of the Universe. On large scales, the latter can be considered as the one of a perfect fluid (this assumption also derives from having assumed the Cosmological Principle):

$$T_{\mu\nu} = (p + \rho) u_\mu u_\nu / c^2 + p g_{\mu\nu} \tag{1.4}$$

where u_μ is the quadrivelocity of the fluid , ρ its energy density and p its pressure. From the 00 tensor component of the Einstein equations and assuming FRW metric one can derive the *Friedmann* equation (1922) for

$a(t)$:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\sum_i \rho_i - \frac{kc^2}{a^2}. \tag{1.5}$$

The sum $\sum_i \rho_i$ runs over the energy densities of all components, assumed to be perfect fluids with their equation of state $p(\rho)$ and represents the mean total energy density in the Universe. It is worth defining the *density parameter* for the $i-$th component :

$$\Omega_i = \frac{8\pi G\rho_i}{3c^2H^2} \tag{1.6}$$

and the *curvature density parameter* :

$$\Omega_k = -\frac{kc^2}{a^2H^2} \tag{1.7}$$

and the *total density parameter*

$$\Omega(t) = \sum_i \Omega_i(t). \tag{1.8}$$

where $H = \dot{a}(t)/a(t)$ is the *Hubble parameter*. With these definitions the *Friedmann* equation becomes:

$$1 = \Omega + \Omega_k. \tag{1.9}$$

Written in this form *Friedmann* equation shows clearly that the values of Ω e Ω_k are not independent (this arises from Einstein's equations coupling). If $\Omega < 1$ then $k < 0$ (open universe); if $\Omega > 1$ then $k > 0$ (close universe), if $\Omega = 1$ then $k = 0$ (flat universe).

The value of the energy density for which $\Omega = 1$ is called *critical density* and is given by:

$$\rho_c = \frac{3c^2H^2}{8\pi G} \tag{1.10}$$

So the density parameter, written in terms of ρ_c, is:

$$\Omega(t) = \frac{\sum_i \rho_i(t)}{\rho_c(t)}. \tag{1.11}$$

From now on, we use Ω_i for the present day energy density parameters, and explicitate the time dependence only for $t \neq t_0$. The *Friedmann* equation evaluated at present day becomes:

$$1 = \frac{8\pi G\rho_0}{3c^2H_0^2} - \frac{kc^2}{H_0^2} \tag{1.12}$$

where subscript 0 indicates the present epoch and , as said before, $a(t_0) = 1$. $H_0 \equiv H(t_0)$ is referred to as *Hubble "constant"*. If present energy density of

each component and H_0 are known then is possible to obtain the curvature. Then, if we want to know also the time evolution of the *scale factor* $a(t)$, a second differential equation is needed. It is obtained setting to 0 the covariant derivative of energy-momentum tensor obtaining the conservation equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0. \tag{1.13}$$

Combining equation 1.5 and 1.13 another significant equation, *acceleration equation*, known also as *Friedmann* equation too, can be derived:

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

The sum is over each component contributing to the energy density of the Universe, Ω. This equation implies that ordinary components with positive pressure, like radiation and non relativistic matter, determine a decelerated expansion $\ddot{a} < 0$. A negative expansion can be obtained if a component with $p < -(1/3)\rho$ is present and dominates the energy budget. Now equations 1.5 and 1.13 can be solved for $a(t)$ and $\rho(t)$, if the *equation of state*, namely a relation between pressure and density , is known for each component. Writing the equation of state in the following form, proposed by Zel'dovich:

$$p = w\rho \tag{1.15}$$

where w is an adimensional *equation of state parameter*, a solution for the scale factor and energy density in function of proper time can be obtained. The evolution of the energy density in function of the scale factor can be obtained from equation 1.13 for constant w and for each component and is given by:

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

where $\rho_{w,0}$ is the present day energy density.

In 1929 Edwin Hubble discovered that galaxies receed from us with a velocity proportional to their distance , measuring the redshift z of these objects. Such experimental law agrees with the hypothesis of an homogeneus and isotropic expansion of the Universe and is fully consistent with the FRW metrics, where a positive evolution of the scale factor would determine a homogeneus and isotropic expansion of the system, as observed. This evidence 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. The temperature associated to each component (reducing to a single temperature when these are coupled) also increases backward in time. Extrapolating the density, pressure and temperature back in time the model predicts 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. Predictions reveal that non negligible quantities of light elements would have been produced (notably He and D) that we have indeed observed at hight redshifts in the intergalactic medium, hence providing additional observational support to the Big Bang model. At high density and temperature baryons and photons were highly coupled. Decoupling occurs when the temperature drops enough for the hydrogen ionization finishes to decrease significantly. This occurs at $z \sim 1100$ (i.e. $3{\times}10^8$ years after the Big Bang) when the baryons decouple from photons and the Universe becomes transparent to the radiation that , because of the adiabatic cooling, we observe now as a Cosmic Microwave Background radiation. The commonly accepted model for the present day Universe as depicted by a variety of cosmological observations is expanding, approximatively flat, accelerating ($\ddot{a} > 0$). In the energy density budget, nonrelativistic matter constitutes $\sim 1/3$ of the total of which 15% is ordinary matter (defined conventionally *baryonic*) and 85% is *cold dark matter* (not electromagnetically but only gravitationally interacting). A small contribution, less than 10^{-4}%, is the relativistic matter in form of photons and neutrinos, generically called *radiation*. Finally, the remaining two thirds is believed to be made of *dark energy*, with negative pressure responsable for acceleration (see eq. 1.14). The simplest model of dark energy is the *cosmological constant* Λ, which is characterized by a constant energy density and equation of state parameter $w = -1$. This model of universe, with a cosmological constant and non relativistic *cold* dark matter is generally referred to as ΛCDM.

Table 1.1 shows the best fit energy density parameters for the ΛCDM model from the Planck experiment (Planck collaboration, 2013). Ω_M is the matter density parameter at present time, Ω_b and Ω_c are respectively the baryonic matter and cold dark matter density parameters.

	best fit	68% limits
H_0	67.11	67.4±1.4
$\Omega_b h^2$	0.022068	0.02207±0.00033
$\Omega_c h^2$	0.12029	0.1196±0.0031
Ω_M	0.3175	0.314±0.020
Ω_Λ	0.6825	0.686±0.020

Table 1.1: Results for cosmological parameters for ΛCDM model. Planck collaboration 2013. h represents the Hubble constant in unity of $100 Kms^{-1}Mpc^{-1}$.

The parameter of state w for non relativistic matter, simply referred to as *matter*, is ~ 0, therefore, according to 1.16, matter density evolves as a^{-3} decreasing with the increase of volume due to the expansion. The parameter of state for radiation is $1/3$ so radiation density evolves as a^{-4}. The extra factor a^{-1} comes from the increase of the radiation wavelenghts

due to expansion. The nature of dark energy is matter of debate: if it is in the form of a cosmological constant then $w = -1$ and its the energy density remains constant with time. Values of $w < -(1/3)$ either constant or time dependent characterize more general forms of dark energy.

In the limit $a \to 0$ the component with the higher value of w dominates, for $a \to \infty$ the component with the lower value of w dominates. In a model of expanding universe with radiation , matter and dark energy , radiation dominates for $a \to 0$ (early times), and dark energy dominates at later times. The present day values of energy densities and other type of observations suggest that radiation-matter domination transition happens for a value of the scale factor $a_{eq} \cong 2.8 \times 10^{-4}$ namely around 47000 year after Big Bang. a_{eq} denotes the value of the scale factor at the epoch when the energy density of the radiation component is equal to the energy density of matter. Accounting for the evolution of the radiation component and the matter component, respectively given by $\rho_r(t_0)a^{-4}$ and $\rho_M(t_0)a^{-3}$, then a_{eq} can be expressed as the ratio between the present day energy densities of radiation and matter:

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

Only recently we should have entered an epoch dominated by a dark energy component. Fig. 1.1 shows the evolution of radiation , matter and dark energy densities with redshift. The latter, as shown in next Section, is related with the scale factor. The figure shows the radiation, matter and dark energy dominated epochs. In each epoch the evolution of the scale factor with time is driven by the dominant component. In particular $a(t) \propto t^{1/2}$ in the radiation epoch, $\propto t^{2/3}$ in the matter dominated epoch and $\propto exp(Ht)$ if dark energy dominates in form of cosmological constant.

1.1.1 Redshift-distance relation

The *proper distance* of an object at time t, $d(t)$, with respect to an hypotetical observer in an expanding homogeneus and isotropic universe is given by the integral of the space interval, i.e. the integrated spatial interval measured instantanously by a chain of observers between the central observer and the object. If the observer is at the origin of the comoving coordinates system (r , θ, ϕ) then , from eq. 1.1:

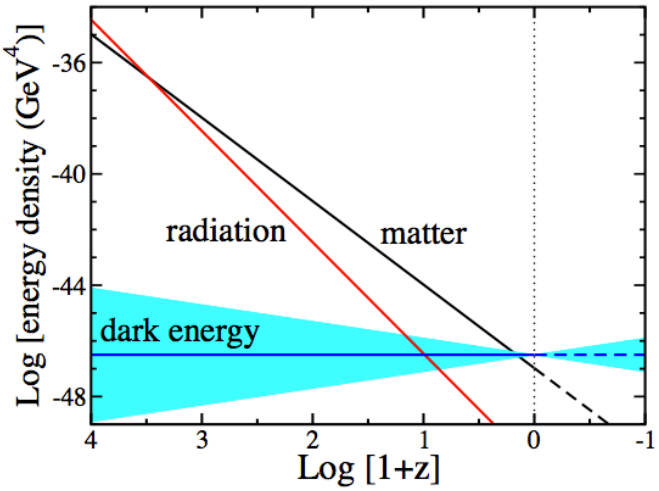
$$d(t) = a(t)r \tag{1.18}$$

where r is the comoving radial coordinate of the object. For the present day t_0, the distance is:

$$d(t_0) = r = \int_{t_e}^{t_0} \frac{cdt}{a(t)}. \tag{1.19}$$

where t_e is the emission time of the light signal received at t_0, the present time. The latter formula can be used to derive the relation between the

Figure 1.1: Evolution of radiation, matter and dark energy with redshift. The blue band represents deviation from cosmological constant state parameter $w = -1 \pm 0.2$. Plot from (Frieman, Turner, Huterer, 2008).



redshift of the light emitted by an object and the scale factor at the time of the light emission $a(t_e)$: $1 + z = 1/a(t_e)$. From this relation and Friedmann equation 1.5, the comoving distance can be written in function of redshift, Hubble parameter and density parameters:

$$d(z) = \frac{c}{H_0} \int_0^z X(z')^{-1/2} dz' \tag{1.20}$$

where

$$X(z') = \Omega_R(1+z')^4 + \Omega_M(1+z')^3 + \Omega_k(1+z')^2 + \Omega_{DE} exp\left\{-3\int_0^{z'} \frac{dz''}{z''}[1 + w_{DE}(z'')]\right\} \tag{1.21}$$

The parameter $w_{DE}(z)$ is the state parameter of the dark energy component, here assumed to be time dependent. As can be seen from equation 1.20, the relation redshift-distance depends on the density parameters and equation of state of the various components. The H_0 value sets the scale lenght. Therefore, observing d at different redshifts allows, in practice, derive all these fundamental quantities. However, the comoving distance is not an obsevable quantity since it would request symultaneous measurements. Fortunately there are two observable quantities that are related to the comoving distance: the *luminosity distance* d_L and the *angular diameter distance* D. The first is obtained by measuring the photon flux F from an

object with known luminosity L:

$$d_L(z) \equiv \sqrt{\frac{L}{4\pi F}} = S_k(d)(1+z), \tag{1.22}$$

where L is the intrinsic luminosity of the object. Analogously one can observe the *angular diameter distance* from the angular size of an object with known size D':

$$D(z) = \frac{D'}{\delta\theta} = \frac{S_k(d)}{1+z}. \tag{1.23}$$

The best known application of 1.22 is with Type Ia Supernovae, an ideal type of standar candles with very similar luminosities. They have successfully used to perform the so called "Hubble-test" that is obtained by rewriting the 1.22 in terms of magnitudes:

$$m - M = 5log_{10}(d_L[pc]/10pc) = 5log_{10}(S_k(d)(1+z)[pc]) - 5 \tag{1.24}$$

in which the apparent magnitude m of Supernovae Ia Type with known absolute magnitude M is used togheter with the observed redshift z to trace $S_k(d)$ out to large distances. This test has provided the first evidence for an accelerating Universe.

In the limit $z \to 0$:

$$d(z) \approx d_L(z) \approx D(z) \approx \frac{c}{H_0} z \tag{1.25}$$

For small redshifts , namely small distances from the observer, this simple relation is a good approximation for the current proper distance.

1.2 Density fluctuations and gravitational instability

The Cosmological Principle is valid only on large scales. Indeed, disomogeneities and anisotropic structures such as galaxies, clusters, superclusters, voids and filaments are observed on increasingly large scales as the Universe evolves. Therefore we need to extend the Standard Model, based on the the Cosmolgical Principle. As already anticipated, the structures in galaxy and matter distribution can be explained as result of growth of initial perturbations of the smooth background . The basic mechanism of this evolution is gravitational instability in density fluctuations of a self gravitating fluid beyond a characteristic scale called "Jeans lenght". This is the mechanism that determines star formation in galaxies. In the cosmological framework, the expansion of the Universe and the coupling of matter perturbations with perturbations of the other components and with perturbations of the metric complicate this picture. The theory of gravitational instability is greatly simplified when density

fluctuations are small and the equation that describe their evoloution can be linearized. This is a feasible approximation since fluctuations were indeed small at early epochs and still are on large scales. On small scales the linear approach breaks down and higher order perturbation theory or numerical simulation are mandatory. In this Section a brief summary of the basic concepts of the linear perturbation theory is given. We set $\hbar = c = k_B = 1$.

Let's linearly perturb the metric tensor with respect to FRW metric of the zero order homogeneus cosmology:

$$\begin{aligned} g_{00}(\vec{x},t) &= -1 - 2\psi(\vec{x},t) \\ g_{0i}(\vec{x},t) &= 0 \\ g_{ij}(\vec{x},t) &= a^2\delta_{ij}(1 + 2\phi(\vec{x},t)). \end{aligned} \tag{1.26}$$

The perturbations to the metric are ψ , the *Newtonian potential*, and ϕ , the *perturbation to the spatial curvature*. They are assumed to be small so that all quadratic terms are neglected and we only retain linear perturbations to the metric. Equations 1.26 contain only scalar perturbations; we focus on them because they are the ones which couple to matter perturbations, even if also vector or tensor perturbations of the metric are possible. These equation are written in conformal Newtonian gauge and this corresponds to set equal to 0 two of the four scalar degrees of freeedom, that in this way reduce to ψ and ϕ. Let us introduce the following variables: the temperature fluctuation $\Theta = \delta T/T$ of the photon fluid, decomposed in its multipole components Θ_l; the distribution function for massless neutrinos, N; the overdensity field of cold dark matter: $\delta(\vec{x},t) = \rho(\vec{x},t)/\bar{\rho}((\vec{x},t) - 1$; the peculiar velocity field of cold 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 $v_b(\vec{x},t)$. The differential equation system that describes the evolution of all these quantities is basically given by: 1) the Boltzmann equations; 2) the two independent equations, analogous to the Friedman equation, that can be derived from Einstein's field equations expressed in terms of the perturbed metrics 1.26.

Let us focus on dark matter perturbations. In this case one can simplify the system of equations described above thanks to the following considerations: the contribution of baryonic matter to fluctuations of the metric can be neglected because baryon mass density is negligible respect to that of the dark matter. The contribution of photons to metric fluctuations is taken into account only in the radiation dominated epoch. During this epoch the photons are tightly coupled with baryons, and the scattering Compton is so efficient that the quadrupole and higher order moments of distributions of photons are neglected. Neglecting the higher radiation moments implies that $\psi = -\phi$. Under the above approximations and neglecting all terms that are quadratic in the fluctuations, the set of differential equations to be solved in order to obtain the evolution of dark matter density perturbation

is given by:

$$\dot{\Theta}_{r,0} + k\Theta_{r,1} = -\dot{\phi} \tag{1.27}$$

$$\dot{\Theta}_{r,1} + \frac{k}{3}\Theta_{r,0} = -\frac{k}{3}\phi \tag{1.28}$$

$$\dot{\delta} + ikv = -3\dot{\phi} \tag{1.29}$$

$$\dot{v} + \frac{\dot{a}}{a}v = ik\phi \tag{1.30}$$

$$k^2\phi + 3\frac{\dot{a}}{a}(\dot{\phi} + \phi\frac{\dot{a}}{a}) = 4\pi G a^2[\rho_c\delta + 4\rho_r\Theta_{r,0}] \tag{1.31}$$

In these equations the apex r indicates generically radiation (photons and neutrinos), 0 and 1 are the multipole indices l of the temperature distribution of radiation. These equations are defined in Fourier space: in linear theory, the Fourier transform produces a set of independent ordinary differential equations that can be solved for each $k-$mode indipendently. In this Chapter the Fourier transform convention is:

$$\tilde{\Theta}(\vec{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\vec{k}\cdot\vec{x}}\Theta(\vec{k}) \tag{1.32}$$

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

$$\eta \equiv \int_0^t \frac{dt'}{a(t')}. \tag{1.33}$$

η is used as evolution variable, analogous to the scale factor a, the redshift z or the proper time t. It coincides (if c is set to 1) with the *comoving horizon*, namely the comoving distance that a photon has traveled from the Big Bang to the epoch t. Regions separated by distances greater than η can not be causally connected. The derivatives in 1.27 have to be estimated with respect to the conformal time.

Now it is possible to solve this set of equations for each comoving wave number $\vec{k}$. A simple qualitative description of the analitic solutions for the epochs in which one of the energy density components dominates and for different comobile scale ranges is given in the next Subsection.

1.2.1 Linear theory solutions

The comoving horizon increases with time incorporating larger and larger comoving scales. Let's examine first the wavenumbers k which enter the horizon in the matter dominated epoch, well after the matter-radiation equivalence. These large scale modes lie beyond the horizon at early times, when radiation dominates the Universe energy density content. As long as these modes are outside the horizon, the gravitational potential ϕ, on the

corresponding scale, is approximatively constant and density perturbations of dark matter δ are constant too. These modes enter the horizon Universe during the matter dominated epoch. During this epoch, when density fluctuations are small and the Universe is approximatively Einstein-de Sitter, namely a flat, matter only Universe, the time derivative of the gravitational potential is zero and the density fluctuations within the horizon has a simple dependence on the scale k and the time variable a:

$$\delta(\vec{k}, a) = \frac{k^2 \phi(\vec{k}, a) a}{(3/2)\Omega_M H_0^2}. \tag{1.34}$$

$d\phi/dt = 0$ implies that the density perturbations grow proportionally to the scale factor a during the matter domination epoch. This is not true at later times when dark energy dominates, and the amplitude of the gravitational potential decreases, $d\phi/dt < 0$.

For the smaller scales that have entered the horizon earlier on in the radiation dominated epoch, the solution to the equations 1.27-1.31 for dark matter perturbations is:

$$\delta(\vec{k}, \eta) \propto ln(\vec{k}, \eta). \tag{1.35}$$

This means that, when radiation dominates, dark matter perturbations grow logaritmically with the scale factor a after the mode has entered into the horizon since η coincides with a (from eq. 1.33 and the relation $a(t) \propto t^{1/2}$ valid during the radiation dominated epoch). The growth is not fast as during the matter era (where $\delta \propto a$) due to the radiation pressure. Later on, when matter starts dominating, perturbations grow faster. The equation which governs the dark matter perturbations after radiation perturbations have become negligible can be derived from equations 1.27–1.31 and has two independent solutions: a growing solution $D_1(\eta)$ (called *growth function*) and a decaying mode $D_2(\eta)$. The general solution for dark matter perturbations is:

$$\delta(\vec{k}, \eta) = C_1 D_1(\eta) + C_2 D_2(\eta) \tag{1.36}$$

where C_1 and C_2 depend on $\vec{k}$ and are set by the initial conditions. At late times, in the matter dominated era, the growing solution scales with a and the decaying mode vanishes.

To summarize: large scales perturbations in the dark matter component which enter the causal horizon at relatively late times, during the matter dominated epoch, grow proportionally to the scale factor a after the crossing. Perturbations on smaller scales, which cross the horizon earlier during the radiation dominated epoch, grow less prominently (only logaritmically) during the radiation era. At later times when matter dominates, this perturbations scale as a, namely in the same way as perturbations on larger scales. Consequently, the evolution of the matter perturbations on all the

modes of cosmological interest becomes scale independent at late times and can be described by a time dependent factor only, namely the growth factor D_1. Finally, at later times ($a > 0.1$ and $z < 10$), when the dark energy dominates the energy budget and the Universe accelerates its expansion, the growth of fluctuations slows down, the quantitative estimate of the effect depending on the equation of state of dark energy.

1.2.2 The linear growth factor and the growth rate

In the matter dominated epoch the evolution of matter perturbations can be described by a scale independent growth factor D_1. In linear regime an explicit expression for D_1 in function of the scale factor a can be derived from the equations of the evolution of the matter density fluctuations also allowing for the possibility of energy other than matter or radiation:

$$D_1(a) = \frac{5\Omega_M}{2}\frac{H(a)}{H_0}\int_0^a \frac{da'}{(a'H(a')/H_0)^3}, \tag{1.37}$$

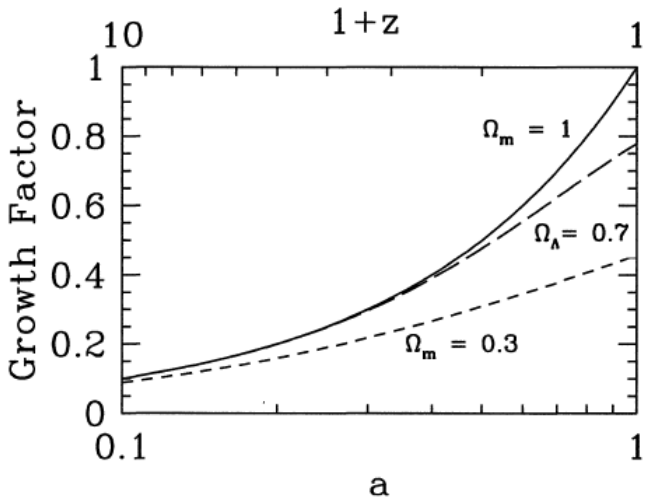
in which the dependence on Ω_M and $H(a)/H_0$ is explicit. The dependence of the growth factor from these cosmological parameters is illustrated in fig. 1.2 for three different models. For flat models that include a dark energy component the behaviour matches that of the flat and matter-only case ($\Omega_M = 1$) out to $z \sim 2$, when the cosmological constant starts outmatching to the energy budget in a significant way. After $z = 2$, the growth is suppressed. A suppression over all cosmic times characterizes the open Universe case (short dashed curve). The growth factor can be used to constrain dark energy models. Different types of dark energy would produce different growth history, i.e. different curves in the plot. This means that an accurate estimate of the growth factor at different epochs can constrain the amount of dark matter, dark energy and its equation of state thanks to the dependence of D_1 on $H(a)$.

An even more effective constrain can be obtained by measuring the time derivative of the growth factor, the so called *growth rate*:

$$f \equiv \frac{dlnD_1}{dlna} = \frac{a}{D_1}\frac{dD_1}{da}. \tag{1.38}$$

The dependence of f from the cosmological parameters is even more transparent. An accurate approximation to the growth rate is given by the following expression: $f = \Omega_M(z)^\gamma$, where the exponent γ depends on the type of dark energy (for ΛCDM model $\gamma \sim 0.55$) and also on the underlying gravity model. This means that measuring f allows, in principle, not only to constrain the nature of the dark energy but also to detect possible deviations from the General Relativity predictions on cosmological scales. This in fact justifies the widespead interest in measuring this quantity from the real data which constitutes one of the main point of this thesis.

Figure 1.2: The black continous curve represents the growth factor for a flat universe without cosmological constant. The intermediate curve is for ΛCDM universe. The short dashed curve represents the growth factor for an open universe. Plot from Dodelson, Modern Cosmology, Academic Press, 2003.



1.2.3 The Matter Power Spectrum in the linear regime

We have seen how, given a set of initial conditions for the density, velocity and metric perturbations, it's possible in principle to obtain analytical solutions of the differential equations which govern their evolution if the linear approximation is valid. In this regime each k-mode of a Fourier expansion of all the fields (density, velocity, potential, etc.) evolves indipendently, i.e constitutes an independent solution to the set of ordinary differential equations 1.27 - 1.31. However, it is neither possible not particularly informative to guess the correct initial condition and provide the precise value of the fluctuation at a specified position and time. Cosmologically interesting information can instead be extracted from the statistical properties of fluctuation's fields and their spatial distributions. Let us focus on density fluctuations. The cosmological model assumes that primordial fluctuations were originally set during the *inflation* era. The cosmological inflationary theory states that the Universe has undergone, in the very early epochs, a period of accelerated expansion, following which the Universe appears highly homogeneus and flat on sub-horizon scales. The residual fluctuations that are the seeds to present day cosmic structures, are quantuum fluctuations, originally produced at microscopic scales and then upgraded to cosmological scales after the inflationary expansion. A variety of inflationary models predict that primordial perturbations obey to Gaussian statistics. Deviations from Gaussianity are also possible, but

constrained to be very small by Cosmic Microwave Background observations (Komatsu et al. 2011). The n-point statistics are useful to characterize the spatial properties of the density perturbations field. Let's focus on the 2-point statistics of matter density fluctuations, namely the *2-point correlation function* defined as an ensemble average over all pairs of points separated by a distance $\vec{r}$:

$$\langle \delta(\vec{x})\delta^*(\vec{x}+\vec{r})\rangle = \xi(\vec{r}) \tag{1.39}$$

and its Fourier transform, the *power spectrum*:

$$\langle \delta(\vec{k})\delta^*(\vec{k'})\rangle = (2\pi)^3 P(\vec{k})\delta_D{}^3(\vec{k}-\vec{k'}). \tag{1.40}$$

If the fluctuations are Gaussian, the ensemble average and 2-point statistics are sufficient to completely characterize the spatial distribution in a statistical sense. This thesis focuses on the power spectrum, which represents the variance (i.e. 2-point ensemble average) of the distribution in Fourier space. Linear perturbation theory plus initial conditions from inflation predict the shape of the power spectrum of density fluctuations simply referred to as *matter power spectrum*, at each time. In practical applications, these predictions are only valid for a limited range of scales in which density perturbations are still small and linearly evolving. Because of the growth of density fluctuations this range reduces with time.

The assumption of isotropy on large scales implies that the matter power spectrum is independent from the direction of the $\vec{k}$ vector. Therefore, under this assumption, the power spectrum can be expressed as function of the module of $\vec{k}$, k.

Equation 1.34 relates the dark matter density fluctuation field δ to the metric peturbations ϕ well after the matter-radiation equivalence, a_{eq}. For a given choice of cosmological parameters this equation allows to obtain an analytical expression for the matter power spectrum. We have seen that for $a >> a_{eq}$, the evolution of perturbations is independent of k so let us define a_{late} as the epoch in which all the Fourier modes of interest start evolving in the same way. The gravitational potential $\phi(\vec{k}, a)$, for each time and mode and $a > a_{late}$, can be written as the product of the primordial power spectrum of the gravitational potential at the end of the inflationary epoch, $\phi_P(\vec{k})$, and a transfer function $T(k)$, which modulates the primordial power spectrum, and the growth function $GF(a)$, which accounts for the late time evolution ($a > a_{late}$), i.e.:

$$\phi(\vec{k}, a) = \phi_P(\vec{k}) \times T(k) \times GF(a) = \frac{9}{10}\phi_P T(k)\frac{D_1(a)}{a}. \tag{1.41}$$

The last equality follows from the fact that in the linear regime for the gravitational potential $GF(a) = D_1(a)/a$. In the matter dominated era $D_1 \propto a$ and GF remains constant.

The transfer function $T(k)$ is set equal to 1 for the largest scales that never cross the horizon, and the additional factor 9/10 accounts for the slight overall decrease of the potential respect to the primordial value. Substituting eq.1.41 in eq. 1.34, we obtain:

$$\delta(\vec{k},a) = \frac{3}{5}\frac{k^2}{\Omega_M H_0^2}\phi_P(\vec{k})T(k)D_1(a). \tag{1.42}$$

The primordial power spectrum is conveniently expressed to highlight possible deviations from *scale-invariance* that implies $k^3 P_\phi(k) = constant$, i.e.:

$$P_\phi(k) \propto k^{-3}\left(\frac{k}{H_0}\right)^{n_s-1}(\Omega_M/D_1(a=1))^2. \tag{1.43}$$

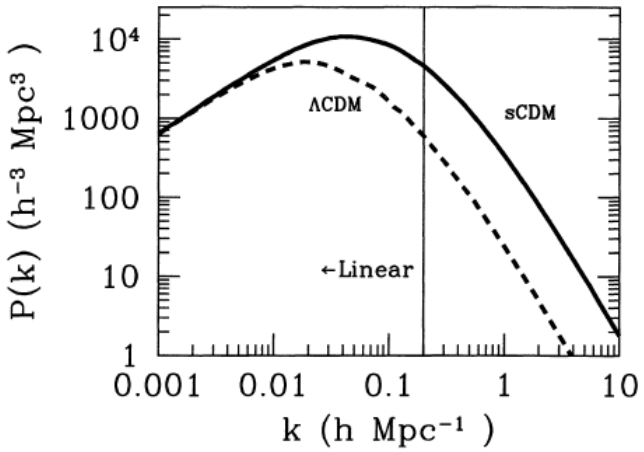
where n_s is the *scalar spectral index*. A scale invariant power spectrum is characterized by $n_s = 1$. Given this expression for the primordial power spectrum for the potential, the power spectrum of matter density fluctuations at late times is:

$$P(k,a) \propto \frac{k^{n_s}}{H_0^{n_s+3}}T^2(k)\left(\frac{D_1(a)}{D_1(a=1)}\right)^2. \tag{1.44}$$

This equation shows that the linear power spectrum of the matter, as we shall from now on refer to it, at $a = a_{late}$, is proportional to $k^{n_s}T^2(k)$, and the growth factor D_1 determines its evolution at later times $a > a_{late}$. Fig. 1.3 shows the linear matter power spectrum at present time ($z = 0$ or $a = 1$) for two different cosmological models. The shape of the power spectum at large scales, is proportional to k^{n_s} as expected for modes that cross the horizon after the radiation-matter equivalence where $T(k) \equiv 1$. On smaller scales, the power spectrum is increasingly suppressed respect to the primordial one k^{n_s}. These scales cross the horizon when radiation still dominates slowing down the growth of the dark matter perturbations.This leads to a turn over in the power spectrum at the scale that enters the horizon at radiation-matter equivalence, k_{eq}. Fig. 1.3 shows also the approximate validity limit of linear theory at $z = 0$ (vertical line).

So far we have focused on dark matter density perturbations, neglecting the contribution of the baryons to the total matter perturbations. Even if the leading role in the total matter power spectrum is played by dark matter, the contribution of baryons alterates the total transfer function shape. The baryon's evolution is different from the dark matter one: they are coupled with photons untill recombination and then turn free to cluster in the dark matter potential wells. The perturbations of the photo-baryon plasma don't grow appreciably with time, but oscillate under the combined effect of radiation pressure and gravitational collapse. Slightly after the recombination, these sound waves freeze, leading to intermediate scales oscillations in the power spectrum (referred to as BAO, acronymous for

Figure 1.3: Power spectrum in two cosmological models: ΛCDM universe and without cosmological constant. They are normalized to agree on large scales. The veritical line evidences the linear scales. Plot from Dodelson, Modern Cosmology, Academic Press, 2003.



baryonic acoustic oscillations) and to an overall suppression of power on intermediate and small scales. The physical scale of the BAO is known and can be used as a standard ruler. The BAO signature has now been detected in the power spectrum and correlation function of galaxies (Cole et al. 2005, Eisenstein et al. 2005), and is now routinely used as a observational test to trace the expansion history of the Universe.

Dependence on cosmological parameters -

Here's a summary of how the cosmological parameters determinate the shape and the amplitude of the linear total matter power spectrum:

1. A_s: represents the amplitude of the primordial curvature perturbations power, namely the amplitude of the primordial power spectrum of the gravitational potential, linked to the matter density fluctuations by 1.42.

2. n_s: the scalar spectral index, it uniquely determines the shape of the primordial power spectrum, which should be preserved on large scales.

3. Ω_M: the dark matter density parameter. Togheter with the radiation density parameter it sets the value of the scale factor at radiation-matter equivalence : $a_{eq} = \rho_R/\rho_M$, namely the scale where the $P(k)$ turns over. It affects also the amplitude of the power spectrum.

4. Ω_{DE}: the dark energy density parameter. Dark energy affects the growth of the perturbations at very late times, contributing at the suppression of the growth and the precise impact depends on the

specific type of dark energy. Another indirect effect of dark energy is that for a flat universe, the higher Ω_{DE} the smaller Ω_M, with a corresponding shift in the turnover scale.

5. Ω_b: the presence of baryons also suppress power on small scales. Indeed, the baryon overdensities are suppressed or frozen until decoupling. After decoupling the baryons fall into the potential wells associated to dark matter fluctuations and then share their growth history. Increasing the baryon fraction with respect to the total mass density reduces the depth of dark matter potential wells, effectively reducing the power in the corresponding scales. In addition baryons impact their BAO feature into the transfer function. Finally photon diffusion damps baryonic oscillations, leaving yet another characteristic signature on the transfer function.

6. σ_8: it is the *rms* fluctuation in total matter in spheres of 8 Mpc/h at z=0, computed in linear theory. It's given by:

$$\sigma_R^2 = \int \frac{dk}{2\pi^2} k^2 P(k) [\frac{3j_1(kR)}{kR}]^2 \tag{1.45}$$

where R =8 Mpc/h and j_1 is the spherical Bessel function of order 1. σ_8 measures the normalization of the power spectrum , and is a derived parameter. σ_8^2 and A_s are related by a proportionality law. The proportionality constant depends on the cosmological model.

7. Ω_ν and N_ν: the mass density in neutrinos and the number of neutrinos fractions modify the value of a_{eq} and further suppress power on scales smaller than that of the turnover.

1.3 The Galaxy Power Spectrum

So far we have seen how the matter distribution on large scale can be characterized in a statistical way and how perturbation theory describes the relevant 2-point statistics in the limit of the linear approximation.

What type of dataset can be used to assess these predictions? 2-point statistics, like 2-point correlation function and power spectrum, can be measured for a discrete sample of objects with known positions and this suggests that we could use galaxy redshift surveys to map the position of extragalactic objects in space, compute their 2-point statistics and compare it with theoretical predictions. Is this comparison meaningful? Let us postpone the potential problem of using galaxy redshifts as distance proxy in Section 1.3.3 and let us focus on the adequacy of galaxies as tracers for the distribution of the (mostly dark) matter. Galaxies are thought to form within gravitational potential wells determined by the dark matter

component and, more specifically, within virialized structures, called *dark matter haloes*. The physical processes which govern the evolution of both these dark matter structures and of the galaxies are very complex. The main issues are two. First of all, dark matter haloes where galaxies reside are nonlinear structures, therefore their evolution and properties can not be simply predicted using linear theory. Describing the growth of density fluctuations in the non linear regime is the first problem to solve to analyze the spatial distribution of galaxies down to small, galaxy-like scales. The second issue is the physics of the baryons and the stellar feedback process that greatly complicate our understanding of how galaxies form and evolve. Both effects are discussed in the next Sections.

1.3.1 Galaxy Bias

Since one typically observes the spatial distribution of galaxies, not mass, some independent estimate or theoretical insight of the mapping from one to the other is mandatory. This mapping, which is commonly referred to as *galaxy bias*, parametrises our ignorance on the physics of galaxy formation and evolution and represents one of the main source of uncertainties in the study of the large scale structure of the universe. Quantifying the galaxy bias is not just necessary to obtain unbiased cosmological information. It also allows to discriminate among competing models of galaxy formation and the physical processes that regulate the evolution of stars and galaxies. Assuming that the bias relation between the galaxy density field and the mass field is local, namely the galaxy density is uniquely determined by the underlying mass density, and linear, i.e. $\delta_g = b\delta_M$, then the galaxy power spectrum and the linear matter power spectrum can be related by a simple direct proportionality law:

$$P_g(k) = b^2 P(k) \tag{1.46}$$

where b is referred to as *linear galaxy bias*. On scales much larger than those affected by galaxy formation processes the linear bias relation is likely to be valid and fluctuations are still in the linear regime, so that the galaxy power spectrum is well approximated by $b^2 P_{lin}(k)$.

This relation, commonly adopoted in the literature, is scale independent and supported by cosmological observations, but valid only on large scales. A step towards more realistic bias model is represented by the *Halo Model* (HM) which provides a theoretical scheme to describe the clustering of dark matter, haloes and galaxies in both the linear and the non linear scales (Cooray et al., 2002 for a review). According to the HM prescriptions, galaxies can form only in the potential wells of virialized dark matter objects, the haloes, that are formed via gravitational collapse in correspondence of the density peaks of the density field. The galaxy properties are strictly related to the mass and angular momentum of the parent halo in which the

baryonic matter collapse forming galaxies. In the HM framework the galaxy power spectrum is given by the sum of two terms: the 1 halo term , which is the power relative to the structure whithin a single halo, and the 2 halo term relative to the correlation between haloes, which dominates on large scales:

$$P_g(k) = P_g^{1h}(k) + P_g^{2h}(k) \tag{1.47}$$

In the large scales limit, the galaxy power spectrum is related to the linear power spectrum of matter by a 1.46-like formula where b is now given by:

$$b(z, m_g, L_g) = \int n(z, m) b_{1,halo}(z, m) \frac{< N_g|m >}{\bar{n}_g} dm. \tag{1.48}$$

In this formula, m is the halo mass, $b_{1,halo}(z, m)$ is the linear bias of the haloes with mass m at redshift z, while $n(z, m)$ is the halo mass function which gives the comoving number density of haloes with mass m at redshift z. Halo mass function and halo bias can be derived assuming a dynamical model for the halo formation in a given cosmological framework. A simple model in which one assumes that haloes form through spherical collapse (Press, Schechter, 1974) has been later improved , providing a better fit to the numerical simulations of gravitational clustering (Sheth & Tormen, 1999). According to this models haloes are generally biased respect to dark matter because they form in the overdensity peaks of dark matter density field and bias grows with the halo mass. Once the halo bias and halo mass function are specified, the bias relation between galaxies and mass can be obtained assuming a halo occupation number that specifies the number of galaxies that populate a halo of a given mass. Its mean value, $< N_g|m >$, represents the mean number of galaxies N_g in a halo of mass m. Different prescriptions for this conditional probability can be found in literature(e.g. Jing, Mo & Borner, 1998; Sheth & Diaferio, 2001;Tinker et al. 2005; Zheng et al. 2005). This probability generally depends on halo mass, redshift, type of galaxy, luminosity and mass of the galaxies. The integral on all halo masses in eq. 1.48 gives the linear galaxy bias. 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). At the same time, galaxies with early morphological type, namely elliptical galaxies in the Hubble morphological classification, preferentially populate the massive galaxy clusters, whereas spiral galaxies are commonly found in lower density environments. Similar differences in clustering are expected for galaxies characterized by different colours. The dipendence of the bias from the redshift is physically motivated by the fact that, in the past, when galaxies started forming , they populated the highest density peaks that are statistically more clustered.

Current limitations in the theoretical models of galaxy evolution and in the HM do not allow to predict galaxy bias with sufficient accuracy in order

to constrain cosmological models so in general, the galaxy bias and non linearities of the bias (see Subsection 1.3.2 and Chapter2) are measured directly from the data, like the power spectrum. The linear bias b is generally set as free parameter in the galaxy power spectrum modellizations. This approach in principle introduces degeneracies with other cosmological parameters that can be broken through combinations with independent experiments or with sufficient accurate future galaxy redshift surveys.

1.3.2 Non linearities

All non linearities are very tricky to predict theoretically. Going back to the matter power spectrum , there's a variety of approaches to the non linear evolution of density fluctuations.

The simplest analytical approach consists in following the evolution of an overdensity with spherical form, constant density inside it, initially small $\delta_i << 1$ and with an initial peculiar velocity on its edges equal to zero. Assuming an Einstein-de Sitter background Universe, at early times the density perturbation increases with the scale factor a, according to linear theory, but at some point the sphere breaks away from the Universe expansion and begin to collapse reaching a final virialized configuration. This simple model is useful to understand the non linear evolution of structures but relies on not realistic assumptions, as the more realistic models of structure formation reveal.

Equations of density and velocity fluctuations 1.27-1.31 can be analitically solved to some higher order perturbative terms. The solution works quite well in the middly non-linear regime (Jeong & Komatsu, 2006, 2009) but not on smaller scales. Another parallel analytical approach is represented by the *Zel'dovich approximation*, (1970): it consists in following the trajectories of a particle distribution. The Eulerian coordinate $\vec{x}$ is given for each time η by the sum the initial unperturbed (Lagrangian) coordinate $\vec{q}$ and a displacement field $\psi(\vec{q}, \eta)$. The particles are assumed to move in their initial direction. The *Zel'dovich approximation* is described as a first order Lagrangian perturbation theory (alternative to the Eulerian perturbation theory which follows the evolution of density and velocity fluctuations). For small displacements, this approximation recovers the Eulerian linear regime, briefly discussed in the section 1.2. The *Zel'dovich approximation* consists in a purely kinematical approach. It doesn't account for close-range forces as well as pressure or shocks and breaks down in the strongly non linear regime.

Different approaches to non linearities come from the use of fits to N-body simulations, among them Hamilton et al. (1991), Peacock & Dodds (1996), Smith et al. (2003). The latter work consists in an improved fitting formula based on the Halo Model, in which a code aimed to compute the non linear matter power spectrum, *Halofit*, is provided. The Halo Model itself,

introduced in Subsection 1.3.1, provides a physical motivation to the linear and non linear clustering of mass, haloes and galaxies but doesn't match perfectly N-body data. Particularly tricky is to predict galaxy non linearities of the bias, both in real space and in redshift space. As already anticipated, the approach to non linearities in the data is typically phenomenological: empirical formulas, based on some physical motivation, like from the HM or simulations, are very often used in literature to model non linearities, for example in the power spectra analysis of the galaxy surveys. The introduction in these formulas of a small number of free parameters, on which in general one marginalizes over, permits to recover the cosmological parameters of interest with a good accuracy (Cole et al. 2005). At the same time, they can be used to estimate the non linearities of the bias themselves from the data, if applied to the future galaxy redshift surveys. We postpone the discussion about non linear models of bias to the next Chapter.

1.3.3 Redshift space

Another complication which must be taken into account if one wants to extract unbiased cosmological information from the galaxy power spectrum and that, at the same time, introduces extra cosmological information respect to the one contained in the linear matter power spectrum is related to the use of the *observed redshift* as distance indicator.

As seen in the Subsection 1.1.1, the comoving distance is linked to the redshift by a model dependent relation. As a consequence the mapping from the measured redshift to the comoving distance d is generally model dependent, and the choice of the model affects the clustering measure. We don't focus on this kind of problem because of the very local nature of the galaxy survey under study, the 2MRS redshift survey. In the limit of small redshifts , equation 1.25 shows that the redshift distance mapping consists in a simple direct proportionality law.

The observed redshift, z_{obs}, doesn't coincide in general with the cosmological redshift, z_{cos}, determined by the expansion of the Universe as if the galaxy was fixed in its comobile radial position.Each galaxy experiences a peculiar motion which contributes to the observed redshift with a Doppler effect due to the radial component of its peculiar velocity vector, $\vec{v}$. A galaxy's velocity, $\vec{v}_{obs}$, is in general given by:

$$\vec{v}_{obs} = \vec{v} + \hat{r} v_{cos} \tag{1.49}$$

where v_{cos} is the expansion velocity , and $\hat{r}$ the radial versor. Focusing on the low redshift limit we have:

$$s_{obs} \equiv c z_{obs} = c z_{cos} + \hat{r} \cdot \vec{v} = H_0 d + v \tag{1.50}$$

where d is the comoving distance and v is the component of the peculiar velocity in the radial direction. When mapping the radial positions from

redshift one in general ignores the peculiar velocity of the object, introducing an error in the radial positioning. The expression *redshift space* means this type of mapping and corresponds to assign distances to galaxies equal to:

$$d_s = \frac{s_{obs}}{H_0} = \frac{cz_{obs}}{H_0}. \tag{1.51}$$

The redshift space mapping constitutes a problem, inducing a bias in the power spectrum measure, but also an opportunity because it contains informations about the growth of the cosmological structures. The peculiar motions of the galaxies are in fact due to the gravitational collapse due to the surrounding density field, so are stricly related with the growth of the perturbations. Focusing on linear perturbation theory, we have seen that, at late times on scales well within the horizon, the gravitational potential is constant with time while the dark matter density fluctuations grow as the growth factor D_1. The continuity equation for dark matter perturbations 1.29 reduces to:

$$\dot{\delta} + ikv = 0. \tag{1.52}$$

So, the peculiar velocity field can be written in function of δ and the growth factor as:

$$v(k,\eta) = \frac{i}{k}\frac{d}{d\eta}\left[\frac{\delta}{D_1}D_1\right] = \frac{i\delta(k,\eta)}{kD_1}\frac{dD_1}{d\eta}. \tag{1.53}$$

In the two previous equations and in Section 1.2 v is written for simplicity without the vector symbol, but it is the Fourier component of the velocity parallel to $\vec{k}$. In terms of the linear growth rate f (eq. 1.38) and scale factor a, the velocity field, is given by:

$$\vec{v}(\vec{k},a) = ifaH\delta(\vec{k},a)\frac{\vec{k}}{k^2}. \tag{1.54}$$

Here the vector symbol is explicitly indicated. This equation shows the relation between peculiar velocities and density field valid in linear theory.

The result of using the observed redshift as distance is that the resulting distribution of objects appears distorted along the line of sight and this induces anisotropy in the correlation function and power spectrum $P(\vec{k})$. Using the linear theory approximation it is possible to quantify this distortion, namely relate the underlying linear power spectrum in real space, which is isotropic under the Cosmological Principle assumption, to the redshift space anisotropic power spectrum. In order to do this, the distant observer approximation is needed: it assumes that the galaxies are so distant from the observer that, dubbed $\hat{z}$ the versor pointing along the line of sight from the observer (pointing to the centre of the galaxy sample under study), and $\hat{r}$ the versor pointing to a galaxy, the projection of the peculiar velocity vector of that galaxy along $\hat{r}$, $\hat{r}\cdot v$, can be approximated by the projection of the peculiar velocity vector along the line of sight, $\hat{z}\cdot v$. In practice the

galaxies under study must be sufficiently near to each other. Assuming the distant observer approximation and linear theory, the relation between linear power spectrum in real space and galaxy power spectrum in redshift space, first derived by Kaiser (1987), is given by:

$$P_g(\vec{k}) = b^2 P_{lin}(k)[1 + \beta\mu^2]^2 \tag{1.55}$$

where μ is the cosine of the angle between the vector $\vec{k}$ and the line of sight, and the parameter β, dubbed *linear redshift distortions parameter*, is given by:

$$\beta = \frac{f}{b} \approx \frac{\Omega_M^{0.55}(z)}{b}. \tag{1.56}$$

The linear bias parameter has been introduced in equation 1.55 in order to account of the bias between galaxies and matter. Equation 1.55 shows that the power spectrum is boosted in the direction of the line of sight, because in that direction the peculiar velocities squash virtually the structures enancing the power. The direction perpendicular to the line of sight is unaffected by peculiar velocities, because the distorsion affects only the radial coordinate that in the distant observer approximation coincides with the direction parallel to line of sight. Let decompose the k vector in the parallel and perpendicular directions respect to the line of sight: $\vec{k} \equiv (k_{//}, k_\perp)$. Fig. 1.4 shows the contour levels of the galaxy power spectrum in the plane ($k_{//}$, $k_\perp$) and shows that , for small k's, namely large scales, where linear approximation is valid, the power spectrum is enhanced in the parallel direction. The spherical average in k-space of the galaxy power spectrum 1.55, namely the monopole of the Legendre decomposition of $P_g(\vec{k})$, is given by:

$$P_g^{(0)}(k) = b^2 P_{lin}(k) \left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right). \tag{1.57}$$

The monopole is boosted of the quantity:

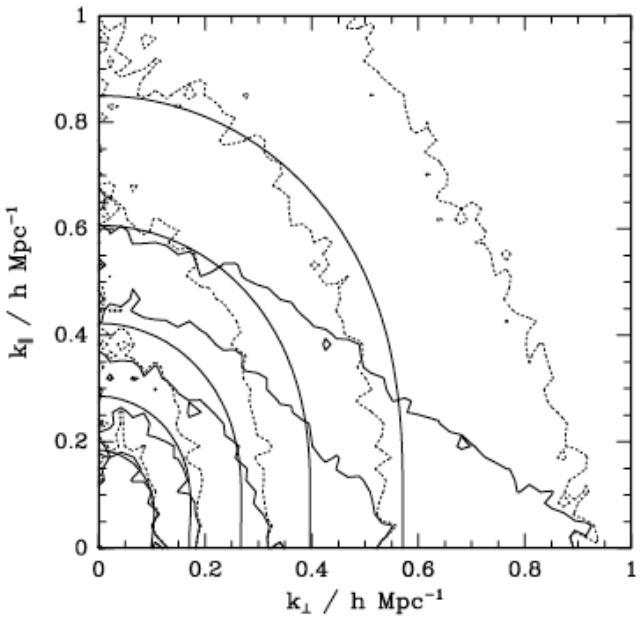
$$\left(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2\right) \tag{1.58}$$

dubbed *Kaiser boost*.

All these relations are valid in the linear regime. On smaller scales the galaxies are collapsing or are part of virialized structures, like the galaxy clusters, and the velocities are random (in case of virialized velocities) and so large that the structures appear elongated in redshift space. So, galaxy power is damped on small scales due to this random motions. This elongation along the line of sight is referred to as *Fingers of God* or simply FoG. The FoG are generally accounted for by empirical damping formulas (e.g. Hamilton, 1997; Hamilton et al. 2001; Blake et al. 2011), that are the Fourier transform of the line of sight component of a random isotropic pairwise peculiar velocities distribution function. The latter is generally

assumed to be an exponential or gaussian in v, characterized by a velocity dispersion parameter σ that is generally let free and marginalized over when one tries to constrain cosmology from data.

Figure 1.4: Contour plots of the galaxy power spectrum in the plane ($k_{//}$, $k_{\perp}$). The light solid contours are for the linear theory redshift space power spectrum: $P_g(\vec{k}) = P_{lin}(k)[1 + \beta\mu^2]^2$; increase of power on large scales along the parallel direction is shown. Heavy solid contours are from N body simulations: the results agree with linear theory for large scales, on smaller scales power is damped along the parallel direction due to FoG. Dotted contours are from the same N-body simulation but removing the FoG: this remotion restores power on small scales. Plot from Cole, 1993.



Besides the random motions due to the virialized structures, random errors on the redshift measures affect the power spectrum on small scales. They are generally taken into account in the same way as FoG.

Chapter 2

The Galaxy Power Spectrum estimate

As seen in the previous Chapter, the power spectrum is the variance of the density fluctuation field in Fourier space, and corresponds to the Fourier transform of its 2 point correlation function. In this Chapter we discuss how we compute the power spectrum of a continous density field of a distribution of discrete tracers. As first step we consider the simple case in which the density field is defined in a cubic box with constant mean number density, and show how an efficient estimator that exploits the Fast Fourier Transform can be implemented. Then we focus on the more realistic case of galaxy redshift surveys characterized by complex geometries and a mean number density of objects that varies in space. The estimator that we have considered and extensively applied throughout this thesis is the one originally proposed by Feldmann, Kaiser and Peacock in 1994. We describe it in detail. Finally we shall discuss the theoretical model of galaxy power spectra from simple linear prescription to nonlinear models for bias and the growth of density fluctuations.

2.1 Estimating the power spectrum

Let's consider the simplified case in which the density field is defined in a cubic box and the mean number density is constant within the all box.

Galaxies are discrete tracers and are assumed to sample the density field following a Poissonian distribution. The galaxy fluctuation field is defined in this way:

$$F(\vec{r}) \equiv \frac{n_g(\vec{r}) - \bar{n}}{\bar{n}} \tag{2.1}$$

where $n_g(\vec{r}) = \sum_i \delta(\vec{r} - \vec{r}_i)$ is the number density of galaxies, with $\vec{r}_i$ being the position of the ith galaxy. $\bar{n}$ is the mean number density of galaxies, constant in this case. In this case, a part from a shot noise term, dubbed

P_{shot} and described below, the ensamble average of the squared modulus of the galaxy fluctuation distribution (in Fourier space) coincides with the true galaxy power spectrum $P(\vec{k})$:

$$< |F(\vec{k})|^2 >= P(\vec{k}) + P_{shot} \tag{2.2}$$

This is true only under the simplified hypotesis of constant mean number density and cubic samples. The shot noise term, P_{shot}, is due to the discrete nature of the density field tracers, like galaxies, and would be measured even if the galaxy density were constant ($P(\vec{k}) = 0$). Since galaxies are Poisson samplers of the underlying density field and if the mean density is constant then the shot noise contribution to power is: $P_{shot} = 1/\bar{n}$.

The estimated power spectrum is then given by:

$$\hat{P}(\vec{k}) = |F(\vec{k})|^2 - \frac{1}{\bar{n}} \tag{2.3}$$

where $\hat{P}(\vec{k})$ indicates the estimator of the power spectrum $P(\vec{k})$. The shot noise power is opportunely subtracted to the measured $|F(\vec{k})|^2$ in order to obtain an unbiased estimation of the power spectrum.

The power spectrum $P(\vec{k})$ is a scalar function of the wavenumber vector. As we shall see it may be convenient to express this quantity as a function of $k_\perp$ and $k_{//}$, the components of $\vec{k}$ across and along the line of sight. In many applications, however, one can assume that galaxy clustering is homogeneus and isotropic and therefore is common to express the monopole of $P(\vec{k})$ i.e. to express the power spectrum as a function of $|\vec{k}|$.

In order to estimate the monopole of the density field the power spectrum, $\hat{P}(\vec{k})$, is averaged on spherical shells in $k-$space:

$$\hat{P}(k) \equiv \frac{1}{V_k} \int_{V_k} d^3k' \hat{P}(\vec{k}'). \tag{2.4}$$

The main advantage of considering cubic samples is that one can decompose the density field in Fourier modes and compute the power spectrum directly from the Fourier coefficients. The latter can be efficiently obtained through Fast Fourier Transform algorithms that make possible to handle very large samples of objects (Press, 1992). This is what makes Fourier-based scheme so attractive even when the geometry of the sample is not cubic.

The practical steps to estimate the power spectrum are tipically the following ones:

1. interpolation of the discrete galaxy distribution to an overdensity field $F(\vec{r})$ defined onto a regular grid. Different interpolation schemes can be used for this purpose;

2. Fourier transform of $F(\vec{r})$ through an FFT algorithm and estimate of the Fourier coefficients of the field $F(\vec{k})$;

3. computation of the squared modulus of the Fourier coefficients: $|F(\vec{k})|^2$;

4. correction for the adopted interpolation scheme;

5. average of the Fourier coefficients in spherical shells in Fourier space to compute the monopole of the power spectrum:

$$\frac{1}{N_k} \sum_{k<|k|<\Delta k} |F(\vec{k})|^2 \tag{2.5}$$

   where Δk is the shell size and N_k is the number of Fourier modes in each shell. Δk is tipically set to $2\pi/L$ where L is the lenght of the computational box.

6. finally, shot noise subtraction:

$$\hat{P}(\vec{k}) = |F(\vec{k})|^2 - P_{shot} \tag{2.6}$$

$$\hat{P}(k) = \frac{1}{N_k} \sum_{k<|k|<\Delta k} |F(\vec{k})|^2 - P_{shot} \tag{2.7}$$

Before describing the technical details of the steps listed above it is worth mentioning that Fourier decomposition is by no means the only way to compute a power spectrum. Sometimes the geometry of the problem makes more natural to adopt different approaches, especially when the focus is on the anisotropies in the clustering and not on the monopole of $P(\vec{k})$. The typical examples are all-sky surveys. In this case the density field traced by galaxies can be more correctly decomposed using spherical harmonics and Bessel functions in what is called "Fourier-Bessel" decomposition (Tadros et al. 1999). The drawback is the speed of the procedure which is much slower than the "standard" Fourier transform and, therefore, difficult to apply to very large datasets.

On the other hand, the adoption of the Fourier decomposition approach poses severe limits on the possibility to estimate anisotropies in galaxy clustering since the underlying assumption is "flat sky" in which all the line of sights to galaxies are considered parallel to each other. Indeed a hybrid approach has also been proposed (Beutler et al. 2013; Yamamoto et al. 2006) in which this approximation is adopted only in a local sense. Yet, this method is significantly more computationally demanding than the Fourier decomposition. For this reason in this thesis we will focus on the latter and on the analysis of the $P(k)$ monopole.

2.1.1 Interpolation schemes

The first step consists in computing a continous density field for a discrete set of objects. This is typically done by interpolating the mass of each object onto a regular grid, before transforming it into the Fourier domain. The use of a cubic grid is dictated by the need of applying the Fast Fourier Transform technique (Press, 1986-1992). Note that most of the commonly used FFT also impose a constrain on the number of gridpoints that should be equal to some power of 2.

There is some freedom in the interpolation scheme. Basically interpolation is performed using some polynomial functions whose order is detemined by the nature of the problem. In this thesis we shall consider two different mass assignement schemes (MAS): the *nearest grid point* (NGP) and the *cloud in cell* (CIC) schemes. The interpolated density field is the convolution between the original density field $F(\vec{r})$ with a MAS function $W_{MAS}(\vec{r})$ given by:

$$W_{MAS}(\vec{r}) = \prod_{i=1}^{3} W_{MAS}(x_i) \tag{2.8}$$

where the coordinate of the generic object $x_i = x, y, z$ for $i = 1, 2, 3$ are expressed in unities of cell size. $W_{MAS}(x_i)$ for NGP and CIC are:

$$W_{NGP}(x_i) = \begin{cases} 1 & \text{if} \quad |x_i| < 0.5 \\ 0.5 & \text{if} \quad |x_i| = 0.5 \\ 0 & \text{if} \quad |x_i| > 0.5 \end{cases} \tag{2.9}$$

$$W_{CIC}(x_i) = \begin{cases} 1 - |x_i| & \text{if} \quad |x_i| \leq 1 \\ 0 & \text{if} \quad |x_i| > 1 \end{cases} \tag{2.10}$$

The Fourier transform of $W_{MAS}(\vec{r})$ is :

$$\tilde{W}_{MAS}(\vec{k}) = \left[\prod_{i=1}^{3} \frac{\sin(\pi k_i / 2k_{Ny})}{\pi k_i / 2k_{Ny}} \right]^p \tag{2.11}$$

where p is the order of the interpolation scheme. For NGP $p = 1$, for CIC $p = 2$. k_{Ny} is the 1-dimensional Nyquist frequency that we describe in the next Section.

After computing the Fourier transform of the interpolated density field and the squared modulus of each Fourier mode $|F(\vec{k})|^2$, we use the convolution theorem to correct for the MAS spurious signal:

$$|F(\vec{k})|^2 = \frac{|F_{NC}(\vec{k})|^2}{\tilde{W}^2_{MAS}(\vec{k})} \tag{2.12}$$

where the subscript NC on the right end side of this expression indicates the Fourier coefficient before being corrected for the MAS.

2.1.2 Aliasing

As for all the discrete Fourier transform methods (DFT) , the output will be affected by *aliasing*. The typical scale of the aliasing problem is the cell size, Δ. This quantity is determined by the size of the box in which the power spectrum is computed, L, and the number of cells/grid points, N, and corresponds to the spatial resolution with which the density field is sampled. Let's consider for simplicity a 1 dimensional Fourier transform. The interpolated density field, obtained as described in the previous Section, corresponds to a function sampled at a finite number of points (in our case the grid points) separated each other by a regular interval Δ, the cell size. Then, the sampling frequency of the density field (the function which must be Fourier transformed), will be given by $1/\Delta$. The computation of the Fourier transform of a function from a discrete sample of points (in our case the regular grid points), implies the periodicy of the Fourier transformed field with period $[0,2f_{Ny}]$, where f_{Ny}, the Nyquist frequency, corresponds to an half of the sampling frequency and is given by $0.5/\Delta$.The corresponding wavenumber k_{Ny} is π/Δ. All the modes out of the range $[0,2f_{Ny}]$ are spouriously moved , *aliased*, into that range. This effect will distort the result. Generalizing to higher dimensions , also the 3D measured Fourier coefficients and spherically averaged power spectrum will be affected by aliasing.

The expectation value of the squared modulus of the Fourier transform of the interpolated density field $|F(\vec{k})|^2$ is in principle given by (Hockney and Eastwood (1981)):

$$< |F(\vec{k})|^2 >= \sum_{\vec{n}} |W(\vec{k}+2k_{Ny}\vec{n})|^2 P(\vec{k}+2k_{Ny}\vec{n}) + \frac{1}{\bar{n}} |W(\vec{k}+2k_{Ny}\vec{n})|^2 \quad (2.13)$$

where $\vec{n}$ is an integer vector. In order to obtain an estimate of $P(\vec{k})$ (corrected for shot noise, aliasing and MAS) one should in principle solve this equation. The first term on the right end side of the equation corresponds to the contribution of the true power spectrum (multiplied by the MAS window function) to the $\vec{k}$ mode plus all the spurious contributions at that k from higher modes $\vec{k} + 2k_{Ny}\vec{n}$, with $\vec{n} \neq 0$. The second term represents the shot noise plus higher modes contribution. As described in Jing et al. (2005), the shot noise term $\frac{1}{\bar{n}}|W(\vec{k} + 2k_{Ny}\vec{n})|^2$ can be explicited through simple analytic formulas, that can be easily computed and opportunely subtracted to each Fourier coefficient $|F(\vec{k})|^2$. Jing et al. (2005) also suggest an iterative method to obtain $P(k)$ from $\sum_{\vec{n}} |W(\vec{k} + 2k_N\vec{n})|^2 P(\vec{k} + 2k_N\vec{n})$. However in this thesis we decided to correct for MAS simply dividing each Fourier mode $|F(\vec{k})|^2$ by the square of the Fourier transform of the MAS function $\tilde{W}^2(\vec{k})$, according to formula 2.12. In this way we are not correcting for aliasing induced by the finite sampling of the density field. This approach, 2.12, is quite standard (Angulo, 2008). The typical scale of the aliasing problem,

$\Delta = L/(N^{1/3})$, is chosen in order to push the Nyquist frequency to values significantly higher than the scales of cosmological interest. In this thesis we set Δ to typical values $\sim 1 - 2$ Mpc h^{-1}, so that the Nyquist wavenumber is equal or greater than 1 h Mpc^{-1}. We see that aliasing (togheter with a not properly exact correction for MAS) distorts significantly the galaxy power spectrum only at $\sim 60\% - 65\%$ of the Nyquist frequency π/Δ, i.e. on physical scales much smaller than that in which we are interested for cosmological analysis.

2.1.3 Geometry and selection effects

All the previous considerations are valid if the mean number density of objects is constant within the sampled volume and if the volume is a cube. However, real galaxy samples are seldom like that, unless they are conveniently trimmed, which typically implies some loss of information. In real world galaxy samples are selected according to observational criteria that have a direct impact on the geometry of the sample and its mean density.

First of all geometry effect. The area of the sky covered by the survey is seldomly regular and, unless the angular size and the redshift is large, close to a box-like shape. On the contrary the geometry can be rather irregular. In this case the use of a Fourier based power spectrum estimator requires to embed the sample within a larger box, zero padding the galaxy distribution outside the suveyed regions. The measured power spectrum will then be the convolution of the true galaxy spectrum and the so called window function of the survey that accounts for its geometry. We shall discuss the effect of convolution and how to account for it in Section 2.3.4.

The second selection effect is typically represented by the flux threshold that is adopted in the survey.

All these selection effects are referred to as 'mask', while we refer exclusively to the geometrical characteristic of a galaxy sample as 'geometrical mask'.

One way to account for the flux cut is to extract a *volume limited* subsample containing all galaxies that would have been included in the sample if placed at a maximum distance. In practical terms given the flux of the object, d_{MAX} a maximum distance and a k-correction, an object at the redshift z is moved at d_{MAX} and removed from the sample if its new flux is below the survey limit. This guarantees a constant mean number density within the sample.

The downside of this option is to remove a significant amount of objects (and information) while increasing the shot noise error. Alternatively one can keep the whole *flux limited* sample and account for the variation in the mean number density either by weighting each object by a statistical factor that accounts for the flux selection or to model this selection effect in the

window function similar to that used to account for geometry effects. In this thesis we don't extract volume limited samples but we shall use two type of weighting schemes: the one described above and the weighting FKP scheme. As we shall see in the next section FKP weights don't account exactly for the flux selection and their effect is modelled in the window function too.

2.2 The FKP estimator

FKP is one of the most commonly used galaxy power spectrum estimators. It's easy to implement and computationally fast since it relies on the FFT technique. The FKP estimator of the galaxy power spectrum can deal with the complex geometries of the galaxy redshift surveys and variations of the mean number density due to selection effects, since density fluctuation are defined with respect to a control sample of unclustered objects (the random catalog) with precisely the same selction effects. The use of an appropriate weighting scheme guarantees that FKP is a minimum variance $P(k)$ estimator.

2.2.1 Method

In the FKP framework, the galaxy fluctuation field is defined as:

$$F(\vec{r}) \equiv \frac{w(\vec{r})[n_g(\vec{r}) - \alpha n_s(\vec{r})]}{[\int d^3r \bar{n}_g^2(\vec{r}) w^2(\vec{r})]^{1/2}} \tag{2.14}$$

where $w(\vec{r})$ is a weighting function introduced in order to minimize the variance of the estimator, α and $n_s(\vec{r})$ are functions that account for the varying mean number density of objects as follows. $n_s(\vec{r})$ represents the number density at the point $\vec{r}$ of a sample of objects that are unclustered but characterized by the same selection effects as the real sample, that is to say, same geometry and redshift distributions. The number of objects in this random catalog doesn't need to match that of the real one. The ratio between the two is set by a parameter α defined as:

$$\alpha = \frac{\int d^3r n_g(\vec{r}) w(\vec{r})}{\int d^3r n_s(\vec{r}) w(\vec{r})} \tag{2.15}$$

which guarantees that the fluctuation field has zero mean $< F(\vec{r}) >= 0$. Since this random catalog is used to compute the window function of the sample it typically consists of a much higher number of objects, $\alpha << 1$.

The weighted mask of the survey $\tilde{M}(\vec{r}) = \bar{n}_g(\vec{r}) w(\vec{r})$. The galaxy fluctuation field can be factorized leading to:

$$F(\vec{r}) = \frac{\tilde{M}(\vec{r})}{N} \left[\frac{n_g(\vec{r})}{\bar{n}_g(\vec{r})} - \frac{\alpha n_s(\vec{r})}{\bar{n}_g(\vec{r})} \right] \tag{2.16}$$

where

$$N = \left[\int d^3 r \tilde{M}^2(\vec{r}) \right]^{1/2}. \tag{2.17}$$

When averaged on large values the second term on the right end side of the 2.16 is equal to 1.

The expectation value of the square of the Fourier transform of $F(\vec{r})$ is given by:

$$< |F(\vec{k})|^2 >= \frac{\int d^3 r \int d^3 r' w(\vec{r}) w(\vec{r'}) < [n_g(\vec{r}) - \alpha n_s(\vec{r})][n_g(\vec{r'}) - \alpha n_s(\vec{r'})] > e^{i\vec{k}\cdot(\vec{r}-\vec{r'})}}{\int d^3 r \bar{n}_g^2(\vec{r}) w^2(\vec{r})} \tag{2.18}$$

Under the Poisson sampling hypotesis equation 2.18 can be written as:

$$< |F(\vec{k})|^2 >= \int \frac{d^3 k'}{(2\pi)^3} P(\vec{k'}) |M(\vec{k} - \vec{k'})|^2 + (1+\alpha) \frac{\int d^3 r \bar{n}_g(\vec{r}) w^2(\vec{r})}{\int d^3 r \bar{n}_g^2(\vec{r}) w^2(\vec{r})} \tag{2.19}$$

where

$$M(\vec{k}) \equiv \frac{\int d^3 r \bar{n}_g(\vec{r}) w(\vec{r}) e^{i\vec{k}\cdot\vec{k}}}{[\int d^3 r \bar{n}_g^2(\vec{r}) w^2(\vec{r})]^{1/2}}. \tag{2.20}$$

The function $M(\vec{k})$ is the Fourier transform of the weighted mask of the survey $\tilde{M}(\vec{r})$. Equation 2.19 is the generalization of equation 2.2. It shows that the expected value of the squared modulus of the galaxy overdensity field in Fourier space is given by the sum of two terms: the convolution between the true galaxy power spectrum $P(\vec{k})$ and the window function of the survey $W \equiv |M(\vec{k} - \vec{k'})|^2$ and a term which represents the shot noise contribution due the discrete nature of the tracers. The first term shows that the window of the survey induce a correlation among different Fourier modes. The second term is the shot noise power, given by:

$$P_{shot} \equiv (1+\alpha) \frac{\int d^3 r \bar{n}_g(\vec{r}) w^2(\vec{r})}{\int d^3 r \bar{n}_g^2(\vec{r}) w^2(\vec{r})}. \tag{2.21}$$

This term includes also the shot noise contribution of the objects of the random catalog, that is in general less important because $\alpha << 1$.

As for the constant density case, the FKP estimator of $P(\vec{k})$ is given by:

$$\hat{P}(\vec{k}) = |F(\vec{k})|^2 - P_{shot}. \tag{2.22}$$

The subtraction of the shot noise term P_{shot} removes contribution to the power simply due to the discrete sampling of the field. The final estimator of $P(k)$ is obtained averaging on spherical shells in k-space:

$$\hat{P}(k) \equiv \frac{1}{V_k} \int_{V_k} d^3 k' \hat{P}(\vec{k'}). \tag{2.23}$$

2.2.2 Variance of the estimator

The variance of the FKP estimator $\hat{P}(k)$ is obtained under the assumption that the Fourier coefficients $F(\vec{k})$ are realizations of a Gaussian distribution. As seen in the previous Chapter, this corresponds to the prediction of most of inflationary models and observationally confirmed by the spectrum of the CMB. The variance can be written as:

$$\sigma_P^2 \equiv < [\hat{P}(k) - P(k)]^2 >= \frac{1}{V_k^2} \int_{V_k} d^3k \int_{V_k} d^3k' < \delta\hat{P}(\vec{k})\delta\hat{P}(\vec{k}') > \quad (2.24)$$

where $\delta\hat{P}(\vec{k}) = \hat{P}(\vec{k}) - P(k)$. Under the assumption that the Fourier coefficients $F(\vec{k})$ are Gaussian distributed it follows that:

$$< \delta\hat{P}(\vec{k})\delta\hat{P}(\vec{k}') >= | < F(\vec{k})F^*(\vec{k}') > |^2. \quad (2.25)$$

Using equation 2.19, $< F(\vec{k})F^*(\vec{k}') >$ can be written as:

$$< F(\vec{k})F^*(\vec{k}') >= \int \frac{d^3k''}{(2\pi)^3} P(\vec{k}'')M(\vec{k} - \vec{k}'')M^*(\vec{k}' - \vec{k}'') + S(\vec{k}' - \vec{k}) \quad (2.26)$$

where the function S is defined as:

$$S(\vec{k}) \equiv \frac{(1+\alpha)\int d^3r \bar{n}_g(\vec{r})w^2(\vec{r})e^{i\vec{k}\cdot r}}{\int d^3r \bar{n}_g^2(\vec{r})w^2(\vec{r})} \quad (2.27)$$

Let's call $\vec{k} - \vec{k}' = \delta\vec{k}$. If the survey is characterized by a compact geometry, then the Fourier transform of the mask $M(\vec{k})$ is a rather compact function that is defined over a range $\delta k \sim 1/D$, where D is the typical size of the survey. If $|\vec{k}| >> 1/D$, the quantity $P(\vec{k}'')$ can be extracted from the integral in the right end side of the expression 2.26, that reduces to:

$$< F(\vec{k})F^*(\vec{k}') > \simeq P(\vec{k})Q(\delta\vec{k}) + S(\delta\vec{k}) \quad (2.28)$$

where $Q(\vec{k})$ is defined as

$$Q(\vec{k}) \equiv \frac{\int d^3r \bar{n}_g^2(\vec{r})w^2(\vec{r})e^{i\vec{k}\cdot r}}{\int d^3r \bar{n}_g^2(\vec{r})w^2(\vec{r})}. \quad (2.29)$$

So

$$< \delta\hat{P}(\vec{k})\delta\hat{P}(\vec{k}') >= |P(\vec{k})Q(\delta\vec{k}) + S(\delta\vec{k})|^2. \quad (2.30)$$

This formula, evaluated for $\vec{k} = \vec{k}'$, represents the variance of the estimator $\hat{P}(\vec{k})$. It is given by the squared module of the sum of two terms: the total power and the shot noise (equation 2.27 evaluated for $\vec{k} = 0$, $S(0)$, corresponds to the shot noise contribution to power).

When the k-shell in which the spherical average is performed is large compared to the coherence lenght $\delta k = 1/D$, namely when the effect of the

mode coupling due to the presence of the mask is limited within each k-shell, so that the different k-bins can be considered approximatively independent, then equation 2.24 reduces to:

$$\sigma_P^2(k) \simeq \frac{1}{V_k} \int d^3k' |P(k)Q(\vec{k}') + S(\vec{k}')|^2 \tag{2.31}$$

Therefore, using the Parseval's theorem , the fractional variance of the power is:

$$\frac{\sigma_P^2(k)}{P^2(k)} = \frac{(2\pi)^3 \int d^3r \bar{n}_g^4 w^4 [1 + 1/\bar{n}_g P(k)]^2}{V_k [\int d^3r \bar{n}_g^2 w^2]^2} \tag{2.32}$$

FKP weights. The optimal choice for the weighting function $w(\vec{r})$ is obtained minimizing the fractional variance 2.32 respect to $w(\vec{r})$ and is given by:

$$w_{FKP}(\vec{r}) = \frac{1}{1 + \bar{n}_g(\vec{r})P(k)}. \tag{2.33}$$

The use of the weighting function 2.33 requires a preliminary evaluation of the $P(k)$. In practice, it is approximated to the mean value of the recovered power spectrum in the k-range of interest.

Substituting equation 2.33 into 2.32 gives

$$\frac{\sigma_P^2(k)}{P^2(k)} = \frac{(2\pi^3)}{V_k} \frac{1}{V_{eff}} \tag{2.34}$$

where V_{eff} represents the effective volume of the survey that can be explicited for clustering analysis as

$$V_{eff} = \int d^3r \left[\frac{\bar{n}_g(\vec{r})P(k)}{1 + \bar{n}_g(\vec{r})P(k)} \right]^2. \tag{2.35}$$

Let's consider for simplicity the case of constant mean number density $\bar{n}_g(\vec{r}) = \bar{n}_g$. In this case the effective volume reduces to:

$$V_{eff} = \left[\frac{\bar{n}_g P(k)}{1 + \bar{n}_g P(k)} \right]^2 V_{survey} \tag{2.36}$$

with V_{survey} equal to the survey volume, and the squared root of the variance of the power spectrum is:

$$\sigma_P = \sqrt{\frac{2\pi^3}{V_k V_{survey}}} \left(P(k) + \frac{1}{\bar{n}_g} \right) \tag{2.37}$$

The term $(P + \bar{n}_g^{-1})$ corresponds to the squared root of the variance in the power for a single k-mode: as already anticipated it is given by the sum of power spectrum and shot noise (see section 2.1 for shot noise in case of constant density). The first term derives from having estimated the variance

of the distribution of the fluctuations, namely the power spectrum $P(\vec{k})$, from one single realization of the distribution, namely the single sampled Fourier coefficient $|F(\vec{k})|^2$. The quantity $V_k V_{survey}/2\pi^3$ corresponds to the number of independent modes within the k-shell in which we average in order to obtain $\hat{P}(k)$. In fact V_k corresponds to the volume of the k-shell and $2\pi^3/V_{survey}$ is the coherence volume in Fourier space. Each volume can be considered approximatively independent from the others (Tegmark 1997). When $V_{survey} \to \infty$ the statistical error on $\hat{P}(k)$ tends to 0 and the estimator tends to the true galaxy power spectrum $\hat{P}(k) \to P(k)$. The overdensity field is real valued, so , the number of independent modes per shell should be, more correctly, equal to $V_k V_{survey}/4\pi^3$, being $\vec{k}$ and $-\vec{k}$ not independent.

The first term of 2.37 :

$$\sqrt{\frac{4\pi^3}{V_k V_{survey}}} P(k) \tag{2.38}$$

represents the *cosmic variance* contribution to the uncertainty. This term dominates the error budget on large scales. It is due to the finite volume under study and constitutes an intrinsic limitation to the precision with which the power spectrum can be determined.

The second term is the shot noise contribution to the variance, due to the discrete and finite sampling of the density field. It dominates on small scales.

2.3 Practical implementation of the FKP estimator

In this Section we show how FKP estimator has been practically implemented. First of all, we show how we measure the mean number density of a galaxy sample. Then, we briefly show how the FKP method is implemented, and finally how one can account for the window survey.

2.3.1 The mean number density estimation

An accurate estimate of the variation of the mean number density of objects across the sample is an important pre-requisite to the building of the mock catalog and to obtain an unbiased estimate of the power spectrum. Here we assume that spatial variation in the mean number density can be factorized in two terms: one that quantifies angular variation and the other that quantifies radial variations. The former is the subject of the next Section in which we assume that it can be modelled with a geometry mask with no angular selection effect within the surveyed area and zero mean density

outside. Here we focus on the radial selection issue and on how to quantify it.

The radial variation of the mean number density can be estimated in different ways. One of them consists in counting objects in redshift bins, derive the histogram of galaxy counts and smooth it to regularize the features that large scale structrures can imprint if the sample is not a fair sample. The advantage of this approach is its simplicity and the downback is the lack of rigorous criteria about the smoothing procedure.

A better method, more statistically robust is the F/T *estimator*, in Davis and Hucra (1982), based on the measure of the selection function.

The technique consists in the evaluation of the selection function from the observed counts. It has been applied on the 2MRS galaxy redshift survey (Branchini et al. 2012) and provides results that are consistent with the methods in which the radial selection function is derived from the luminosity function of the sample (i.e. Schmidt , 1968). For this purpose two functions, dubbed $F(s)$ and $T(s)$ are introduced, where s is the redshift in Km s^{-1}. For a given s value, $T(s)$ represents the number of galaxies that would still be included in the sample (i.e. bright enough) when placed at $s' > s$. $F(s)$ is the number of objects that can only be detected out to $s + ds$. The ratio between F and T gives an estimate of the relative variation of the selection function $S(s)$ at distance s, independent from the local inhomogeneities. The selection function can then be computed by integrating the differential equation:

$$\frac{dlnS(s)}{ds} = -\frac{F(s)}{T(s)} \tag{2.39}$$

The selection function is by definition proportional to the mean number density , a part from a normalization factor A, that can be set directly from the data. As outlined in Branchini et al. (2012), this method assumes no dipendence of the galaxy luminosity on the redshift and, consequently, the selection does not evolve with the redshift. This is not true and a possible luminosity evolution must be allowed for in the modelling and determined by fitting the observed histogram of galaxy counts as a function of redshift.

2.3.2 Random catalog building

As we have seen the random catalog is an essential ingredient of the FKP estimator and consists of a collection of unclustered objects with mean number density equal to $1/\alpha$ than the real catalog.The building of such catalog is made in two steps. First we generate a random distribution of objects in a cubic box with lenght L, encompassing the real catalog. Each object is randomly placed at $\vec{r} \equiv(\ x_{ran},\ y_{ran}\ ,\ z_{ran}\ )$ where each Cartesian coordinate is randomly sampled in the range [0,L] according to a flat probability distribution. This temporary random catalog has a constant density, and is $1/\alpha$ denser than the maximum value of the mean number

density measured in the real sample, $\bar{n}_0$. The factor $1/\alpha$ is typically set to a value greater than 10 to model accurately the survey mask. In the second step, the temporary catalog is diluted by a spatially varying fraction $\bar{n}(\vec{r})/\bar{n}_0$ using the following Monte Carlo rejection procedure: we select an object of the temporary catalog at $\vec{r}$, draw a random number in the interval [0,1] and compare it with $\bar{n}(\vec{r})/\bar{n}_0$. If the random number is larger, than the object is discarded. Othewise, we keep it. The resulting catalog has a mean number density $n_s(\vec{r}) = \bar{n}(\vec{r})/\alpha$ but no clustering properties. This is the catalog used in the FKP estimator.

2.3.3 Steps to the implementation

The steps to estimate $P(k)$ and its uncertainties, with the FKP method are as follows:

1. estimate of FKP weights using 2.33 for both the real and the random objects.

2. The weighted number densities of the real and the random catalog, respectively given by $\sum_{g,i} \delta_D(\vec{r} - \vec{r}_{g,i}) w(\vec{r}_{g,i})$ and $\sum_{s,i} \delta_D(\vec{r} - \vec{r}_{s,i}) w(\vec{r}_{s,i})$, where $\vec{r}_{g,i}$ is the position of the i-th real catalog object and $\vec{r}_{s,i}$ the position of the i-th random catalog object, are separately interpolated onto a cubic grid, using CIC mass assignement scheme.

3. Estimate the difference density field in 2.14 as: $[n_{g,W}(\vec{r}) - \alpha n_{s,W}(\vec{r})]$, where $n_{g,W}(\vec{r})$ and $n_{s,W}(\vec{r})$ are respectively the interpolated weighted number densities of the real and the random catalog. The factor α, specified in 2.15, is estimated as $\alpha = \sum_g w(\vec{r}_g) / \sum_s w(\vec{r}_s)$.

4. Fourier Transform of the difference field using FFT and correction for the mass assignement scheme.

5. Compute the monopole by averaging over spherical $k-$shells:

$$\hat{F}(k) = \frac{1}{N_k} \sum_{k<|k|<\Delta k} |F(\vec{k})|^2 \tag{2.40}$$

6. Compute the normalization factor 2.17:

$$N^2 = \alpha \sum_s \bar{n}_g(\vec{r}_s) w^2(\vec{r}_s) \tag{2.41}$$

7. Estimate the shot noise contribution to be subtracted from the power spectrum:

$$P_{shot} = \alpha(1+\alpha) \sum_s w^2(\vec{r}_s). \tag{2.42}$$

8. Estimate of the power spectrum:

$$\hat{P}(k) = \hat{F}(k) - P_{shot}. \tag{2.43}$$

9. Compute the statistical errors on $P(k)$. The squared root of the variance of the power spectrum FKP estimator (2.24) is computed as:

$$\sigma_P^2(k) = \frac{2}{N_k^2} \sum_{\vec{k}'} \sum_{\vec{k}''} |PQ(\vec{k}' - \vec{k}'') + S(\vec{k}' - \vec{k}'')|^2. \tag{2.44}$$

This expression accounts for possible correlations among the different k-bins due to the window function convolution. $Q(\vec{k})$ is computed as:

$$Q(\vec{k}) = \frac{\alpha \sum_s \bar{n}_g(\vec{r}_s) w^2(\vec{r}_s) e^{i\vec{k}\cdot\vec{r}s}}{N^2} \tag{2.45}$$

and $S(\vec{k})$ as

$$S(\vec{k}) = \frac{\alpha(1+\alpha) \sum_s w^2(\vec{r}_s) e^{i\vec{k}\cdot\vec{r}s}}{N^2}. \tag{2.46}$$

If we assume all Fourier modes are independent, which is not necessarly true when a geometric mask is present, the squared root of the variance reduces to :

$$\sigma_P(k) = \sqrt{\frac{2}{N_k}}(P(k) + S(0)). \tag{2.47}$$

where N_k is the number of modes sampled in each k-shell.

Expression 2.44 is quite tricky to express and it is valid only in the limit of Gaussian distributed Fourier coefficients $F(\vec{k})$. In this thesis we estimate the errors from mock dataset matching the properties of the real sample. Given a set of independent mock catalogs, we evaluate the statistical error from the *rms* scatter of the measured power spectra around the mean:

$$\frac{\sigma_P(k)}{< P(k) >} = \frac{1}{< P(k) >} \sqrt{\frac{\sum_{i=1}^{N_{mocks}} (P_i(k) - < P(k) >)^2}{N_{mocks} - 1}} \tag{2.48}$$

where N_{mocks} is the number of mocks, $P_i(k)$ is the measured power spectrum of the i-th mock and $< P(k) >$ is the mean among the mocks.

2.3.4 Mask Convolution

According to equation 2.19 and 2.22, the FKP estimator of the power spectrum $\hat{P}(\vec{k})$, that from now on we generically refer to as the measured power spectrum, $P_m(\vec{k})$, is the convolution between the true galaxy power spectrum $P_g(\vec{k}')$ and the window function $|M(\vec{k})^2| = W(\vec{k})$ (2.20).

$$P_m(\vec{k}) = \int \frac{d^3\epsilon}{(2\pi)^3} P_g(\vec{k}' + \vec{\epsilon}) W(\vec{\epsilon}) - \frac{W(\vec{k})}{W(0)} \int d^3\epsilon P_g(\vec{\epsilon}) W(\vec{\epsilon}) \tag{2.49}$$

The second term on the right end side of this expression represents the *integral constraint* , and accounts for the fact that the measured power spectrum is set to zero at $k = 0$ to guarantee a zero-mean density fluctuation field.

The spherically averaged power spectrum , i.e. the measured monopole of the galaxy distribution, is:

$$P_m(k) = \frac{1}{V_k}\int_{V_k} d^3k' \left[\int \frac{d^3\epsilon}{(2\pi)^3} P_g(\vec{k}' + \vec{\epsilon})W(\vec{\epsilon}) - \frac{W(\vec{k})}{W(0)}\int d^3\epsilon P_g(\vec{\epsilon})W(\vec{\epsilon})\right] \tag{2.50}$$

Let's focus on the first term on the right end side of the 2.50 that quantifies the convolution effect of the window function that needs to be accounted for to extract the physical information encoded in $P_g(\vec{k}')$ from the measurement $P_m(\vec{k})$. The deconvolving procedure from which one obtains $P_g(\vec{k}')$ from $P_m(\vec{k})$ is notoriously challenging and plagued by numerical instabilities. For this reason the common procedure, also adopted in this thesis, is to model $P_g(\vec{k}')$, estimate the integral on the right end side of 2.50 and compare the results with $P_m(\vec{k})$. The accuracy of this procedure critically depends on the precision with which we can define $W(\vec{k})$.

We refer to the power spectrum obtained convolving a generic theoretical model $P_g(\vec{k})$ with the window function as $P_{conv}(\vec{k})$, the convolved model power spectrum.

In general the window function of the sample is not isotropic and one should account for the explicit dependence on the single vector $\vec{\epsilon}$. However, the geometry of the sample is reasonably regular (like in the case of all sky-surveys) and one is interested in the monopole of the power spectrum then one can consider the spherically-averaged window function W_s (Percival et al. 2001, Cole et al.2005,Ross et al. 2013), defined as:

$$W_s(\epsilon) = \frac{1}{4\pi}\int d\Omega_{\epsilon'} W(\vec{\epsilon}')\delta(r_{\epsilon'} - r_\epsilon) \tag{2.51}$$

where $d\Omega_{\epsilon'}$ is the solid angle element: $\sin(\theta)d\theta d\phi$ with θ polar angle and ϕ azimuthal angle. $r_\epsilon = |\vec{k} + \vec{\epsilon}|$. W_s does not depend on angles and can be used to link the monopole of the true and measured power spectra:

$$P_m(k) = \int \frac{d^3\epsilon}{(2\pi)^3} P_g(|\vec{k} + \vec{\epsilon}|)W_s(\epsilon) \tag{2.52}$$

where the direction of $\vec{k}$ is kept fixed and the integral is over $\vec{\epsilon}$. A possible drawback of this simplifying assumption is the systematic error that it may be induced on possible clustering anisotropies, like redshift-space distortions, that may compromise important cosmological information. In these case the magnitude of this effect needs to be carefully quantified.

In this thesis we verify that this approximation does not compromise the extraction of cosmological paramenters form the clustering of the 2MRS galaxies.

In order to perform the integral 2.52 we follow Ross et al. 2013 (appendix A) and perform the convolution integral with a matrix-based approach. The actual estimate of $W_s(|\vec{\epsilon}|)$ is obtained directly from the random catalog by computing the monopole power spectrum of the random sample using FFT:

$$W_s(\epsilon) = \frac{1}{N_\epsilon} \sum_{\epsilon<|\epsilon|<\epsilon+\delta\epsilon} [W(\vec{\epsilon}) - P_{shot,W}] \tag{2.53}$$

where $W(\vec{\epsilon})$ is the squared modulus of the Fourier transform of the interpolated density field of the random catalog, corrected for the MAS, $P_{shot,W}$ is the shot noise and N_ϵ is the number of Fourier modes in the shell $\delta\epsilon$. Both the spherically averaged window function W_s and the power spectrum of the real catalog $P_m(k)$ are measured in discrete k-bins. As a result, the integral convolution can be estimated as:

$$P_{conv}(k_i) = \sum_j W[k_i][k_j] P_g(k_j) - \frac{W_s(k_i)}{W_s(0)} \sum_j W[0][k_j] P_g(k_j) \tag{2.54}$$

where $W[k_i][k_j]$ is the convolution matrix:

$$W[k_i][k_j] = \frac{1}{2\pi^2} \int\int W_s(\epsilon)\epsilon^2 \Theta(r_\epsilon, k_j) \sin(\theta) d\theta d\epsilon \tag{2.55}$$

where $r_\epsilon \equiv |\vec{k_i} + \vec{\epsilon}| = \sqrt{k_i^2 + \epsilon^2 - 2k_i\epsilon\cos(\theta)}$. $\Theta(r_\epsilon, k_j)$ is equal to 1 if r_ϵ is within the k_j bin and 0 otherwise. The second term on the right end side of this expression is the spherical average of the *integral constraint*.

2.4 Modelling galaxy bias and redshift space distortions

In this Section we focus on the modelling of the bias and redshift space distortions. A model of the galaxy power spectrum is required to compare the measured quantity with the theoretical predictions represented by the convolution integral. The model galaxy power spectrum in 2.19, $P(k')$, needs to account for galaxy bias, i.e. the fact that galaxies do not necessarly trace the mass, possible non linear effects that affect the modelling of the mass power spectrum and for the fact that observed redshift are typically used as proxy for the galaxy distance, hence inducing spurious redshift distortions in galaxy clustering. We adopt the standard phenomenological approach of using some simple but physically motivated models for the galaxy power

spectrum that account for all these effects. These models are characterized by a small number of free parameters.

A fairly general model for the monopole of the galaxy power spectrum at $z = 0$ can be expressed as follows:

$$P_g(k) = b^2 P_{lin}(k, P_1) K(\beta) F(k, P_2) G(k, \sigma) \tag{2.56}$$

where $P_{lin}(k, P_1)$ is the mass power spectrum in the linear regime that depends on a number of cosmological parameters P_1, b is the galaxy bias that is assumed to be deterministic and linear and $F(k, P_2)$ is a term that accounts for middly non linear and scale-dependent effects both in the galaxy bias and in the growth of density fluctuations. It is specified by a number of free parameters P_2 and can be included in the bias expression as well as in the matter power spectrum. $G(k, \sigma)$ is the damping term that accounts for highly non linear motions characterized by a virial pair velocity dispersion σ. $K(\beta)$ is the Kaiser boost, eq. 1.58. We shall use the formalism described by eq. 2.56 in Chapter 3 and 4.

The galaxy power spectrum on large scales should be sufficiently well reproduced by:

$$P_g(k) = b^2 P_{lin}(k, P_1) K(\beta) \tag{2.57}$$

where the linear theory prediction $P_{lin}(k, P_1)K(\beta)$ is multiplied by a scale independent bias.

In order to model the power spectrum in the middly non linear regime, several authors have proposed different analytic models. One of the most diffuse in literature is the model proposed by Cole et al. 2005:

$$P_g(k) = b_K^2(k) K(\beta) P_{lin}(k) \tag{2.58}$$

where

$$b_K(k) = b \left[\frac{1 + Q(k/k_1)^2}{1 + A(k/k_1)} \right]^{1/2} \tag{2.59}$$

This expression, commonly referred to as Q-model, corresponds to a k-dependent correction to the linear model 1.57. Without lack of generality we fix k_1 at 1 h Mpc^{-1}.

Fig. 2.1 shows the power spectrum measures of the N-body Hubble Volume simulation, both for dark matter and galaxies, from Cole et al. 2005. The upper panel shows the linear theory power spectrum $P_{lin}(k)$, computed for the input cosmology of the simulation, denoted by a continous line. The dotted curve represents the power spectrum of galaxies in real space: the square of the large scale linear bias b is equal to 1.061 while for $k > 0.12h$ Mpc^{-1} the bias becomes scale dependent. The dashed red line represents the power spectrum of galaxies in redshift space. On large scales the power spectrum is boosted due to the coherent motions (Kaiser boost, dubbed K) and the product b^2K is equal to 1.441. On smaller scales

the power spectrum is damped respect to the real space one , due to the incoherent motions (FoG). The lower panel shows how the power spectum of the Hubble volume galaxies differs from the linear theory, both in real and redshift space. The continous line shows what happens with the power spectrum of the dark matter particles of the Hubble Volume simulation. It differs significantly from linear theory on small scales. As explained in Cole et al. 2005, the Hubble Volume simulations are realistic in some respects but not completely physically motivated. This figure is shown here only to give a qualitative idea of the differences between galaxy power spectrum and linear theory.

The k-dependent correction 2.59, provides a good fit to the predictions of semi-analitic models of galaxy formation, as shown in the paper. It has been used in order to fit the full shape of the power spectrum of the 2dF Galaxy Redshift Survey. In order to do this, the authors propose to fix the A parameter to the result of the fits to the numerical simulation, $A = 1.4$ for the redshift space case ($A = 1.7$ for real space). The robustness of this assumption is quite well supported by the Halo Model predictions, as explained in the paper. Differently, the Q parameter is treated as nuisance parameter, allowed to vary from 0 to approximatively the double of the recovered value from the simulations and marginalized over when fitting the 2dF power spectrum measures.

One limitation of the Q-model is that it doesn't account for the dependence of the A parameter on the galaxy type, unless one considers A as additional free parameter. Another limit of the Q-model is that it doesn't account for an eventual non-Poisson shot noise term, due to the fact that, if galaxies form only in the dark matter haloes, they are not Poissonian samplers of the density field.

Other possible models of bias and redshift distortions, that we use in this thesis, are:

$$P_g(k) = b^2 K(\beta) P_{lin}(k)(1 + Q(k/k_1)^2) \tag{2.60}$$

and

$$P_g(k) = b^2 K(\beta) P_{lin}(k)(1 + Q(k/k_1)^{3/2}) \tag{2.61}$$

The first model is the Q-model for $A = 0$. The second one is analogous to the previus one but shows a different dependence on k. In both cases k_1 is set to 1 h Mpc^{-1}.

Another simple model for the bias is:

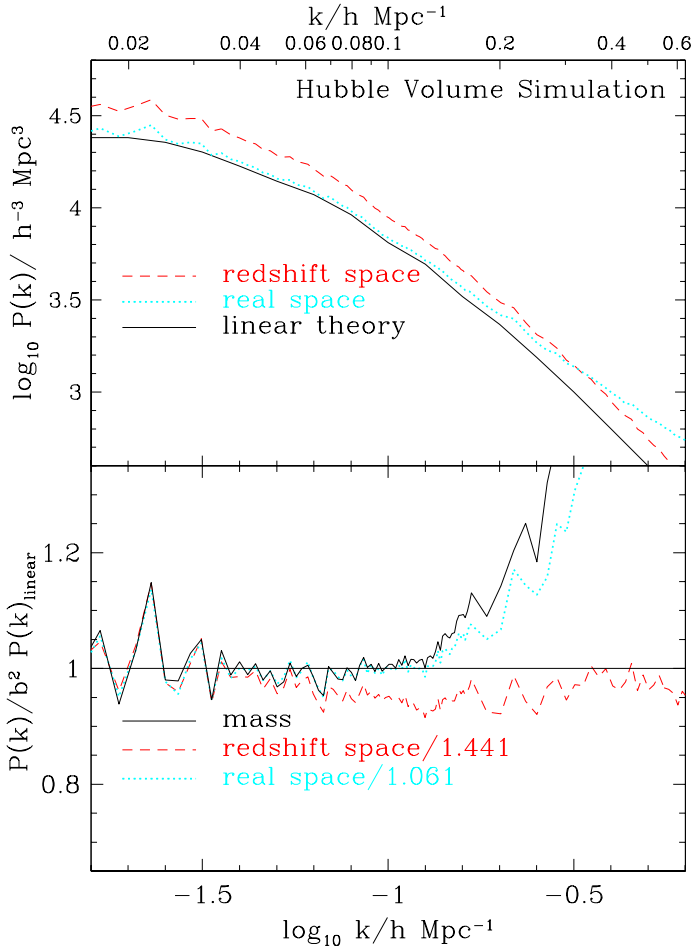
$$P_g = b_K^2(k) K(\beta) P_{lin}(k) \tag{2.62}$$

with

$$b_K(k) = b + b_1(k/k_1)^n \tag{2.63}$$

where b and b_1 are constant. We refer to this model as $P-$model, a shorcut for Power-law model. It is similar to the one obtained from a Taylor

Figure 2.1: The upper panel shows the power spectrum of galaxies of the Hubble Volume simulation in real space (dotted line) and redshift space (dashed line). The continous line is denotes the linear theory power spectrum computed for the input cosmological parameters of the simulation. The lower panel shows the ratio between the galaxy power spectrum and $b_s^2 P_{lin}(k)$ both in real and redshift space, respectively dotted and dashed line ($b_s^= b^2$ in real space, $b_s = b^2 K$ in redshift space). The continous line is denotes the dark matter particles of the simulation. Plot from Cole et al., 2005.



expansion of the mass density contrast (Fry and Gaztanaga, 1993). We note that this model is also similar to the one proposed by Seo and Eisenstein (2005) whose validity is limited to ranges of wavenumber where the power spectrum can be well approximated by a power law. Note that for $A \ll 1$ and $n = 1$ the Q-model and the P-model are the same. Also in this case we fix k_1 at 1 h Mpc^{-1}.

Other models have been proposed with stronger theoretical motivations

(Smith et al. 2007 and references therein). These models account, for example, for eventual non-Poisson shot noise term. Smith et al. 2007 propose a complex starting point for empirical modelling , which includes non linear corrections galaxy type dependent and non Poisson shot noise, that, for $k << 1h/\mathrm{Mpc}$, is:

$$P_g(k,T) = b^2(T)\{P_{lin}(k)[1 - A_1(T)k^2] + A_2 k^{m(T)}\} + A_0(T) \tag{2.64}$$

where T denotes the galaxy type. The factor $[1 - A_1(T)k^2]$ term in the right end side of 2.64 accounts for the damping effect of BAO. The term $A_2 k^{m(T)}$ accounts for non linear bias and increase of power due to the correlations among the k-modes. The term $A_0(T)$ accounts for all possible sources of constant power.

Another type of model of the galaxy power spectrum that we adopt consists in using the non linear matter power spectrum output of the *Halofit* code, $P_{nl}(k)$, and modelling dynamical non linearities with the damping factor, as follows:

$$P_g(k) = b^2 K(\beta) P_{nl}(k) G(k,\sigma) \tag{2.65}$$

where

$$G(k,\sigma) = \frac{\pi^{1/2}}{2}\frac{erf(\sigma k)}{\sigma k}. \tag{2.66}$$

The damping factor in the 2.65 is obtained averaging over all angles the line of sight Fourier transform of the velocity distribution, given by: $exp\left[-(\mu k\sigma)^2/2\right]$, where μ is the cosine of the angle between $\vec{k}$ and the line of sight. We show that also expression 2.65 provides a good model for the galaxy power spectrum.

Chapter 3

Assessing Code Performance

The goal of this Chapter is to assess the performance of the code that we use to estimate the power spectrum of the galaxy distribution. We do this in two step. First we test the performance of the code by comparing the results with those of other implementations of the same FKP estimator. This code contest was performed using the same test of ideal, mock dataset. Second, focusing on our code, we assess the impact of the geometry and selection effects typical of the catalog that we want to study, the 2MASS Redshift Survey, on the power spectrum estimate. The goal is that of selecting the best range of wave numbers in which one should perform the analysis of the real data and obtain a first assessment of its accuracy.

3.1 Comparing different FKP codes

As we have seen, in our analysis of 2MRS we use the FKP method to estimate the power spectrum of the galaxy distribution. An effective way of testing the performance of our implementation of the method is that of comparing the results obtained by our numerical code with those obtained by other codes, independently estimated by other authors. We have done this in the more general framework of the code contest that has been set up to select the best performing FKP code to be further developed within the Euclid Consortium and that it will be used to estimate the power spectrum of the galaxies in the Euclid spectroscopic survey (Laurejis et al 2012).

The codes considered in this contest besides ours were:

- the FKP code developed by A.Balaguera-Antolinez, dubbed B-code, and used to analyse the spatial distribution of Galaxy Clusters in the Reflex survey (Balaguera-Antolìnez et al. 2011);
- the FKP code developed by W. Percival and L. Samushia (S-code) and extensively used to estimate the power spectrum of galaxies in the SDSS and BOSS surveys.

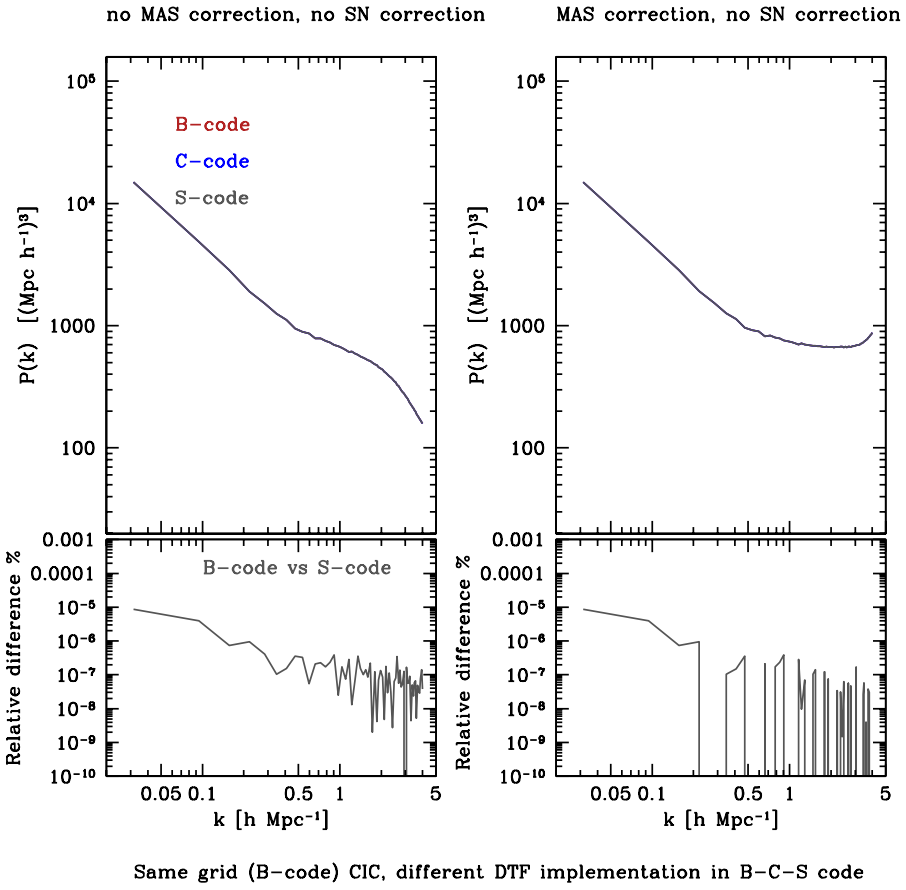
All codes are based on the Fast Fourier Transform technique. Unlike these two codes, which are based on the FFTW public libraries (Frigo & Johnson, 2005), our code (dubbed C-code) uses the FFT routine described in the Numerical Recipes libraries (Press, 1992). Furthermore, B and S-code use NGP, CIC and TSC mass assignment schemes (MAS) and their corrections to interpolate the density field at the grid positions. TSC is an higher order mass assignement scheme. Our code uses NGP and CIC mass assignement schemes. The three codes have been used to analyze the same simulation distribution of 12798 galaxies in a cubic box of 200 h^{-1} Mpc with a density field interpolated onto a grid 256^3. The sample has been extracted from the so called "100 deg^2" mock galaxy catalogs [1] obtained by applying the Semi-Analytic galaxy formation model "GALFORM" (Cole et al. 2000, Baugh et al. 2005, Bower et al. 2006) to the outputs of the Millennium N-body simulation (Springel et al. 2005). The simulated volume consists of a lightcone with a sky coverage of 100.206 deg^2 spanning the redshift range [0.0, 2.0]. Mock galaxies are characterised by angular position, redshift, comoving distances (according to the cosmological model of the parent N-body simulation) and several observational properties among which the luminosity of the H_α line. The cubic sample used for these tests consists in a volume limited sample extracted along the line of sight from the lightcone at redshift 0.75.

The code-by-code comparison has been performed considering the partial output (interpolated density field onto a grid, its Fourier transform, the MAS corrections, the shot-noise correction) as well as the final one in the form of spherically averaged power spectrum. To illustrate the outcome of the contest we focus on the final output only. Figure 3.1 shows the output of the three codes (B, S and C) when the same density field is considered (i.e. skipping the mass interpolation step). The upper panels show the monopole power spectra of the three codes whereas the bottom panels illustrate the relative % differences between the B-code and S-code. The difference between our code (C-code) and the B-code are not shown in the figure: they are below 10^{-10}% and probably purely numerical. Plots on the right panel show the effect of the MAS (CIC) correction. Clearly the relative agreement among the codes is excellent. Differences are below 10^{-5}%. The agreement is also excellent when, starting from the same distribution of objects rather than the same density field, we take into account of the impact of the CIC interpolation schemes.

Based on this result we have decided to proceed with the analysis using our own code.

[1] The catalogs are available at http://astro.dur.ac.uk/$\sim$ 40qra/lightcones/EUCLID/

Figure 3.1: Cross comparison of the three FKP codes considered in the contest. Upper panel: power spectrum. Lower panel: relative differences (in %). In the upper panel different colours are used for the various codes, as illustrated by the captions. B-codes: Balaguera Antolinez code; S-code: Percival-Samushia code; C-code: our code. All codes have been applied to the same density field. The lower panel shows the relative differences between B and S code. Differences between our code and B-code are below 10^{-10}%. *Plot courtesy of A.Balaguera Antolinez.*



3.2 Tests with BASICC simulation: real space

The previous test used a distribution of objects within a cubic box. However, here we are interested in the power spectrum of very wide and yet not completely all-sky survey with a well defined flux limit and incomplete sky coverage. Both effects can be encoded in the window function, that we have introduced in the previous Chapter 2. Here we adopt this option in order to account simultaneously for these effects. The combined effect of these selections can be evaluated and eventually accounted for by means of mock catalogs of galaxies to which we apply the same selection effects as the 2MRS catalog. The assessment is then obtained by comparing the power spectrum of the ideal, i.e. with no selection applied, galaxy distribution to the one of the realistic mock sample in which the selection has been properly quantified. A statistically sound estimate of the effect requires a large number of mock catalogs. Therefore the parent catalog needs to be large and dense enough to allow a reliable estimate of the reference power spectrum and to extract a sufficiently large number of mock catalogs. These are quite demanding constraints that, in practice, will not allow an accurate modelling of the 2MRS selection. Therefore, in this Chapter, we aim at a first assessment of the effect of the window function, mainly aimed at spotting possible systematic errors and identifying the range of scales suitable for clustering analysis in the 2MRS, and postpone a more precise assessment of the 2MRS random errors to the next Chapter in which we use a more realistic set of mock samples.

3.2.1 Parent simulation and mock catalogs

Mock galaxy catalogs for the analysis presented here were extracted from the high resolution N-body simulation BASICC, acronymous for Baryon Acoustic Simulation performed at the ICC (Angulo et al., 2008) using the publicly available GADGET-2 code (Springel 2005). Initial conditions were set up from a Gaussian realisation of the CAMB [2] (Lewis et al., 2000) power spectrum of density fluctuations. The gals-like particle distribution was then evolved using the Zel'dovich approximation up to $z = 60$. The cosmological model adopted is the same as the Millennium Simulation: ΛCDM cosmology with $\Omega_M = 0.25$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$ and $h = 0.73$[3].

The BASICC simulation is dark matter only and consists of 1448^3 particles with mass $m = 5.4910^{10} h^{-1} M_{solar}$ loaded within a cubic volume of side 1340 h^{-1} Mpc. Outputs are stored at different redshifts. We only consider the one corresponding to $z = 0$. Dark matter haloes are identified using the Friend of Friend algorithm with linking length equal to 1/5th of

[2] camb.info/

[3] h represents the Hubble constant in unity of $100 Kms^{-1} Mpc^{-1}$

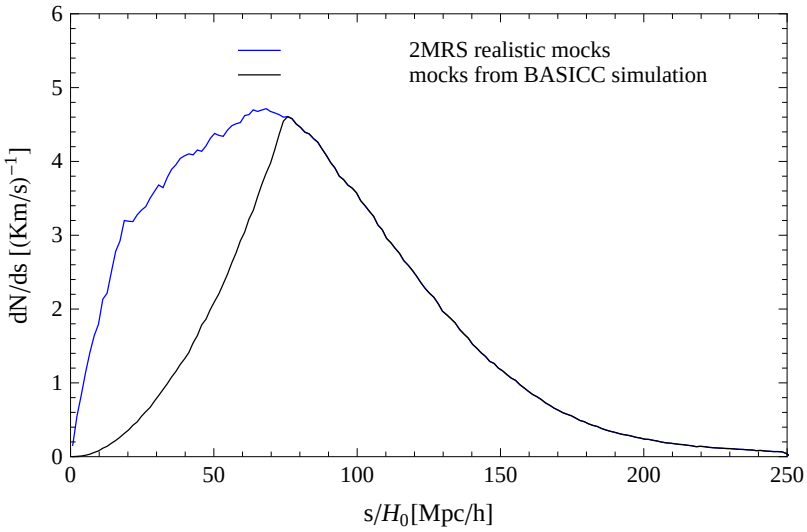
the mean interparticle separation. Since we do not aim at realistic mocks but only at assessing the impact of the window function we make the simplistic assumption that each halo contains only one 2MRS galaxy, sitting at its center, and we build 2MRS mocks by considering all dark matter halos with more then 10 particles which allows to have enough objects (17258579 in total corresponding to a number density of $7.2 * 10^{-3}\mathrm{Mpc}^{-3}h^3$) to match the number density of 2MRS galaxies. We notice that with a threshold of 10 particles per halo we expect that a non negligible fraction of these haloes will be simply a temporary clump of unbounded objects rather than genuine virialized structures. This would affect the estimate of the true power spectrum. However, this is not a concern here where we rather aim at assessing the relative difference between power spectra.

We divide the computational cube at $z = 0$ into 125 not overlapping smaller cubes with size 256 h^{-1} Mpc now containing dark matter haloes only. Small variations in the mean number density in each of the cubes reflects the effect of cosmic variance. For the aim of this analysis each catalog is characterised by its comoving coordinates only, ignoring peculiar velocities. As a result the clustering analysis presented in this Section is performed in real space. Redshift distortions will be accounted for in the next Section. As we describe in detail in the next Chapter, the 2MRS window function is quite simple: the 2MRS is an almost all sky survey with no objects within a latitude of ~ 5 degrees from the Galactic plane. The region within that latitude is dubbed Zone of Avoidance. To mimic the effect of the 2MRS window function we extract, from each sub-cube, all haloes within a distance of 128 h^{-1} Mpc from the center. This radius, within which the mean number density of 2MRS galaxies is $\sim 10^{-3}\mathrm{Mpc}^{-3}h^3$, is close to the maximum distance that we shall consider in the real catalog analysis. We then randomly dilute the distribution of haloes to match the dN/ds (the redshift $s = cz$ is expressed in km s^{-1}) of the more realistic 2MRS mocks that we will describe and extensively use in the next Chapter. The result of this second step is illustrated in fig. 3.2. The blue line shows the mean dN/ds over 135 *realistic* mocks that here we use to average out the effect of cosmic structures and obtain a dN/ds representative of a generic galaxy sample. The mean dN/ds estimated in this way is sufficient for the purposes of these preliminary tests, and we don't adopt the F/T method here. The mean number density of dark matter haloes is similar but smaller than that of the 2MRS galaxies in the *realistic* mocks. As a result, when extracting random subsamples we undersample objects near the center of the spheres. To minimise the effect we keep all the haloes within 7000 km/s and randomly select the others according to the reference dN/ds. The result is the black curve, in which the discrepancy below 7000 km/s illustrates the effect of the mismatch in the mean number densities. However, this mismatch is not large (it affects a limited number of nearby objects) and its effect is to overestimate the impact of the window function and artificially

increases the shot noise level.

As a final realistic touch, we model the Zone of Avoidance of the 2MRS by excluding all objects within ±5 degrees from the Galactic plane that we define by laying to a set of spherical coordinates centred at the center of the cube.

Figure 3.2: dN/ds of 135 2MRS *realistic* mocks (blue curve) vs. dN/ds of the mocks extracted from the BASICC simulation.



The main scope of this analysis is to assess systematic errors driven by the window function. For this purpose and to assess the impact of the various selection effects, we compute and then compare different types of power spectra listed below:

1. The reference power spectrum obtained considering all haloes within the BASICC simulation cube with no selection applied: P_0.

2. The mean power spectrum computed from haloes in the 125 sub-cubes when no selection is applied: P_C.

3. The mean power spectrum computed from haloes in the 125 spheres extracted from the sub-cubes when no selection is applied: P_S. The corresponding window function is a top hat sphere.

4. The mean power spectrum computed from haloes in the 125 spheres with the dN/ds selection: P_{SR}. The corresponding window function is spherically symmetric with a radial profile dN/ds in fig. 3.2.

5. The mean power spectrum computed from haloes in the 125 spheres with the dN/ds selection and the Zone of Avoidance: P_{SRZ}. The corresponding window function is axis-symmetric with a $\pm 5^o$ disk-like void and the rest in which the halo radial distribution is described by the dN/ds.

In all cases, the power spectrum was estimated using a random sample of objects with the same window function as the halo distribution.

In Table 3.1 we summarise the characteristics of the cubic boxes used to compute the power spectra and their grids.

Table 3.1: Characteristics of the computational box for each subsample and step. $P(k)$=type of power spectrum. N = number of grid points. L = side of the computational box in h^{-1}Mpc. $\Delta k = k$-bin in h Mpc^{-1}.

$P(k)$	N	L	Δk
P_0	1024^3	1340	0.0094
All the others	256^3	256	0.024

Table 3.2 shows the characteristics of the various mocks subsamples for which the power spectra have been computed.

Table 3.2: Characteristics of the BASICC subsamples. $P(k)$=type of power spectrum. L = side of the computational box in h^{-1}Mpc. R = maximum distance from the center in h^{-1}Mpc. $n(\bar{s})$ mean number density in h^3 Mpc^{-3}. ZoA = Type of the zone of avoidance (degrees above a below the Galactic Plane

$P(k)$	L	R	$n(s)$	ZoA
P_0	1340	n.d.	$7.2 * 10^{-3}$	$\pm 0°$
P_C	256	n.d.	$7.2 * 10^{-3}$	$\pm 0°$
P_S	256	128	$7.2 * 10^{-3}$	$\pm 0°$
P_{SR}	256	128	2MRS-like dN/ds	$\pm 0°$
P_{SRZ}	256	128	2MRS-like dN/ds	$\pm 5°$

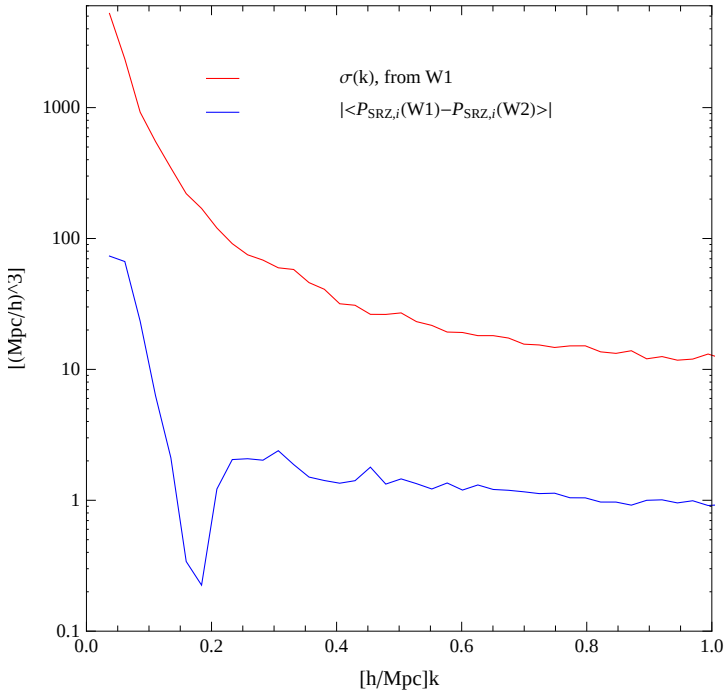
3.2.2 Assessing the impact of the weighting scheme

Each power spectrum has been computed with the two different weighting schemes described in Chapter 2: the first method, dubbed W1, consists in weighting each halo with the inverse of the local mean number density , the

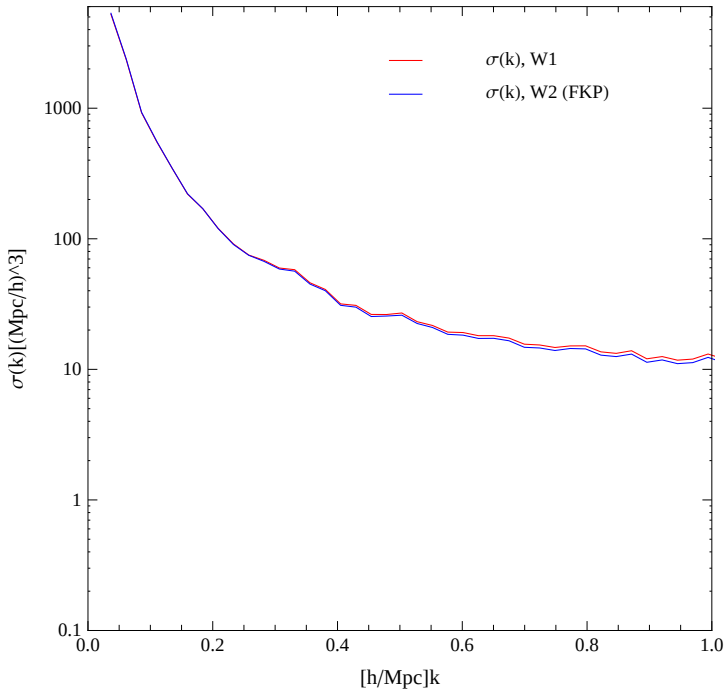
second method, dubbed W2, consists in using the standard FKP weights. The aim of this Section is to evaluate possible differences between the two methods in order to assess the best method to be used in the 2MRS catalog analysis. FKP weights, explicited in equation 2.33, depend on the power spectrum and, in principle, should be computed iteratively. However, this has little impact on the final result and it is common practice to adopt the value of a reasonable (i.e. ΛCDM) power spectrum evaluated at the scale relevant for the analysis. Our 2MRS analysis will be limited to the wavelength range $\sim 0.03 - 0.4\ h$ Mpc^{-1} to compromise between cosmic variance and nonlinear effects . We set $P \equiv 8000h^3$ Mpc^{-3} that corresponds to the mean value of a ΛCDM power spectrum at $z = 0$ in the approximate k range [0.05,0.15] h Mpc^{-1}. We have verified that the results do not change appreciably when using values larger than $5000h^3$ Mpc^{-3}.

Fig. 3.3(a) shows in blue the mean difference between P_{SRZ} computed with W1 and P_{SRZ} computed with W2, $| < P_{SRZ,i}(W1) - P_{SRZ,i}(W2) > |$, where i varies from 1 to 125 and the mean is over the all 125 subsamples. The *rms* scatter of the measures $< (P_{SRZ,i} - < P_{SRZ,i} >)^2 >^{1/2}$ from W1 is plotted in red.

Figure 3.3(a) shows that systematic differences between the two methods are well below the scatter of the estimator and that the scatter of both methods are very similar (overlapping red and blue curves in Fig. 3.3(b)). This result shows that, in this case, the two weighting schemes are quite similar. Which is not surprising since the mean number density, and therefore the quantity $\bar{n}P$, is sufficiently large to guarantee that the FKP weights are close to the asymptotic value $1/\bar{n}$, i.e. the W1 weighting scheme. From here on we decide to use the standard FKP weighting scheme for the 2MRS analysis because FKP weights, built to minimize variance of the estimator, give a slightly smaller scatter as can be seen in fig. 3.3(b), as expected.



(a)



(b)

Figure 3.3: Upper plot: mean discrepancy between the power spectra obtained with method 1 and method 2 (blue curve) vs. the *rms* scatter among the mocks of the power spectrum evaluated with method 1 (red curve). Bottom plot: *rms* scatter of method 1 (red) vs. *rms* scatter of method 2 (blue).

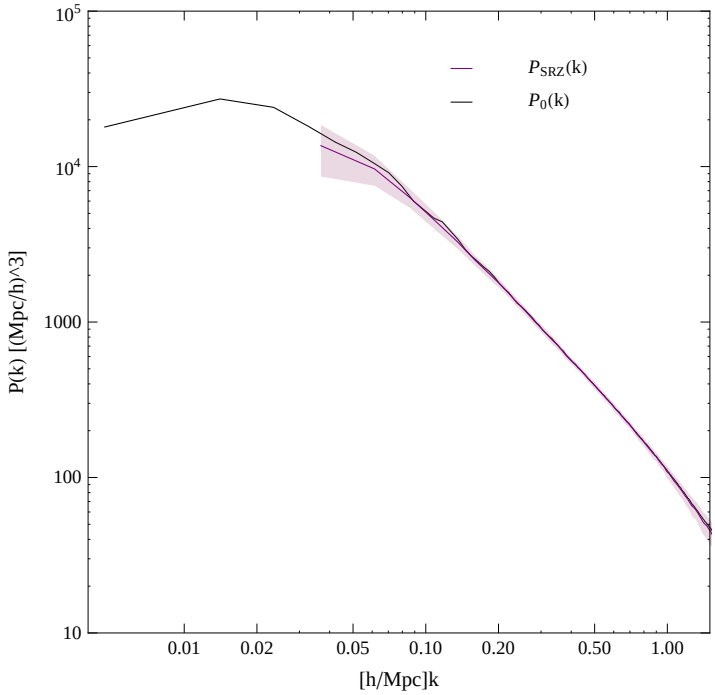
3.2.3 Assessing window-induced systematic errors

Limited sky coverage and flux cuts can, if not properly accounted for, induce systematic errors in the power spectrum estimate. Their amplitude depends on the geometry of the survey and the number density of the objects. The effect of the latter can be minimised by limiting the depth of the survey and by using an appropriate weighting scheme, as shown in the previous Section. The former is also supposed to have a limited impact thanks to the simple geometry of 2MRS that, because of its very large sky coverage, is close to spherical. The impact of the window function is illustrated in fig. 3.4 where the reference power spectrum $P_0(k)$ (black curve) is compared with the mean power spectrum measured in the 125 spheres $P_{SRZ}(k)$ (purple solid) and its *rms* scatter (purple shaded area). All spheres are embedded in a cube with no objects beyond the sphere radius. Power spectra were measured with the FKP estimator and the mean density used in the weighting scheme is the one measured within each sphere (for the $P_{SRZ}(k)$ case). The fact that the minimum wavenumber k_{min} are different for the two spectra simply reflects the different side of the two boxes used: $k_{min} = 2\pi/256h$ Mpc^{-1} for $P_{SRZ}(k)$ and $k_{min} = 2\pi/1340h$ Mpc^{-1} for $P_0(k)$. For the same reason the resolution is different in the two cases. Since for $P_0(k)$ we set the shell size in which the power spectrum is computed $\Delta k = 2\pi/1340h$ Mpc^{-1} whereas for the $P_{SRZ}(k)$ the shell size is $\Delta k = 2\pi/256h$ Mpc^{-1}, for the purpose of comparing the two spectra, we degrade the resolution of $P_0(k)$ to that of $P_{SRZ}(k)$ by averaging over the same Δk shells.

At small k_s the power $P_{SRZ}(k)$ is systematically below $P_0(k)$, which is reasonable since in this case the volume sampled is smaller. However, the significance of the mismatch is well below 1 σ defined from the *rms* mock scatter shown in the plot.

Fig.3.5 illustrates the individual effect of all the contribution to the window function. The top-left panel shows the relative difference between $P_0(k)$ and $P_C(k)$, averaged over the sub-cubes. The only mismatch is at small values of k and derives from the fact that, on those scales, the BASICC cube contains more modes that those in the 125 non-overlapping cubes. On the top right we show the effect of considering a spherical, rather than a cubic volume. The red curve shows the relative difference between $P_0(k)$ and $P_S(k)$ averaged over all spheres. Discrepancies are still small but larger than in the previous case. The fact that the volume covered by the sphere is smaller than that of the corresponding cube is reflected in the loss of large scale power, which is now more prominent than in the top-left panel. The bottom-left plot shows the effect of diluting the spherical sample to match the observed dN/ds. FKP weights compensate for the loss of objects due to the radially-varying galaxy distribution. Finally, the bottom-right panel shows the cumulative effects of all selections, including that of the Zone of

Figure 3.4: $P_0(k)$ estimated with FKP (black, solid curve) vs. P_{SRZ} (k) (purple, solid) and its *rms* scatter from the mocks.



Avoidance whose main effect is to enhance the loss of large scale power. The total systematic error induced by the window function is about 10 % at $k \sim 2k_{min}$.

Fig. 3.6 shows the ratio between the estimated systematic error due to the survey window and the statistical estimated error, $\sigma_{SRZ}(k)$. The former is computed as shown before, as difference between "reference" power spectrum, $P_0(k)$, and mean power spectrum of the sampled spheres, characterized by geometrical and selection properties of the real survey, $P_{SRZ}(k)$. The latter is computed as standard deviation of the $P_{SRZ}(k)$ respect to the mean.

This ratio is smaller than one at any k , meaning that the effect of the mask on power will be in any case smaller than the precision of the measure, but can not be considered negligible respect to the statistical uncertainties. It's more than 30% around k_{min}, and in general around one third of the scatter at the smallest sampled k_s. This suggest to correct for this systematic. The strategy to account for systematic errors induced by the 2MRS is outlined in Section 3.2.5.

Figure 3.5: Relative errors induced by the various selection effects in 2MRS. **Top left:** effect of finite volume. Relative difference between $P_0(k)$ and $P_C(k)$, averaged over the sub-cubes. **Top right:** effect of the spherical window. Relative difference between $P_0(k)$ and $P_S(k)$ averaged over all spheres. **Bottom left:** effect of flux limit and its dN/ds. Relative difference between $P_0(k)$ and $P_{SR}(k)$ averaged over all spheres. **Bottom right:** cumulative effect of volume, flux and Zone of Avoidance. Relative difference between $P_0(k)$ and $P_{SRZ}(k)$ averaged over all spheres.

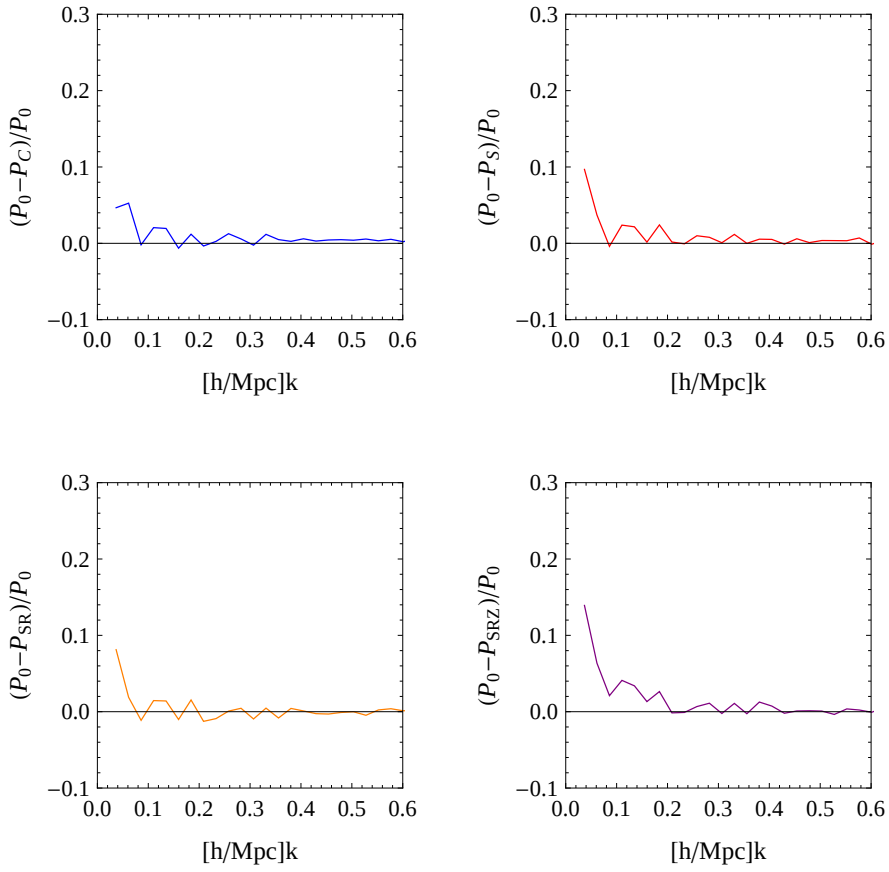
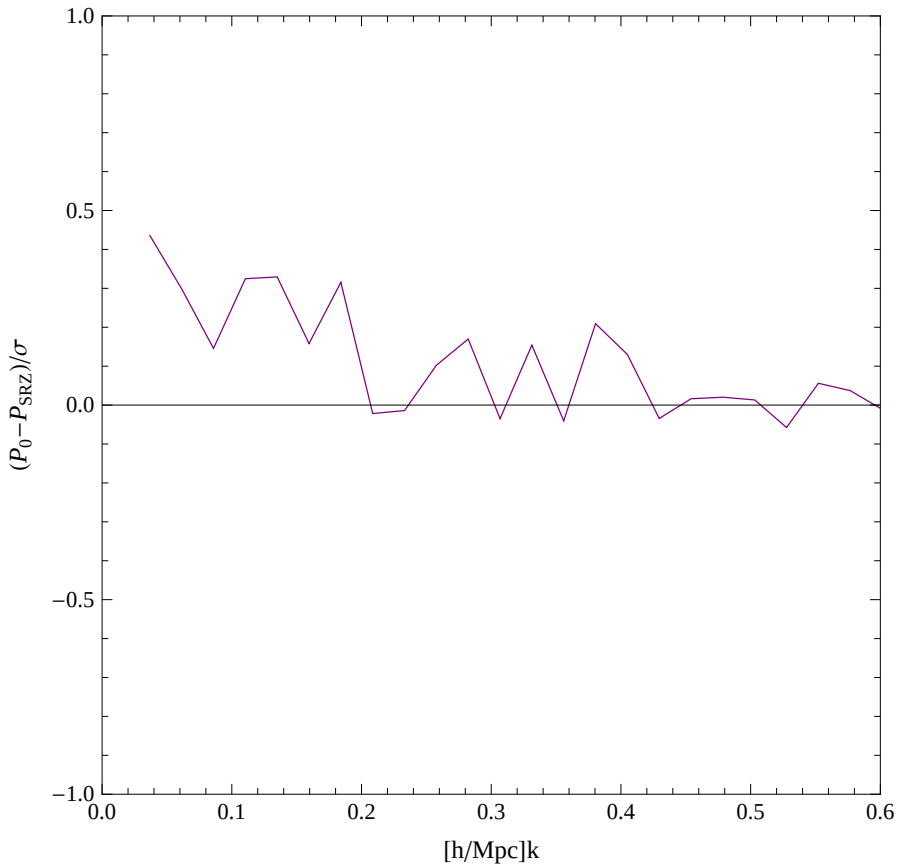


Figure 3.6: Ratio between the estimated systematic error due to the survey window and the statistical estimated error, $\sigma_{SRZ}(k)$.



3.2.4 Setting the $k-$range of the analysis

The range of wavenumbers appropriate for our analysis is determined by both the scientific goal of our analysis and the expected accuracy. At large scales the dominant source of error is cosmic variance. At small scales is shot noise. In addition, at small scales the cosmological interpretation of the results is hampered by nonlinear effects. Let us consider the two scales and their errors separately.

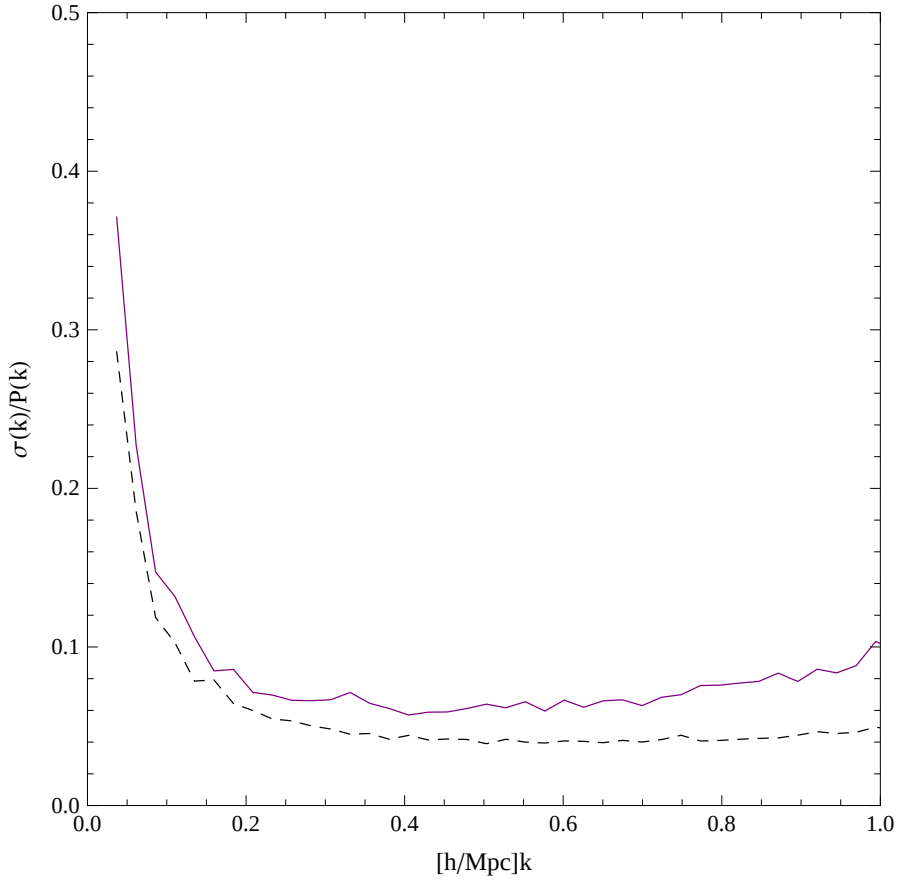
As anticipated in Section 2.1.2, the smallest accessible scales are those corresponding to the 65% of the Nyquist frequency that in our case is: $\pi(N^{1/3})/L = 3.14h$ Mpc^{-1}, where N is the number of grid points of the computational box, and L the side of the box (here $N = 256^3$ and $L = 256$ h^{-1} Mpc, see Table 3.1). At $z = 0$ this scale is well within the nonlinear regime. Therefore the maximum wavenumber, k_{max}, to which we can push our analysis will be set by the scale where nonlinear effects become dominant and difficult to model. We shall address this point in the next Chapter. Here we anticipate that the maximum non linear scale that we use in the analysis is $k_{NL} \sim 0.4 - 0.5h$/Mpc. To determine the maximum scale we compute relative *rms* scatter P_{SRZ} to the mean value, defined in eq. 2.48.

This noise to signal ratio is shown in fig.3.7 as a function of k. The continuous curve shows the ratio estimated from the mocks. The dashed one is the theoretical prediction from FKP (eq.2.47). This expression assumes independent Fourier modes and Gaussian random field. None of these assumption is fully valid in our case. First, the geometry of the sample and the nonlinear evolution mix the various Fourier modes. In addition, a nonlinear evolved density field is also non Gaussian. These effects justify the fact that the actual errors in the plot are larger than FKP predictions that assume independent Fourier modes. Indeed, because of the nonlinear evolution or indeed any deviations from Gaussianity, non-vanishing off-diagonal terms arise in the power spectrum covariance matrix due to the mode coupling. These contributions are fully described by the trispectrum, whereas contributions from the bispectrum vanish (Scoccimarro et al. 1999). These additional contributions to the power spectrum variance include the so called beat coupling effect. In our case the fact that we use the mean number density in each sphere should guarantee that these off-diagonal terms are small, but not completely negligible as demonstrated by de Putter et al. (2012).

The results indicate that the relative random errors can be kept at the 20% level for $k_{min} \geq 0.06$ h Mpc^{-1}, increasing to $20 - 30\%$ in the $0.03 < k < 0.06h$ h Mpc^{-1} interval.

To summarize, we shall restrict the power spectrum analysis of the 2MRS to the interval $[k_{min}, k_{NL}]$, where $k_{NL} \sim 0.5$ and $k_{min} \sim 0.06$, although we occasionally push the analysis out to $k \sim 0.03$.

Figure 3.7: Noise to signal ratio obtained from the *rms* scatter of $P_{SRZ}(k)$ (solid curve) vs. the theoretical expression from FKP (dashed).



3.2.5 Modelling the effect of the window

In this Section we assess the ability to account for the effect of the window function and to recover the underlying power spectrum. We have seen in subsection 3.2.3 that the systematic window induced error is around 40% - 30% of the *rms* scatter for the smallest sampled k's so we have to reduce it taking into account of the convolution with the window function.

We do this in two steps. First we consider the case of a simple window function that can be described analytically and assess the numerical accuracy of the convolution integral. This is a relevant test since, as we have said, the 2MRS window function is close to a spherical top hat, for which an analytic expression is available. Then we repeat the exercise using the numerical estimate of the same window function to check whether our numerical modelling of the window function introduces additional errors.

This test is performed with the same samples described in section 3.2.1 and extracted from the BASICC simulation.

To assess the accuracy of the convolution model we use $P_0(k)$ as the "true" power spectrum and an analytic or numerical model for the window function, as described below.

For the first test we consider purely spherical samples with measured power spectrum $P_S(k)$. In this case the window function is a simple spherical top hat with radius 128 Mpc h^{-1} and its Fourier transform is:

$$W_{th}(k) = ((4\pi/k^3)\frac{(sin(kR) - kRcos(kR))^2}{((4/3)\pi R^3)} \tag{3.1}$$

Then we perform the convolution integral 2.52 using $P_0(k)$ as underlying power spectrum, as anticipated, and compare the result with the measured $P_S(k)$, the mean power spectrum of the 125 spherical samples.

To numerically estimate the convolution 2.52 we use the numerical estimator 3.2. We verify that the effect of the integral constraint is negligible in the $k-$range explored so that our estimate of the convolved power spectrum is:

$$P_{conv}(k_i) = \sum_j W[k_i][k_j]P_0(k_j). \tag{3.2}$$

The elements of the convolution matrix $W[k_i][k_j]$ are computed as in eq. 3.3:

$$W[k_i][k_j] = \frac{1}{2\pi^2}\int\int W_{th}(\epsilon)\epsilon^2\Theta(r_\epsilon, k_j)\sin(\theta)d\theta d\epsilon \tag{3.3}$$

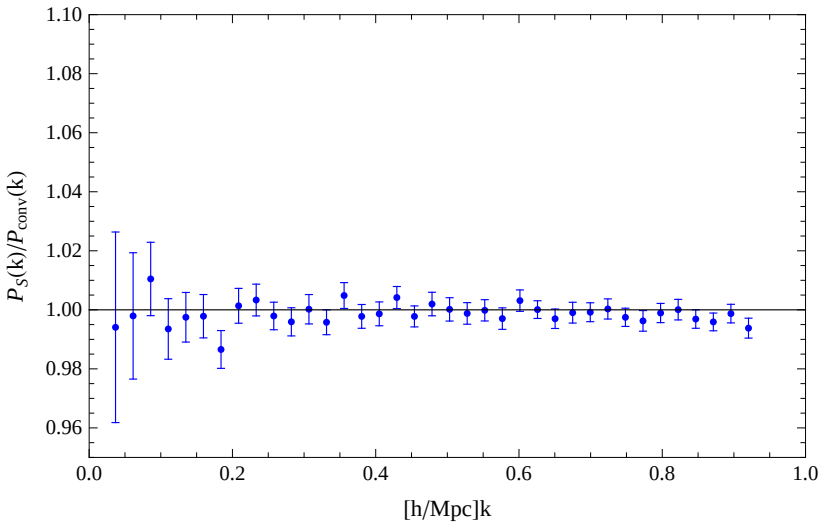
where $W_{th}(\epsilon)$ corresponds to the analytical expression 3.1.

The numerical estimator 3.2 is computed for discrete k-bins, Δk_i and Δk_j. The size of Δk_i corresponds to the resolution of the measured power spectrum , in our case $P_S(k)$, with which we want to compare the convolved power spectrum $P_{conv}(k_i)$. In our case $\Delta k_i = 2\pi/L = 0.024\ h$ Mpc^{-1}, where

$L = 256$ Mpc h^{-1} is the size of the boxes in which $P_S(k)$ is computed, and that corresponds to the diameter of the spherical samples. This resolution can't be improved because of cosmic variance. Δk_j is typically related to the resolution with which the window function is measured. We adopt $\Delta k_j = 0.0047$ h Mpc^{-1} both for the analytical top hat and the measured window function. The power spectrum $P_0(k_j)$ is truncated at $k = 1.79$ h Mpc^{-1}. Beyond that scale we can't neglect aliasing, being the Nyquist wavenumber equal to 2.4 h Mpc^{-1} (see table 3.1 for details on box size and number of grid points).

The comparison of the convolved power spectrum with the mean $P_S(k)$ is in fig.3.8. Errorbars reflect the *rms* scatter of P_S among the mocks, normalized by $P_{conv}(k)$. The result shows that our convolution procedure does not introduce systematic errors down to k_{min}, where the window function effects are expected to be larger.

Figure 3.8: Convolution with analytic spherical top hat (R=128 Mpc/h) and comparison with $P_S(k)$.



In the second test we have repeated the same exercise using again $P_S(k)$ and a numerically estimated spherically averaged window function, $W_s(k)$, that we have computed as follows. We use the random catalog used to measure the power spectrum of the constant mean number density spheres: it consists in a top hat sphere with radius R equal to 128 Mpc/h filled with random points. We compute $W_s(k)$ from this distribution of random objects in the same way we compute the power spectrum of the spheres extracted from the simulation: we interpolate the density field on grid 1024^3, with CIC mass assignement scheme, compute Fourier coefficients with FFT technique and then spherical average in k bins. In order to improve the resolution

of $W_s(k)$, the computation has been carried on in a box of side L equal to 1340 Mpc/h greater than the spherical volume sample. The sampling of modes is finer and the window is computed in bins $\Delta k = 2\pi/1340h$ /Mpc $= 0.0047$. We deduced that this resolution is sufficient in the sense that increasing resolution does not affect the results. The spherically averaged window function $W_s(k)$ is truncated at $k \simeq 65\%$ of the Nyquist wavenumber k_{Ny} to avoid aliasing impact on the measure. The comparison between the analytical top hat function and its numerical estimate obtained from the top hat spherical distribution of points is shown in fig. 3.9. As can be seen the measured window (red curve) matches quite well the analytical top hat (blue curve). The numerical estimate in fig. 3.9 shows some discontinuity that can affect the numerical integration. To avoid this we fit a spline to the spherically averaged window function and obtain a more regular function. The result of the convolution is then compared to the expected power spectrum $P_S(k)$ in fig. 3.10 which is analogous to fig. 3.8 and we find that the accuracy of the numerically estimated window function is good enough to avoid introducing systematic errors except, perhaps at the smallest scales.

Figure 3.9: Top hat window function: analytical function (blue) vs. numerical estimate (red).

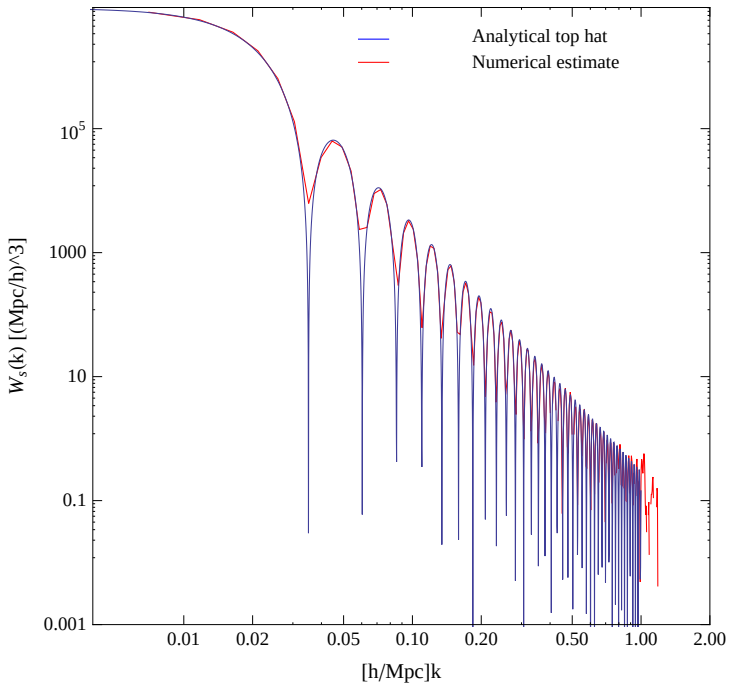
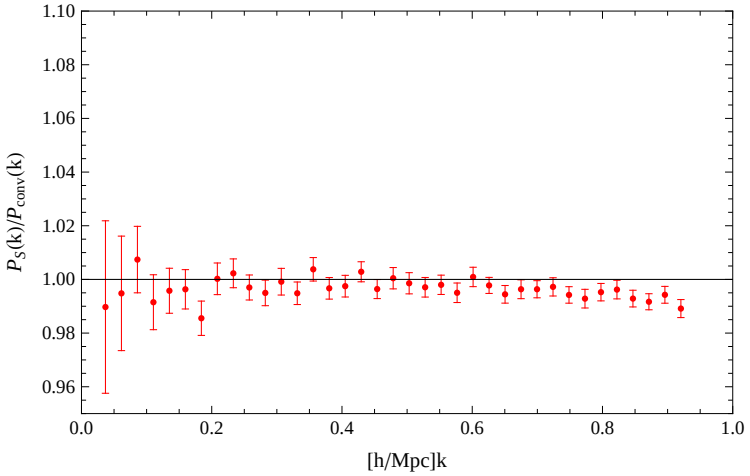


Figure 3.10: Convolution with the top hat computed window and comparison with $P_S(k)$.



Finally, we focused on the more realistic case $P_{SRZ}(k)$. We have repeated the same procedure to compute the new window function that now accounts for the dN/ds, FKP weights and Zone of Avoidance that we plot in fig. 3.11 togheter with the top hat one. The similarity between the two window functions shows what we anticipated: the 2MRS window function is quite simple and not too different from the cases of an ideal all-sky survey. The result of the convolution are compared with the expected $P_{SRZ}(k)$ in fig. 3.12 which is analogous to 3.8. No systematic deviations from unity are detected for $k < 0.7\ h$ Mpc^{-1}. Discrepancies of 1%-2% are seen at smaller scales at which, however, we do not perform our analysis. The discrepancies are probably due to the fact that we have truncated the window function at $k_{max} = 1.2\ h$ Mpc^{-1} to avoid aliasing. We don't see any systematic when the analytic top-hat window function is used (fig.3.8), for which, thanks to the fact that aliasing is not an issue in this case, we set $k_{max} = 2.0\ h$ Mpc^{-1}. We verified that reducing progressively the value of k_{max} and performing the convolution integral using both the analytical expression for the spherical top hat and the numerical one, the ratio $P_S(k)/P_{conv}$ decreases progressively, especially at large k_s. This possibly explains the differences between fig. 3.8 and fig. 3.10. The discrepancies at small scales are more pronounced when the realistic window function computed numerically is used (fig. 3.12), although, in this case, we have truncated the window function at the same maximum wavenumber $k_{max} = 1.2$ Mpc^{-1} as in the case of the numerical estimate of a top hat window function (fig. 3.10). Although very similar, the realistic window function don't coincide with with a simple top hat window, and this could justify the differences between fig. 3.10 and 3.12.

Figure 3.11: Realistic computed window.

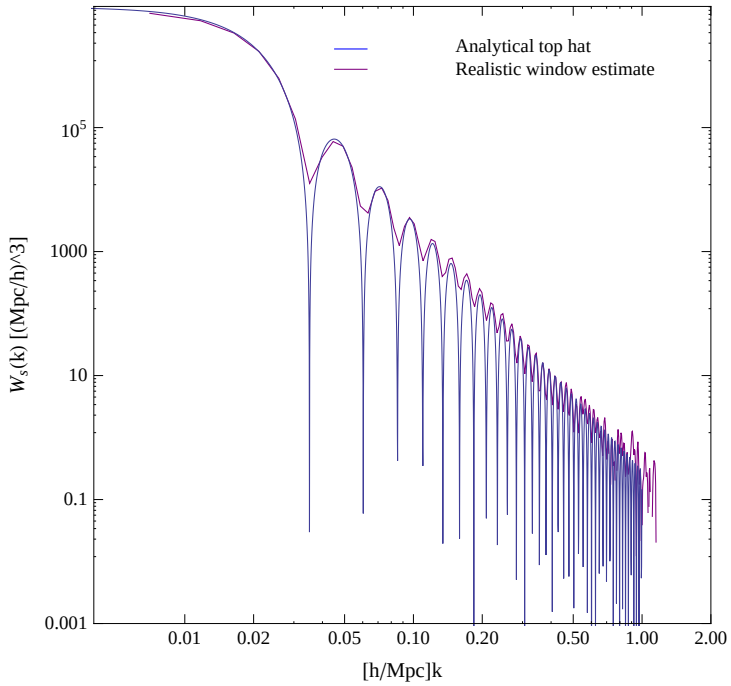
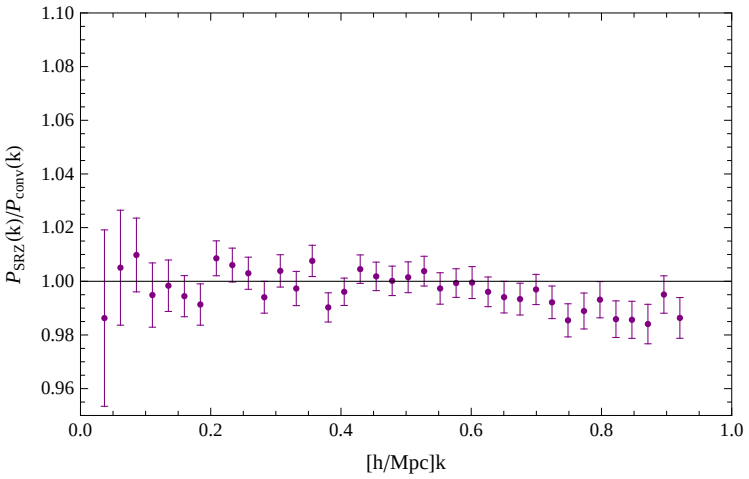


Figure 3.12: Convolution with computed window and comparison with $P_{SRZ}(k)$.



3.3 Tests with BASICC simulation: redshift space

In this Section we account for redshift space distortions. As described in section 1.3.3, redshift distortions modify the amplitude (in the linear regime) and the shape (in the non linear one) of the power spectrum and need to properly accounted for when comparing model with data. Our goal is to verify our ability to model redshift space distortions (dubbed $z-$distortions) and to assess the performance of the convolution technique when $z-$distortions are present. For this purpose we measure the monopole of the halo power spectrum in redshift space. The halo positions in $z-$space, d_s, where obtained by adding the line of sight peculiar velocity component, v, to their comoving distance d: $d_s = d + v$, both expressed in Mpc h^{-1} at $z = 0$. The haloes were selected using the same criteria as in the real space case. And, in analogy with the procedure described in section 3.2.1, we obtain 3 types of power spectra:

1. $P_{0,z}$. This is analogous to P_0 and represents the $P(k)$ of almost all the haloes in the computational cube. The observer position is set at the center of the cube of the simulation. Peculiar velocities were added along the line of sight with respect to the central observer. This has the effect of blurring the edges of the original cube, as a consequence we consider only the objects within a smaller cube , with size L=1200 Mpch^{-1} centered at the center of the parent cube.

2. $P_{C,z}$. This is the analogous of P_C and represents the mean power spectrum of 125 subcubes. Subcubes in $z-$space were obtained by selecting a central observer, adding radial velocities to each object distance and select all haloes in a $L = 256$Mpc/h cube.

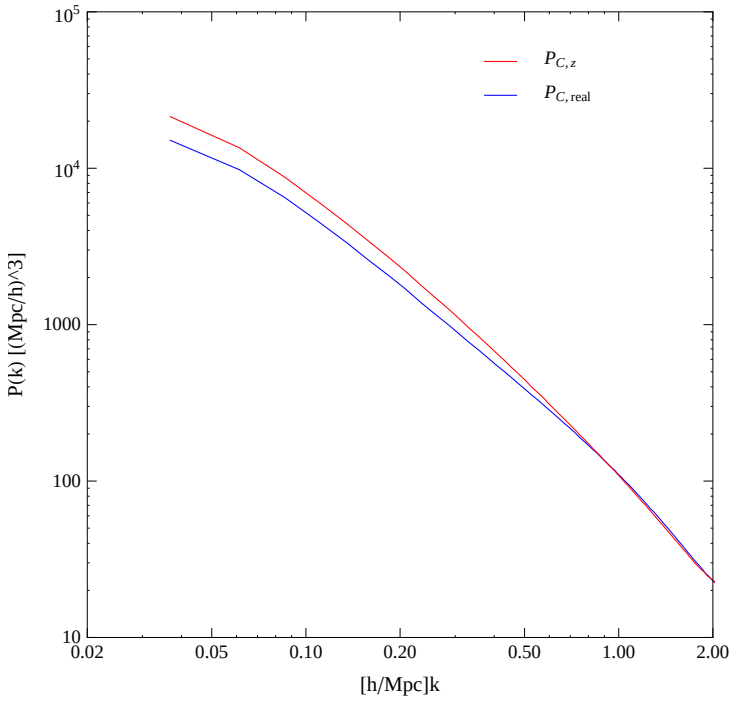
3. $P_{SRZ,z}$ is the analogous of P_{SRZ} and the corresponding objects were obtained by extracting spherical samples from each of the 125 subcubes in $z-$space. The spherical samples match the dN/ds of the 2MRS mocks shown in fig. 3.2 and contain a Zone of Avoidance of $\pm 5^0$ respect to a plane passing through the center of the sphere.

For each power spectrum mentioned above we have the corresponding power spectrum in real space: $P_{0,real}$, $P_{C,real}$ and $P_{SRZ,real}$. The last two are already shown in Section 3.2; here we're adding the subscript *real* to distinguish them from the corresponding redshift space ones.

Fig. 3.13 shows in red the mean power spectrum of the 125 cubes with constant mean number density in redshift space, $P_{C,z}$. In blue, for comparison, there's the mean power spectrum of the 125 cubes in real space $P_{C,real}$. The mean of the 125 ratios between redshift space and real space power spectrum, $< P_{C_i,z}(k)/P_{C_i,real}(k) >$, where the subscript i indicates

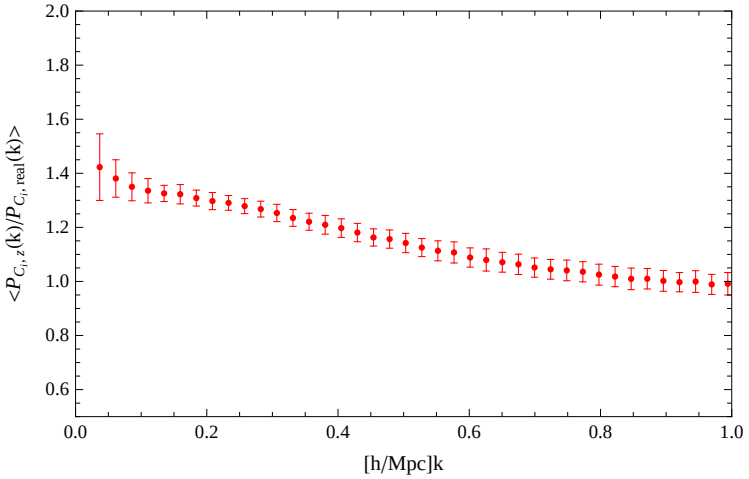
one of the 125 samples, with error bars given by the scatter over the 125 cubes, is shown in fig 3.14.

Figure 3.13: Mean power spectrum of 125 redshift space cubes exctracted from the BASICC simulation, $P_{C,z}$, with side L=256h/Mpc and constant density, in red. Mean power spectrum of 125 real space cubes exctracted from the BASICC simulation, $P_{C,real}$, side L=256h/Mpc and constant density, in blue.



As expected on large scales the redshift space power spectrum is higher than the real space one due to coherent velocities in the linear regime (Kaiser 1987). On smaller scales, non linear velocities have the opposite effect of reducing the power. In fact, as pointed out by Angulo et al. (2008), (see fig. 9, left panel, in that paper), the small scale damping in this case in mostly due to halo merging rather than non linear motions since the halo finder is such that only virialized haloes and not their sub-structures are selected. As a consequence highly non linear effects represented by sub-halo motions within virialized structures are artificially smoothed out, leaving no imprint in the power spectrum. We shall come back to this point when we consider realistic mock galaxy catalogs where, instead, non linear motions are present.

Figure 3.14: Mean of the ratios redshift space real space power spectrum of 125 cubes with constant density, $< P_{C_i,z}/P_{C_i,real} >$.



3.3.1 Linear redshift space distortions

The first set of tests is carried on with the cubic subsamples with constant mean number density. These tests are aimed to check the ability of modelling redshift space distorsions before introducing the complication of the window function.

Linear theory predicts that the amplitude of power spectrum in redshift space is higher than the real space power because of the coherent bulk motions. As described in Section 1.3.3, the boost is scale independent and when averaged over all directions amount to $K(\beta) = 1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2$, quantity usually referred to as "Kaiser boost". This is the first ingredient in the $z-$distortion model and should be sufficient to reproduce the observed behaviour on large scales. To check whether this is true and the maximum scale at which linear theory holds, we estimate the Kaiser boost from the ratio between the real and the $z-$space power spectra, like in fig. 3.14, and compare β with the expected value which we get from the bias parameter previously estimated in real space and described below and the Ω_M value of the BASICC simulation. The best fit β values were obtained by minimizing the χ^2 difference between data and model, that in this case is simply a constant function, with error bars obtained from the mocks.

The linear bias b is obtained from the ratio between the $P_{0,real}$ and the linear matter power spectrum computed using the CAMB package, P_{CAMB}, for the cosmological parameters of the BASICC simulation. In practice we fit a constant function to the ratio $P_{0,real}/P_{CAMB}$ up to $k_{max} = 0.15h$ Mpc^{-1} to keep non linear effect small and we obtain b=0.86. The result of the

fit is very robust respect to variations of k_{max}: variations under $\sim 1\%$ are obtained when varying k_{max} in the range [0.05,0.15] h Mpc^{-1}.

We have obtained β by fitting a measured power spectrum with different models.

1. **MODEL** NL0: fitting $P_{0,z}$ with $P_{0,real}K(\beta)$.

   In this test we exploit the size of almost the full cube to extend the analysis well into the linear regime. The maximum scale probed here is ~ 5 times larger than the size of the 2MRS sample. Errors in this case are set equal to the variance of the FKP estimator (eq. 2.47). They contribute to both $P_{0,z}$ and $P_{0,real}$ and are propagated to the ratio $P_{0,z}/P_{0,real}$ ignoring the fact that taking the ratio eliminate the contribution of the cosmic variance to first order. The resolution is set by the fundamental mode of the box $\Delta k = 2\pi/1200$ h Mpc^{-1} and represents the size of the spherical shell used to estimate P_0.

2. **MODEL** L0: fitting $P_{0,z}$ with $P_{CAMB}b^2K(\beta)$.

   The aim of this test is to check the adequacy of using a linear model to the halo power spectrum in real space ($P_{CAMB}b^2$). Discrepancies with respect to the results obtained in the previous test would indicate departures from linear theory both in the evolution of density fluctuations and in the bias. The resolution is the same as in the previous test. Errors, indeed, are only contributed by $P_{0,z}$ and set equal to the variance of the estimator. No uncertainty is associated to the model P_{CAMB} and, in this respect, this is analogous to the case of the real data analysis.

3. **MODEL** NLC: fitting $P_{C,z}$ with $P_{C,real}K(\beta)$.

   This is analogous to the $P_{0,z}/P_{0,real}$ case but now we consider the average power spectra measured in the smaller cubes. The goal of this test is to assess the effect of reducing the $k-$range usable for the linear theory fit to the intent that can be probed with the 2MRS sample. Errors on $P_{C,z}$ and $P_{C,real}$ are obtained from the scatter among the cubes and propagated to their ratio. We note that the scatter of the ratio $< P_{C_i,z}/P_{C_i,real} >$ is about 4 times smaller, indicating that the cosmic variance dominates the error budget in this $k-$range, as expected. The resolution is $\Delta k = 2\pi/256$ h Mpc^{-1}, lower than in the P_0 case.

4. **MODEL** LC: fitting $P_{C,z}$ with $P_{CAMB}b^2K(\beta)$.

   This is analogous to model L0 and is aimed at assessing the impact of non linear effects in a sample with a size similar to that of the 2MRS catalog.

In all cases we have obtained β by fitting a constant function to the ratio between the estimated and the model spectrum.

Fig. 3.15 shows the results of these tests, each one of them indicated with a different symbol.

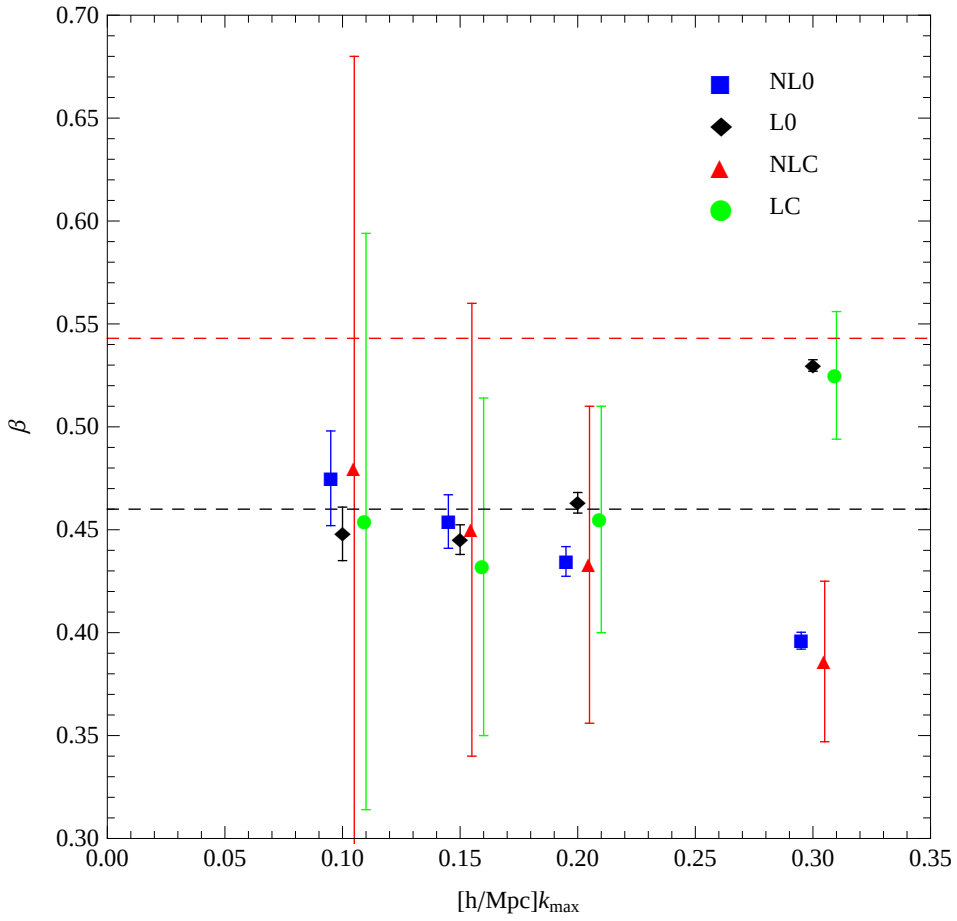


Figure 3.15: The best fit β values and their *rms* uncertainty for the various tests as a function of k_{max}. Black dashed line the β "reference" expected value. Red dashed line is β_{theory}=$\Omega_M^{0.55}/b$.

On the $y-$axis we show the best fit β value while, on the $x-$axis, we show the maximum value of k used to compute the χ^2, i.e. the fit is performed in a k-range $[k_{min}, k_{max}]$ where $k_{min} = 2\pi/L$ and k_{max} is the value displayed on the $x-$axis. To avoid overlapping the plot, the results of these tests that refer to the same k_{max} have been slightly shifted. The black dashed line, $\beta = 0.46$, represents the best fit value obtained in the $P_{0,z}/(P_{CAMB}b^2)$ test using $k_{max} = 0.05$ h Mpc^{-1}, not shown in the plot, well into the linear

regime. We take this as the reference β value. The horizontal red-dashed line is the expected value $\beta = \Omega_M^{0.55}/b$ where $\Omega_M = 0.25$ is the BASICC simulation value and $b = 0.86$. β values plotted in fig.3.15 and their *rms* uncertainties are listed in Table 3.3.

Table 3.3: The best fit β values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	NL0	L0	NLC	LC
0.3	0.3961±0.0041	0.5298±0.0028	0.386±0.039	0.525±0.031
0.2	0.4346±0.0072	0.4631±0.0051	0.433±0.07	0.455±0.056
0.15	0.454±0.013	0.4452±0.0072	0.45±0.11	0.432±0.082
0.1	0.475±0.023	0.448±0.013	0.48±0.20	0.454±0.14

The results show that the value of β obtained from the tests are all consistent with each other up to k =0.2 h Mpc^{-1}. On smaller scales, the β values obtained by fitting the data with the CAMB linear model are systematically higher than those obtained with the power spectra measured in the simulation, indicating that non linear effects can not be neglected on these scales. Focusing on test 4, the more realistic one, we see that, using linear theory only, one could expect to estimate the value of β with an uncertainty $< 15\%$ if one pushes the analysis out to $k_{max} = 0.2h$ Mpc^{-1}, $\sim 20\%$ if one uses the more conservative limit of $k_{max} = 0.15h$ Mpc^{-1}. On these scales the error budget is dominated by cosmic variance, as we have seen. Therefore, to improve the accuracy, one needs to push the analysis to larger k_{max}, i.e. one needs to model non linear effects.

Before doing this, however, we note that our measured β values are significantly different from the expected one. This may indicate that errors are predominatly systematic, not random. We therefore need to understand the nature of such discrepancy. The fact that the best fit β value doesn't vary significantly for $k_{max} < 0.2h$ Mpc^{-1} and with the type of test performed, speaks against systematic errors. Indeed such discrepancy has been alredy noticed by Bianchi et al. (2012) and Marulli et al. (2012) who performed clustering analysis using the same BASICC simulations. In particular Bianchi et al. (2012) pointed out that the β value obtained from the clustering analysis is significantly smaller than the expected values and that the discrepancy increases when selecting haloes of prograssively smaller masses. This mismatch originates from the fact that not all the structures identified as haloes are genuinely virialized structures. Some of them are transient clumps of matter. The fraction of fake haloes increases when decreasing the mass of the structure and is large when haloes are made by 10 particles only, as in our case. Then, our halo catalog contains a significant fraction of objects that, in fact, are simply collection of particles that are

basically unbiased respect to the underlying mass field. Hence the effective bias is smaller and the β value larger.

3.3.2 Non linear redshift space distortions

In this Section we test our ability to model non linearities in both the bias and the gorowth of dansity fluctuations. Our goal is to fit $P_{C,z}$ with the different models described below:

1. MODEL NLC1: $P(k) = b^2 P_{CAMB} K(\beta)(1 + Qk^2)$, where the last term corresponds to the one originally proposed by Cole et al. (2005) (eq.2.59) and where we fix $A = 0$. We use model NLC1 to match the measured power spectrum up to $k_{max} = 0.6h$ Mpc^{-1}. b is set to the fiducial value 0.86 whereas β and Q are treated as free parameters.

2. MODEL NLC2: $P(k) = b^2 P_{CAMB} K(\beta)(1 + Qk^{3/2})$, that is analogous to NLC1 but uses a different model for the middly non linear effect.

3. MODEL NLC3: $P(k) = P_{C,real} K(\beta) \frac{\pi^{1/2}}{2} \frac{erf(\sigma k)}{\sigma k}$. In this case we use $P_{C,real}$ to account for the halo bias and non linear effects. The Kaiser boost accounts for the coherent motions whereas the term $\frac{\pi^{1/2}}{2} \frac{erf(\sigma k)}{\sigma k}$ accounts for highly non linear motions (see Section 2.4), that, as we have pointed out, are undersampled by the BASICC simulation, and therefore, poorly modelled by NLC3. The rationale for this NLC3 model is to reduce the uncertainties related to the bias and middly non linear models and assess the impact of highly non linear effects. Free parameters of the model are β and σ.

The fitting procedure is similar to that of the previous Section. In this case we fit the ratios $P_{C,real}/b^2 P_{CAMB}$ and $P_{C,z}/P_{C,real}$ with the proposed model, taking into account the errors on the measured $P_{C,real}$ and $P_{C,z}$. The best fit β values obtained from models NLC1, NLC2, NLC3 are shown in fig. 3.16 and listed in Table 3.4 as a function of k_{max} with their 1-σ uncertainties. Model NLC3, which is intrinsically more accurate, provides a good fit to the estimated one and successfully recovers the correct β value up to $k_{max} = 0.6h$ Mpc^{-1}. The stability of the result is indeed remarkable. The accuracy increases with k_{max} since cosmic variance decreases with k. As a result, at $k_{max} = 0.6h$ Mpc^{-1} β is estimated with 10% accuracy. The results obtained with the more realistic NLC1 and NLC2 show that the estimated value of β have similar accuracy at $k_{max} = 0.6h$ Mpc^{-1}. However, model NLC2 systematically underestimates β, although the discrepancy is within the error bars.

We conclude that the proposed non linear models are adequate to estimate the value of β well into the non linear regime, with an accuracy of about 10% and with no significant systematic errors. This accuracy is higher

than that obtained with the linear model LC. The results also suggest model NLC1 performs better than NLC2. Table 3.5 shows the best fit values for the remaining parameters Q and σ. The best fit Q value are smaller than unity and determined with an accuracy of $15-20\%$ at $k_{max} = 0.6h$ Mpc^{-1}. The parameter σ, typically interpreted as the 1 dimensional pair velocity dispersion, is of the order of 150 Km s^{-1}. This is rather small if compared with the value typically obtained from real galaxy samples and confirms the fact that BASICC haloes underestimates non linear motions.

Table 3.4: The best fit β values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	β / NLC1	β / NLC2	β / NLC3
0.2	0.44±0.13	0.43±0.16	0.48±0.19
0.3	0.408±0.079	0.382±0.096	0.462±0.098
0.4	0.431±0.053	0.392±0.065	0.460±0.074
0.5	0.456±0.041	0.408±0.050	0.447±0.053
0.6	0.483±0.033	0.430±0.041	0.437±0.046

Table 3.5: The best fit Q and σ values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	Q / NLC1	Q / NLC2	σ $[Mpc/h]$ / NLC3
0.2	0.5±3.5	0.2±1.7	2.0±3.7
0.3	1.5±1.0	0.92±0.61	1.58±0.96
0.4	1.20±0.40	0.85±0.29	1.56±0.51
0.5	0.97±0.21	0.76±0.17	1.45±0.25
0.6	0.78±0.13	0.69±0.11	1.39±0.18

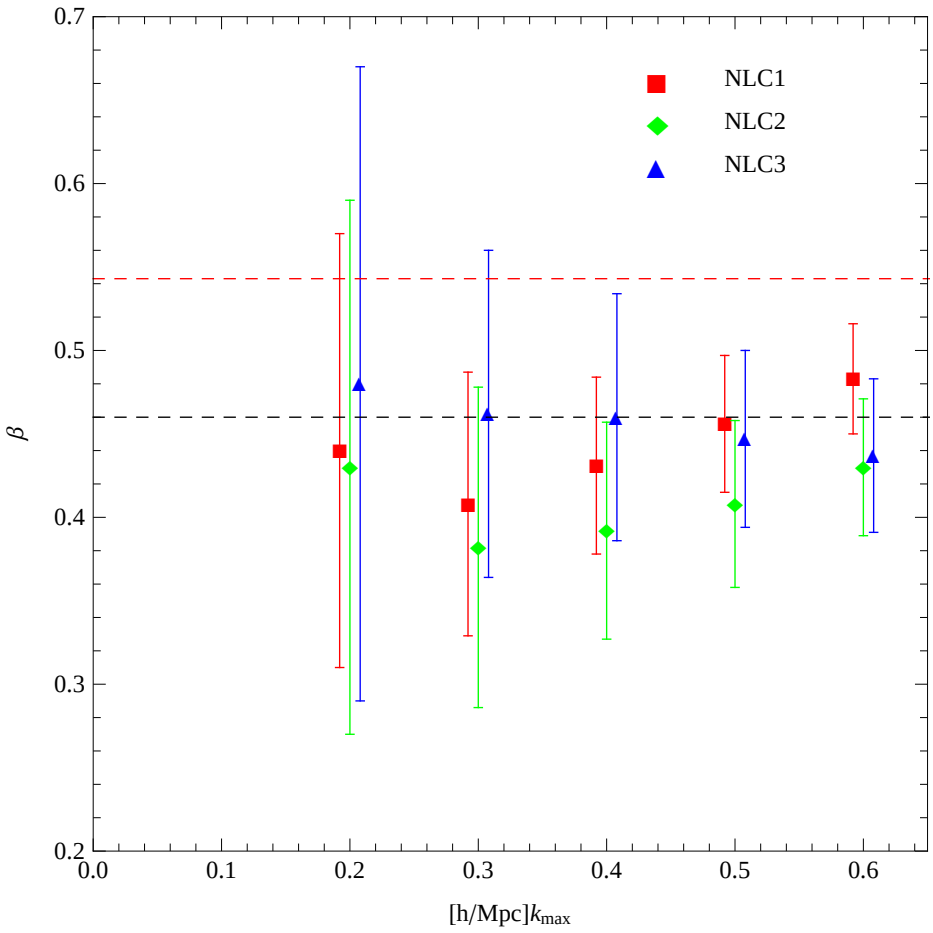


Figure 3.16: The best fit β values and their *rms* uncertainty for the various tests as a function of k_{max}. Black dashed line the β "reference" expected value. Red dashed line is β_{theory}=$\Omega_M^{0.55}/b$.

3.3.3 Including the effect of the window function

The second set of tests is carried on with the spherical subsamples that match the 2MRS window function. The goal of these tests is to check the ability to model the redshift space distortions when the effect of the window function is included. We account for the window function convolving models of power spectrum in redshift space with the spherically averaged window function, as described in 2.3.4. We have obtained β by fitting the measured power spectrum $P_{SRZ,z}$, or the ratio $P_{SRZ,z}/P_{SRZ,real}$, with the following models.

1. MODEL NLW: fitting $P_{SRZ,z}/P_{SRZ,real}$ with $K(\beta)$.

This model is analogous to NLC, described in Section 3.3.1. The goal of this test is to check if the results obtained with the NLC model are confirmed when real space and redshift space power spectra are affected by the same window function. In these case $P_{SRZi,z}$ and $P_{SRZi,real}$ are indeed characterized by the same window function. Errors on $P_{SRZ,z}$ and $P_{SRZ,real}$ were obtained from the scatter among the spheres and propagated to their ratio.

2. MODEL LW: fitting $P_{SRZ,z}$ with $[P_{CAMB}b^2K(\beta)] * WF$.

   This is analogous to LC model, but now the fitting function accounts for the window function, more precisely, the model corresponds to the convolution between the linear $[P_{CAMB}b^2K(\beta)]$ and the measured spherically averaged window function WF, computed from a random distribution of objects that match the 2MRS selection properties, shown in fig.3.11. As for the previous tests b is set to 0.86 and the CAMB parameters are set to the BASICC simulation values. The errors come from the scatter among the spheres only. This model is more realistic respect to NLW. The goal of this test is to verify if we are able to model redshift distortions when the window function is present using the convolution with the sherically averaged window function, described in Section 2.3.4 and a linear model. Deviations with respect to the expected β value would indicate departures from linear theory or non correct modelling of the window function.

3. MODEL NLW1: $P(k) = [b^2P_{CAMB}K(\beta)(1 + Qk^2)] * WF$.

   This is the analogous of NLC1 model. We use it to model the middly non linear effects and the window function. We set β and Q as the only free parameters.

4. MODEL NLW2: $P(k) = [b^2P_{CAMB}K(\beta)(1 + Qk^{3/2})] * WF$.

   This is the analogous of NLC2 model. This is a different model to account for the middly non linear effects and the window function. β and Q are free parameters.

We fit the ratio $P_{SRZ,z}/P_{SRZ,real}$ with NLW. We fit the measured power spectrum $P_{SRZ,z}$ with models LW, NLW1 and NLW2. In all these cases we estimate β fitting data up to different k_{max} values. The best fit β values obtained from models NLW, LW, NLW1 and NLW3 are shown in fig. 3.17 and listed in Table 3.6 as a function of k_{max} with their 1-σ uncertainties.

The results obtained with model NLW are very similar to the ones obtained with model NLC. We can recover the β value up to $k_{max} = 0.2$ h Mpc^{-1}. These results indicate that the ratio $P_{SRZ,z}/P_{SRZ,real}$ is similar to $P_{C,z}/P_{C,real}$, and that a k-independent model for redshift distortions is valid on large scales.

The more realistic model LW indicates that we recover β with an uncertainty of $\sim 15\%$ if we push the analysis out to $k_{max} = 0.15\ h\ \mathrm{Mpc}^{-1}$, and 20% for the more conservative choice $k_{max} = 0.15\ h\ \mathrm{Mpc}^{-1}$. Non linearities in the growth of density fluctuations and bias affect the result for $k = 0.20h\ \mathrm{Mpc}^{-1}$. These results are similar to the ones obtained with LC model.

When using non linear models NLW1 and NLW2 we recover the expected β value with an uncertainty of $\sim 10\%$ fitting data up to $k_{max} = 0.6\ h\ \mathrm{Mpc}^{-1}$.

From these results we conclude that we are able to account correctly for the combination of redshift distortions and 2MRS-like window function. The latter, as shown in Section 3.2.5, is a rather compact function and affects mostly the largest scales, where we are able to account for it. The introduction of the convolution with such a window function doesn't change the ability to model redshift space distorsions respect to the simpler case without window function, seen in the previous Section. The method of convolving with the spherically averaged window function (described in Section 2.3.4) works well when using the redshift space power spectrum monopole as true underlying power spectrum because the 2MRS window function is rather compact, almost isotropic and affects principally scales where the redshift distortion can be modelled with k-independent model.

Table 3.6: The best fit β values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	β / NLW	β / LW	β / NLW1	β / NLW2
0.1	0.48±0.26	0.494±0.19	×	×
0.15	0.45±0.15	0.464±0.10	×	×
0.2	0.43±0.10	0.469±0.073	0.46±0.19	0.45±0.23
0.3	0.386±0.054	0.533±0.042	0.42±0.10	0.40±0.13
0.4	×	×	0.443±0.072	0.407±0.088
0.5	×	×	0.464±0.055	0.420±0.067
0.6	×	×	0.489±0.046	0.441±0.056

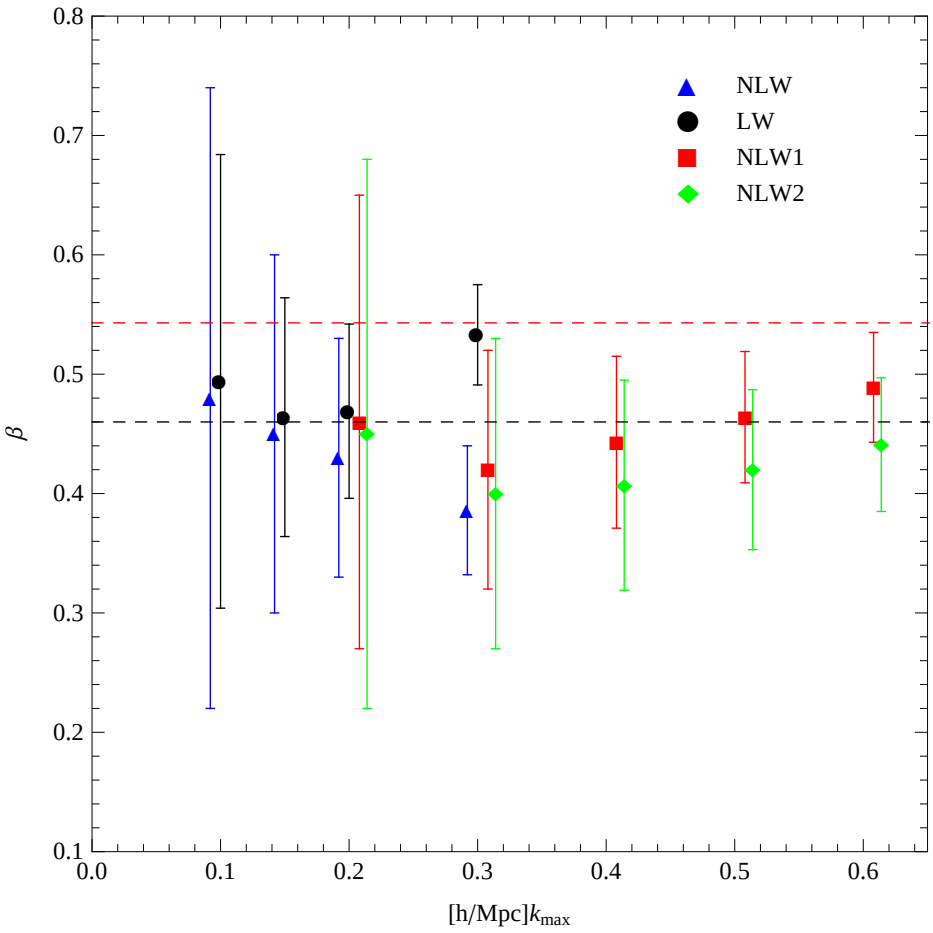


Figure 3.17: The best fit β values and their *rms* uncertainty for the various tests as a function of k_{max}. Black dashed line the β "reference" expected value. Red dashed line is β_{theory}=$\Omega_M^{0.55}/b$.

Chapter 4

The 2MRS analysis

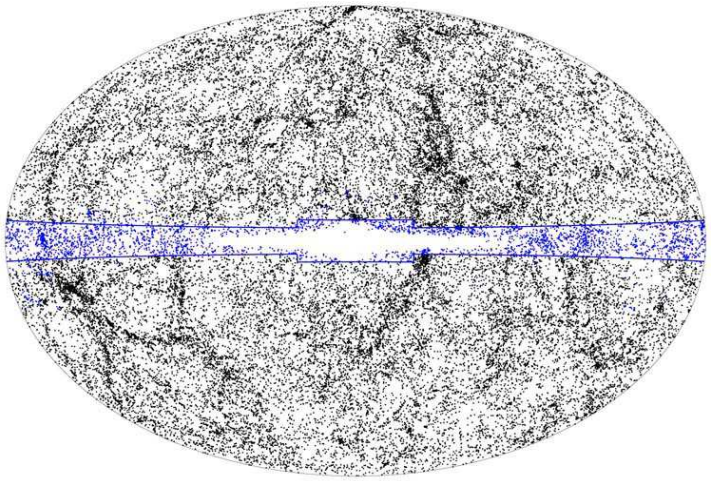
In this Chapter we describe the 2MRS catalog, to which we have applied our own FKP code. We describe also the mock catalogs used in order to estimate the uncertainties in the power spectrum estimate. We use the mock catalogs also in order to select the optimal type of model of bias and non linearities to be adopted in the true analysis and to select the wavenumber range in which these models are valid, allowing to recover unbiased values of the cosmological parameters. Finally we compute the $P(k)$ of the real 2MRS galaxies, divided in subsamples. One of the goals of this Chapter is to see how well we constrain the bias using FKP estimator on a local all sky survey. An accurate bias estimate is important per se and a pre-requisite to obtain unbiased estimates of the cosmological parameters. At the same time, it provides information about the galaxy evolution models. We decide to carry on the analysis as follows: first, we study the clustering properties of different galaxy groups, obtained splitting the main 2MRS sample in subsamples according to a specific and physically motivated criterion. In this part of the work we are interested to the relative bias between the different groups of objects and to the precision with which we are able to recover it, and compare results with that obtained in previous works. Finally we use the full shape of the power spectrum of the main sample in order to obtain an estimate of the linear bias parameter b and the matter density parameter, Ω_M.

4.1 The 2MRS catalog

The 2MASS Redshift Survey (2MRS, Hucra et al. 2012) is a selection of galaxies from the 2MASS Extended Source Catalog (XSC; Jarrett et al. 2000) and 2MASS Large galaxy Atlas (LGA) , version 2.5 (Jarrett et al. 2003). The 2MASS Extended Source Catalog contains approximatively 1 million extended objects from the 2MASS, 2 Micron All Sky Survey (Skrutskie et al. 2006), with K_s <13.5 mag, detected as extended sources

through dedicated algorithms (XSC; Jarrett et al. 2000). The 2MASS is an all sky imaging survey in the three near infrared photometric bands J ($1.25\mu m$), H ($1.65\mu m$), and K_s($2.16\mu m$). It used the twin 1.3-m telescopes located at Mount Hopkins, Arizona and Cerro Tololo, Chile. The effects of interstellar extinction, are strongly reduced with respect to the optical or far infrared surveys and this allows a wider coverage of the sky. The complete coverage of the sky for the galaxies with $K_s \leq$11.75 of the 2MASS XSC is limited only by a confusion region around the galactic center, as shown in fig.4.1.

Figure 4.1: 2MASS galaxies in Galactic coordinates, Aitoff projection, with $K_s \leq 11.75$ mag. Blue dots correspond to the galaxies in the area not covered by the 2MRS. Figure from Hucra et al., 2012.



The 2MASS Redshift Survey is the result of a program aimed to obtain the spectroscopic redshifts of a magnitude limited sample selected from the 2MASS XSC started in September 1997. In this thesis we use the 2.4 version of the catalog, publicly released on 16 December 2011, and available at http://tdc-www.cfa.harvard.edu/2mrs/. The reference paper is Hucra et al. (2012). The selection criteria are the following:

1. $K_s \leq 11.75$ mag and detected at $H-$band
2. $E(B-V) \leq$1.0 mag
3. $|b| \geq$5 deg for $30 \leq l \geq 330$ deg; $|b| \geq$8 deg otherwise.

where K_s is the apparent isophotal magnitude, measured at K_s =20 mag/arsec^2 isophote. The isophote curves denote the curves of equal apparent surface brightness. In this thesis, we use only the isophotal K_s band magnitudes that are the ones on which the selection has been applied in the

2MRS and not the total ones. The excess $E(B-V)$ measures the effects of interstellar absorption on the B-V colour where B and V are two magnitude bands of the Johnson's photometric system. b and l are respectively the Galactic latitude and longitude. The third selection criterion excludes, as anticipated in the previous Chapter, objects within the region near the galactic plane, referred to as Zone of Avoidance. The resulting catalog contains 44599 objects, the redshift is provided for 43533 of them. For each galaxy selected according to the criteria enumerated above, a multiple query to the following databases have been done: SDSS (data release 8), 6dF Galaxy Survey, NASA Extragalactic Dabase and Hucra's personal compilation of redshifts (ZCAT). For the galaxies without a match in any database at the time of the query, the redshift have been measured, at the Fred L. Whipple Observatory 1.5-m telescope for the northern hemisphere part of the survey, at the Cerro Tololo Interamerican Observatory 1.5-m telescope for the southern hemisphere. The observations have been carried on between 1998 July and 2011 January. The redshifts were obtained through longslit spectroscopy. The median redshift of the survey is 0.03. Fig. 4.2 shows the Table with the main characteristics of the telescopes and instruments used in the survey.

Figure 4.2: Telescopes and instruments used for the redshift measures for the 2MRS. Table from Hucra et al., 2012.

Observatory/Telescope		Camera	Grating (l/mm)	Coverage (Å)	Res. (Å)	N gal with K_s < 11.75	> 11.75
Fred L. Whipple	1.5-m	FAST	300	3500-7400	5	7,590	2,596
Cerro Tololo	1.5-m	RCSpec	300	3700-7200	7	3,245	238
McDonald	2.1-m	es2	600	3700-6400	4	114	50
Cerro Tololo	4-m	RCSpec	527	3700-7400	3	48	
Hobby-Eberly	9.2-m	LRS	300	4300-10800	9	3	

For each galaxy the catalog contains various informations, among them:

1. Galactic coordinates l,b;
2. K_s, J , H isophotal magnitudes, extinction corrected;
3. K_s, J , H total magnitudes, extinction corrected;
4. uncertainties in the magnitude estimates;
5. galaxy type;
6. observed redshift in Km s^{-1} in the heliocentric reference frame.

Galaxy types are provided for a 20860 objects with K_s <=11.254 mag and $|b| > 10$, plus around 5000 fainter objects.

4.2 2MRS mock catalogs

In addition to the real 2MRS catalog we use a set of 135 2MRS mock catalogs in order to estimate the statistical errors on the measured power spectrum. As detailed in Chapter 2, the uncertainties are quite tricky to be formalized analitically due to the coupling between the modes induced by the window function and non linearities. The mock catalogs have been also used in order to test models of non linear bias and non linear dynamics to be adopted in the true analysis and to set the wavenumber range in which they can be applied. These mock catalogs match approximatively the general properties of the 2MRS. Given the volume of the parent simulation's box only 15 of them are independent. They have been extracted from the Millennium Simulation volume. 2MASS catalogs were generated through the semi analytic model of galaxy formation of De Lucia & Blaizot (2007) from which we finally extracted our 2MRS mocks. Each mock galaxy in the catalog is identified by:

1. Galactic coordinates l,b;
2. the Cartesian components of the peculiar velocity along the line of sight with respect to the Hubble flow in Kms^{-1} , v_{pec};
3. peculiar velocity along the line of sight to which a Gaussian random error has been added to simulate errors on the observed redshift, $v_{pec,obs}$;
4. observed redshift in Kms^{-1} , cz, (including $v_{pec,obs}$);
5. comoving Cartesian coordinates of comoving distance d in Mpc h^{-1};
6. apparent magnitude in K band , extinction corrected;
7. total on stellar mass of the Galactic bulge;
8. absolute magnitude in K, J and H band;
9. absolute magnitude in u, g, r, i, z optical bands. The latter are not available in the real 2MRS catalog. These will be used to colour-select galaxy subsamples.

4.3 Analysis of the mocks

In this Chapter we compare the measured power spectrum of 2MRS galaxies with theoretical predictions. We estimate the free parameters of the model by minimising the χ^2 differences between model and data using the errors computed from the mocks. In so doing we consider different models that can be divided in two categories: linear models and nonlinear models. The

former assume that density fluctuations are small and, consequently, we restrict model vs. data comparison on large scales. For the latter we push the comparison to smaller scales.

In both cases we make an important assumption: the covariance matrix is purely diagonal, i.e. the power measured in different k-bins is uncorrelated. This hypothesis holds true only when the field is Gaussian, density fluctuations are small and the spectrum is measured in a cubic region with periodic boundary conditions. Breaking any of these hypotheses will produce non-zero off-diagonal elements in the covariance matrix. Since none of these hypotheses are rigorously valid in our case, we make the assumptions that off-diagonal terms are small compared to the diagonal ones and can be neglected in the error budget.

This sounds reasonable when we consider the linear regime. In this case we focus on large scales, where linear theory should hold, and we consider a sample with a simple window function that is close to the spherical top hat, as we have verified. For the nonlinear model, the validity of this hypothesis should be tested. However, such assessment would require a large number of independent mock 2MRS catalogs, whereas only 15 of them among the 135 we have extracted from the Millennium Simulation can be considered statistically independent. This is beyond the scope of this thesis and our current computational capabilities. Therefore, we shall regard our errors, estimated from the *rms* scatter among the mocks, as a lower limit to the true uncertainties. However, it is reassuring that, as we shall see, the errors on the cosmological parameters estimated from the linear model are close to those estimated form the nonlinear ones. A clue that off-diagonal errors in the nonlinear case cannot be much larger than those in the linear model.

4.3.1 Accounting for redshift space distortions : methodology

As in Section 3.3 we need to model $z-$distortions. We can't rely on those results since 2MRS mock catalogs are more accurate in this respect since they allow for multiple galaxy within the same halo, each one with its own peculiar velocity. A better model for the mock galaxies' velocity field requires a better model for redshift distortions. For this purpose we use 95 of the 135 mock realizations described in Section 4.2 and consider objects within a distance of 140 Mpc h^{-1} from the observer to mimic the cut applied in the real catalog, justified by the fact that shot noise becomes high beyond that radius. We also exclude objects within the Zone of Avoidance, $|b| < 5^o$, in order to match the real catalog geometry. We apply a luminosity cut at the redshift $s_{cut} = 3000 Kms^{-1}$ to minimize incompleteness. This luminosity cut is set to extract a semi-volume limited sample, as follows. We eliminate all the objects that, if placed at s_{cut} would have a K-band magnitude above the constant limit K_s =11.75 mag or, which is a more operable definition,

that have absolute K_s magnitude:

$$M_K > M_{K,cut} = 11.75 - 5log(d_{Lcut}) - 25 \quad (4.1)$$

where d_{Lcut} is the luminosity distance at s_{cut}, defined in 1.22. The luminosity distance has been computed from the comoving distance as $d_{cut}(1 + z_{cut})$, with d_{cut} =30 Mpc h^{-1}. We keep all the galaxies beyond s_{cut}. The result is a volume limited sample up to $s = 3000Kms^{-1}$ and flux limited beyond this distance. When we compare the redshift distribution of galaxies in these mocks, dN/ds, with that of the real catalog, we notice that mock galaxies are more concentrated around the central observer. To eliminate this mismatch and possible incompleteness in the real catalog that may miss nearby, extended objects, we decided to exclude all objects within 30 Mpc h^{-1}.

4.3.2 Comparison real space - redshift space

The mean power spectrum of the 95 2MRS mocks is shown in fig.4.3 (blue solid line) togheter with its 1-σ scatter (blue filled region). In this case distances where computed from observed redshifts that include peculiar velocities, i.e. the spectrum is computed in redshift space and dubb it $< P_{mock_i,z} >$.

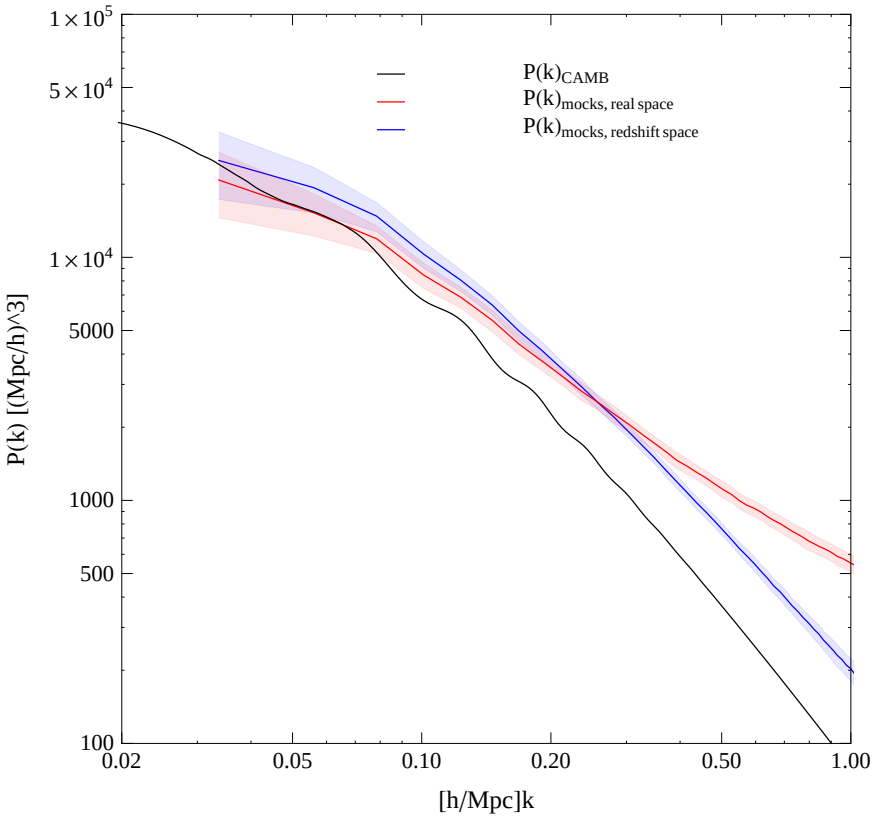
We also compute the real space power spectrum $< P_{mock_i,real} >$ using the comobile distances provided in the mocks. $< P_{mock_i,real} >$ and its scatter are also shown in fig.4.3 (red line and red filled region) togheter with linear theory predictions obtained from CAMB (black line). From the comprehensive comparison between $< P_{mock_i,z} >$ and $< P_{mock_i,real} >$ and CAMB shown in fig.4.3 we can appurate the small scale boost induced by non linear effects in $< P_{mock_i,real} >$, the small scale damping induced by non linear motions in $< P_{mock_i,z} >$ and the large scale Kaiser boost in $< P_{mock_i,z} >$ induced by coherent motions.

The ratio between real space and redshift space power spectra is shown in fig. 4.4. Linear theory would predict a constant $K(\beta) = (1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2)$ for this ratio. The plot shows that the ratio is a decreasing function of k, which indicates that non linear effects cannot be neglected for $k > 0.1$ h Mpc^{-1}.

4.3.3 Modelling the 2MRS power spectrum

A second ingredient to extract cosmological parameters from the measured 2MRS power spectrum is a model for the window function of the catalog. The window function is computed from a random catalog matching the selection effects of the 2MRS mocks but with no clustering on 1024^3 grid with size L = 1400 Mpc h^{-1} to match a resolution of $\Delta k = 2\pi/1400h$ Mpc^{-1}. We now put everything togheter (non linearities, bias , window

Figure 4.3: Mean redshift space power spectrum (blue line) and mean real space power spectrum (red line) of the 95 2MRS mock catalogs with their 1-σ scatter (blue and red shaded areas, respectively). Linear matter power spectrum for mocks cosmology (black solid line) obtained from CAMB.

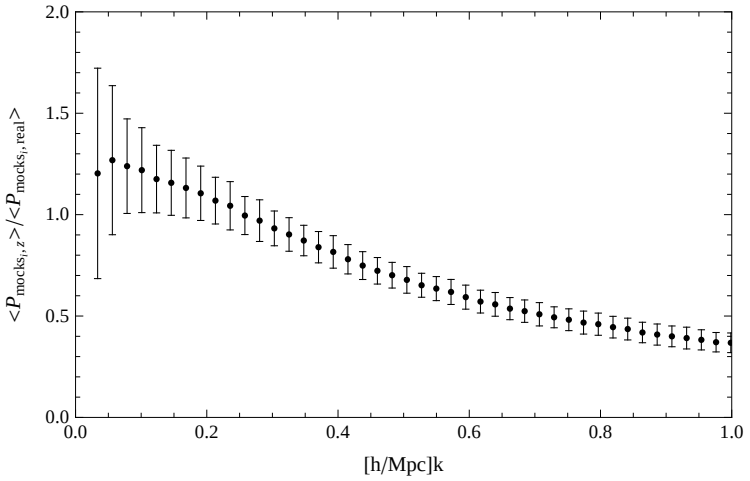


function) to fit the power spectrum of the mock galaxies. The goal is to select the best model for the $P(k)$ and the corresponding $k-$range, that we shall use in the real analysis. As in Section 3.3 we have obtained an estimate of the β parameter by fitting the measured power spectrum $< P_{mock_i,z} >$, or the ratio $< P_{mock_i,z} > / < P_{mock_i,real} >$, with the following models.

1. MODEL NLW: fitting $< P_{mock_i,z} > / < P_{mock_i,real} >$ with $K(\beta)$.

   To put the impression given by fig. 4.4 on a more quantitative ground in this case we fit a constant function, the so called Kaiser boost, $K(\beta)$, to the measured ratio $< P_{mock_i,z} > / < P_{mock_i,real} >$. In this model the errors are computed for $< P_{mock_i,z} >$ and $< P_{mock_i,real} >$ from the *rms* scatter among the 95 mocks and then are propagated to the ratio.

Figure 4.4: Ratio between mean redshift space power spectrum and mean real space power spectrum: $< P_{mock_i,z} > / < P_{mock_i,real} >$. The error bars are computed for $< P_{mock_i,z} >$ and $< P_{mock_i,real} >$ from the rms scatter among the 95 mocks and then propagated to the ratio.



2. MODEL NLW$_G$: fitting $< P_{mock_i,z} > / < P_{mock_i,real} >$ with $K(\beta)G(k,\sigma)$ where $G(k,\sigma) = \frac{\pi^{1/2}}{2}\frac{erf(\sigma k)}{\sigma k}$.

   This model corresponds to the previous one with the additional introduction of a damping term. β and σ are free parameters. The error bars are computed as before.

3. MODEL LW: fitting $< P_{mock_i,z} >$ with $P(k) = [P_{CAMB}b^2K(\beta)] * WF$.

   P_{CAMB} is the linear matter power spectrum from CAMB with the same cosmological parameters of the simulation and WF is the measured spherically averaged window function of the mocks. b is obtained from the ratio between the real space mean power spectrum of the mocks and linear matter power spectrum from CAMB (with input parameters fixed to the simulation ones). The ratio is fitted up to $k_{max} = 0.1h$/Mpc to guarantee linearity obtaining b =1.077±0.056. Here we fix b to the best fit value 1.077. β is free parameter. In this model, as in the following ones, the error bars come from the *rms* scatter among the mocks only. This model is more realistic with respect with the previous ones, and accounts for linear terms only.

4. MODEL NLW1: $P(k) = [b^2P_{CAMB}K(\beta)(1 + Qk^2)] * WF$.

   This model is used to account for the middly non linear effects and the window function. β and Q are fre parameters.

5. MODEL NLW2: $P(k) = [b^2P_{CAMB}K(\beta)(1 + Qk^{3/2})] * WF$.

Analogous to the previous one but with a different dependence on k. β and Q are fre parameters.

6. MODEL NLW3: $P(k) = [P_{Hf} b^2 K(\beta)] * WF$.

   In this model we use the non linear matter power spectrum from the *Halofit* code, with input cosmological parameters fixed. β is the only free parameter.

7. MODEL NLW3$_G$:$P(k) = [P_{Hf} b^2 K(\beta) G(k, \sigma)] * WF$.

   This model is analogous to the previous one but contains the Gaussian damping term which accounts for dynamical non linearities.

The best fit β values were obtained by minimizing the χ^2 difference between data and model with error bars from the mocks.

Fig. 4.5 shows the best fit β as a function of the maximum wavenumber k_{max} considered, when fitting data with model NLW and NLW$_G$. The plot for model NLW shows non linear effects beyond $k = 0.15$ h Mpc^{-1}. Introducing a damping term, with model NLW$_G$, has the effect of increasing the best fit β-value. The results have been compared with the expected β value, dubbed β_{theory}, given by $\Omega_M^{0.55}/b$, where Ω_M is the one used in the simulation. b is estimated from the ratio between real space power spectrum and P_{CAMB}, as described above, obtaning b =1.077±0.056 and β_{theory} =0.433±0.023, respectively. This is higher but still compatible within 1-σ with the best fit value obtained both with linear theory ($\beta = 0.29 \pm 0.14$ at k=0.15 h Mpc^{-1}) and non linear damping term ($\beta = 0.35 \pm 0.13$ at k=0.3 h Mpc^{-1}) and possibly indicates that additional non linear effects, non modelled by a simple damping term, are present on these scales (Marulli et al. 2012).

The results are also summarized in Table 4.1.

Fig. 4.6 shows the best fit β value as a function of k_{max} when the $z-$space power spectrum is simply modelled using linear theory (model LW), i.e. using P_{CAMB} as the spectrum for the matter, $K(\beta)$ to model $z-$distortions and $b = 1.077$ as linear bias parameter. The linear model provides an acceptable fit (i.e. in agreement with expectations β_{theory} =0.433±0.023, represented by the thick black line) if $k_{max} < 0.2h$ Mpc^{-1} and overestimates the correct value on smaller scales. For $k_{max} = 0.15h$ Mpc^{-1} the β parameter is estimated with an accuracy of 19%. When an additional degree of freedom is added to model middly non linearities with the Q-factor $F(k, Q) = (1 + Qk^2)$ (model NLW1) the accuracy on the β measurement decrease, as expected, but systematic difference between the true and the measured $\beta-$values are within random errors out to $k_{max} = 0.4h$ Mpc^{-1}, well into non linear scales. These results are shown in fig.4.7. The accuracy at $k_{max} = 0.4h$ Mpc^{-1} is $\sim 15\%$ as a result of the largest number of $k-$modes sampled. All $\beta-$values plotted in the figure for NLW1 model are computed

in correspondence of the best fit Q-values, shown in Table 4.3. In the figure we show the results obtained with model NLW2. The results can not be distinguished from the previous ones.

An alternative way to model non linear effects is to use non linear $P(k)$ model obtained from *Halofit*. The results are shown in fig. 4.8, model NLW3. They are less satisfactory than the ones obtained with the model NLW1. β is systematically underestimated meaning that *Halofit* is overestimating power on large scales, which is not surprising since we are not modelling the damping effect of small scales virial motions on the $P(k)$ measured in $z-$space. Adding a Gaussian damping term (model NLW3$_G$), characterized by a velocity dispersion parameter σ, improves the quality of the fit, as expected. Fig. 4.9 demonstrates this point. The correct β value is recovered all the way down to $k_{max} = 0.4h$ Mpc^{-1}, with an accuracy of 16%, which is quite remarkable.

The results of this exercise provide important guidelines for the analysis of the real data. First of all, the analysis of the mocks allows to estimate uncertainties on the parameter of the fit, here we focused on the parameter β, and indicates the range of validity of each of the models we have used for the $z-$space power spectrum. We shall use this information, which we summarize in Table 4.2, in the analysis presented in the following Sections.

Table 4.1: The best fit β values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	β / NLW	β / NLW$_G$
0.1	0.33±0.23	×
0.15	0.29±0.14	0.36±0.28
0.2	0.247±0.099	0.35±0.20
0.3	0.103±0.058	0.35±0.13
0.4	×	0.33±0.10
0.5	×	0.325±0.089
0.6	×	0.310±0.080

Table 4.2: The best fit β values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	β / LW	β / NLW1	β / NLW2	β / NLW3	β / NLW3$_G$
0.1	0.32±0.17	×	×	0.42±0.17	0.43±0.44
0.15	0.446±0.083	0.22±0.24	0.15±0.32	0.356±0.082	0.41±0.13
0.2	0.484±0.064	0.33±0.16	0.28±0.19	0.302±0.062	0.36±0.13
0.3	0.578±0.042	0.381±0.097	0.33±0.12	0.152±0.040	0.44±0.10
0.4	×	0.449±0.067	0.390±0.081	×	0.472±0.076
0.5	×	0.495±0.054	0.433±0.066	×	0.485±0.068
0.6	×	×	0.472±0.054	×	×

Table 4.3: The best fit non linear parameters values and their *rms* uncertainty for the various tests (columns) as a function of k_{max} (rows).

$k_{max}[h/Mpc]$	Q / NLW1	Q / NLW2	σ[Mpc/h] / NLW$_G$	σ[Mpc/h] / NLW3$_G$
0.15	10±13	4.6±7.5	3.7±5.2	0.0±3.5
0.2	4.6±4.9	2.5±2.7	3.4±2.6	3.1±1.6
0.3	3.1±1.5	1.89±0.95	3.34±0.79	3.61±0.72
0.4	1.90±0.55	1.38±0.40	3.23±0.46	3.83±0.32
0.5	1.39±0.31	1.095±0.25	3.18±0.32	3.91±0.24
0.6	×	0.90±0.16	3.10±0.25	×

Figure 4.5: The best fit β values and their *rms* uncertainty for the models NLW and NLW$_G$ as a function of the maximum wavenumber k_{max}. The results are obtained by fitting the ratio between the mean redshift space and the mean real space power spectrum of 95 mocks with the Kaiser boost factor $K(\beta)$ (model NLW), and Kaiser boost multiplied by Gaussian damping $K(\beta)G(k,\sigma)$ (model NLW$_G$). The k_{max} values are slightly artificially shifted respect to the orginal value. The β "reference" expected value β_{theory}=$\Omega_M^{0.55}/b = 0.433$ (with the value of Ω_M given by the one used in the mocks simulation and b measured from the real space power spectrum, see text for details), is represented by the black thick dashed line with $1-\sigma$ error (shaded area).

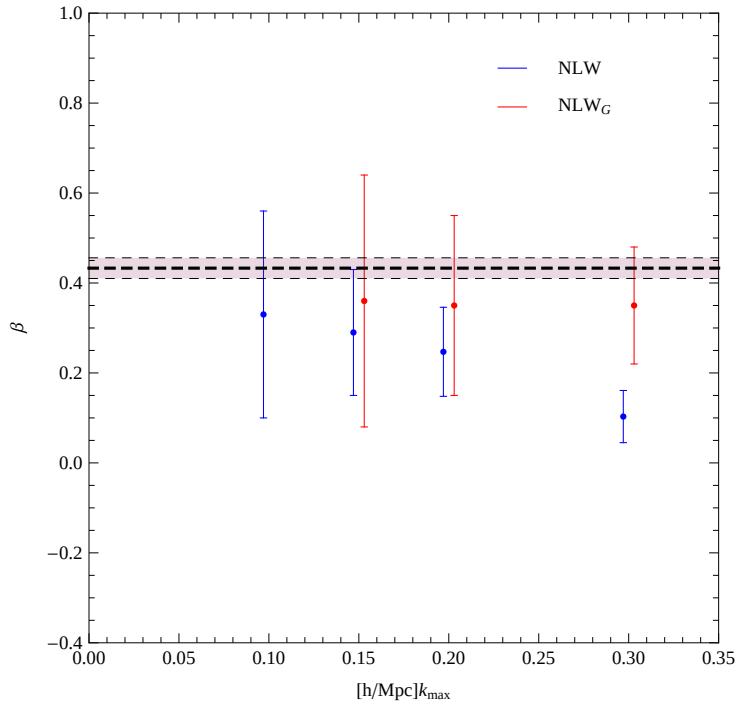


Figure 4.6: The best fit β values and their *rms* uncertainty for the model LW as a function of the maximum wavenumber k_{max}. The results are obtained by fitting the redshift space mean power spectrum of 95 mocks with $P(k) = [P_{CAMB}b^2K(\beta)]*WF$ (where $K(\beta)$ is the Kaiser boost, b is the bias measured from the real space power spectrum, P_{CAMB} is the linear matter power spectrum from CAMB, WF denotes the window function of the mocks). The β "reference" expected value β_{theory}=$\Omega_M^{0.55}/b = 0.433$ (with the value of Ω_M given by the one used in the mocks simulation, see text for details), is represented by the black thick dashed line with $1-\sigma$ error (shaded area).

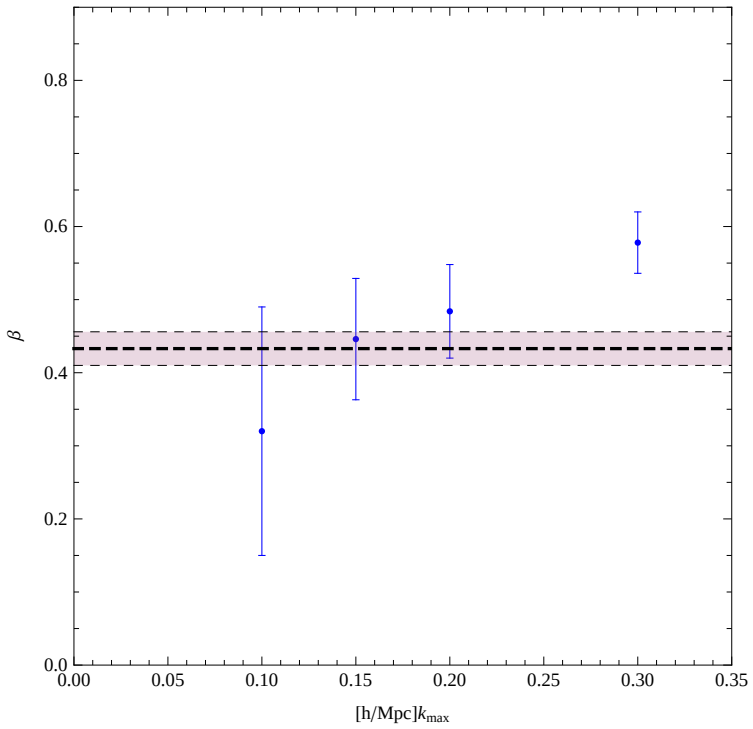


Figure 4.7: The best fit β values and their *rms* uncertainty for the models NLW1 and NLW2 as a function of the maximum wavenumber k_{max}. The results are obtained by fitting the redshift space mean power spectrum of 95 mocks with $P(k) = [b^2 P_{CAMB} K(\beta) F(k,Q)] * WF$ (where $K(\beta)$ is the Kaiser boost, b is the bias measured from the real space power spectrum, P_{CAMB} is the linear matter power spectrum from CAMB, $F(k,Q) = (1 + Qk^2)$ for model NLW1 and $F(k,Q) = (1 + Qk^{3/2})$ for model NLW2, WF denotes the window function of the mocks). The k_{max} values are slightly artificially shifted respect to the orginal value. The β "reference" expected value β_{theory}=$\Omega_M^{0.55}/b = 0.433$ (with the value of Ω_M given by the one used in the mocks simulation), is represented by the black thick dashed line with $1-\sigma$ error (shaded area).

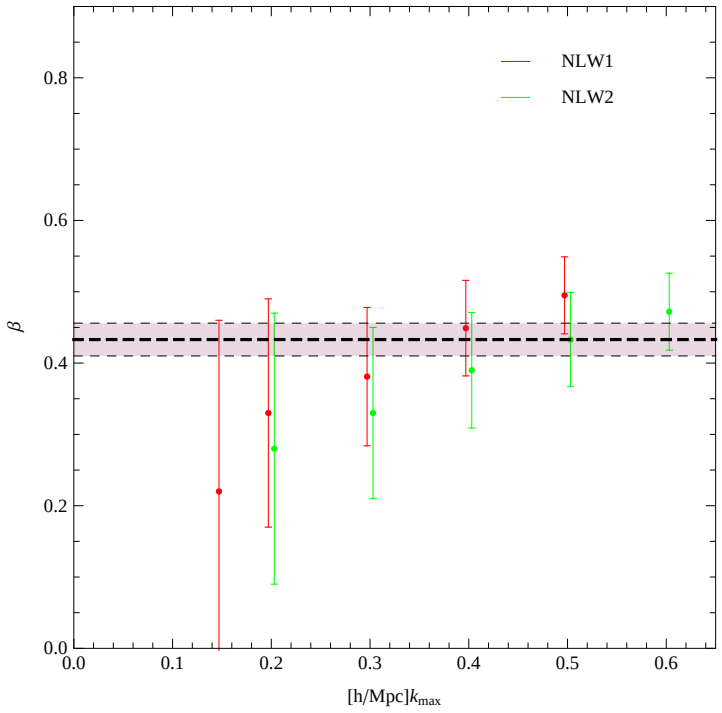


Figure 4.8: The best fit β values and their *rms* uncertainty for the model NLW3 as a function of the maximum wavenumber k_{max}. The results are obtained by fitting the redshift space mean power spectrum of 95 mocks with $P(k) = [P_{Hf} b^2 K(\beta)] * WF$ (where $K(\beta)$ is the Kaiser boost, b is the bias measured from the real space power spectrum, P_{Hf} is the non linear matter power spectrum from the Halofit code, WF denotes the window function of the mocks). The β "reference" expected value β_{theory}=$\Omega_M^{0.55}/b = 0.433$ (with the value of Ω_M given by the one used in the mocks simulation), is represented by the black thick dashed line with $1-\sigma$ error (shaded area).

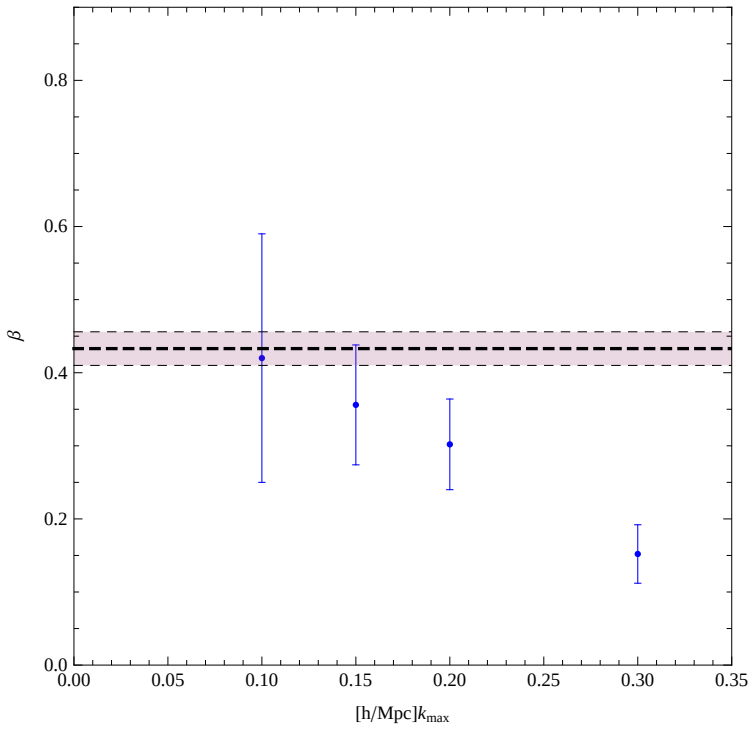
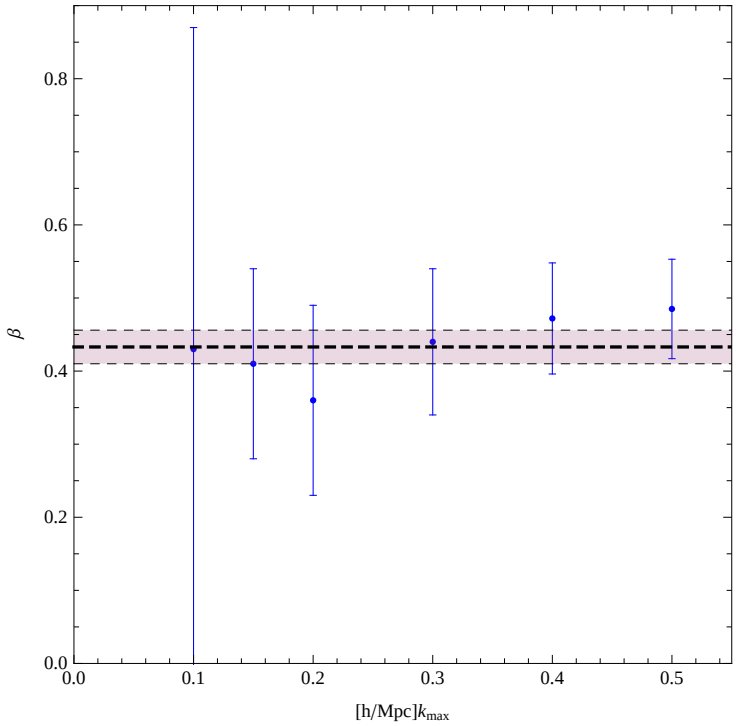


Figure 4.9: The best fit β values and their *rms* uncertainty for the model NLW3$_G$ as a function of the maximum wavenumber k_{max}. The results are obtained by fitting the redshift space mean power spectrum of 95 mocks with $P(k) = [P_{Hf} b^2 K(\beta) G(k,\sigma)] * WF$ (where $K(\beta)$ is the Kaiser boost, b is the bias measured from the real space power spectrum, P_{Hf} is the non linear matter power spectrum from the Halofit code, $G(k,\sigma) = \frac{\pi^{1/2}}{2}\frac{erf(\sigma k)}{\sigma k}$ is the Gaussian damping term accounting for non linear motions, WF denotes the window function of the mocks). The β "reference" expected value β_{theory}=$\Omega_M^{0.55}/b = 0.433$, (with the value of Ω_M given by the one used in the mocks simulation), is represented by the black thick dashed line with $1-\sigma$ error (shaded area).



4.4 The 2MRS power spectrum

4.4.1 Power spectrum measure

In this Section we compute the power spectrum of the 2MRS galaxy sample and show the result.

The analysis is performed in redshift space therefore, before computing the power spectrum , we convert the redshift in the catalog, expressed in the heliocentric frame, into the Local Group frame using the transformation of Yahil et al. 1977:

$$v_{LG} = v_{HE} + (-79 \cos l \cos b + 296 \sin l \cos b - 36 \sin b) \tag{4.2}$$

where v_{HE} denotes the heliocentric velocity of the galaxy in Km s^{-1} and v_{LG} is the new velocity in the Local Group frame in Km s^{-1}. This conversion does not have great impact on the power spectrum measurement but could affect redshift and luminosity cuts used in the sample selection. In addition, we collapse Fingers-of-God in the 7 main nearby clusters with known redshifts. In practice we assign the redshift of the cluster to galaxies members. Cluster membership is assigned to all galaxies that are within a specified distance (the cluster virial radius) from the center of the cluster. This should help to reduce errors in the estimate of the absolute magnitudes and to reduce local dynamical non linearities. The total number of objects assigned to the 7 clusters is 1966.

Redshifts are then converted into distances. This require specifying cosmological parameters. However, since we are considering a very local sample extending out to d_{MAX} =140 Mpc $h^{-1} = 0.046$, we can assume a locally Euclidean Universe and simplify eq. 1.20 with $d = cz/H_0$, where d is the comoving distance. Note that in our analysis we do not need to specify H_0 as long as we use velocity units (Km s^{-1}) to measure distances.

The maximum distance of 140 Mpc h^{-1} is set to guarantee a sufficiently high number density of objects to avoid a too large shot noise error at the edge of the sample. We also adopt: conservative cut excluding objects with $|b| < 8$ deg , which is significantly larger than the Zone of Avoidance of the 2MRS and is set to guarantee high completeness in the sample. We also exclude the objects that are nearby. We exclude all Local Group members within 3 Mpc h^{-1} and apply the luminosity cut at the redshift $s_{cut} = 3000 Kms^{-1}$, described in Section 4.3.1, to minimize incompleteness as already done for the mocks.

The final number of objects in the catalog is 34904. We refer to this galaxy sample as the *main sample*.

The redshift distribution of galaxies in the sample, dN/ds, is estimated using the so called F/T estimator (see Section 2.3.1). Since we need this dN/ds to build the random catalog used by the FKP estimator, i.e. a sample of objects with no clustering, we need to estimate all features in the measured

dN/ds that can be attributed to local structures. In other words, we use the observed dN/ds to determine a redshift distribution representative of the cosmic mean. This is the aim of the F/T estimator. We first compute the selection function $S(s)$ integrating equation 2.39 in bins $\Delta s = 150 Kms^{-1}$ and then we obtain the dN/ds as:

$$\frac{dN}{ds} = \frac{S(s)dV}{\Delta s} \frac{N_{norm}}{\int_0^{s_{norm}} S(s)dV} \tag{4.3}$$

where dV is the survey volume between the redshift s and $s + \Delta s$, N_{norm} is the total number of galaxies up to s_{norm} of 12000 Km s^{-1}. The selection function is proportional to the mean number density, so the dN/ds is proportional to $S(s)dV$. In order to compute the equation 2.39, as well as in order to apply the semi-volume limited cut, one needs an estimate of the absolute magnitude of the galaxies. To do this we need to account for the $k-$correction and for possible additional evolutionary correction in the luminosy of the galaxies. Following Branchini et al. (2012) we use $k(z)$ of 2MRS galaxies determined by Kochanek et al. (2001):

$$k(z) = -6log(1 + z), \tag{4.4}$$

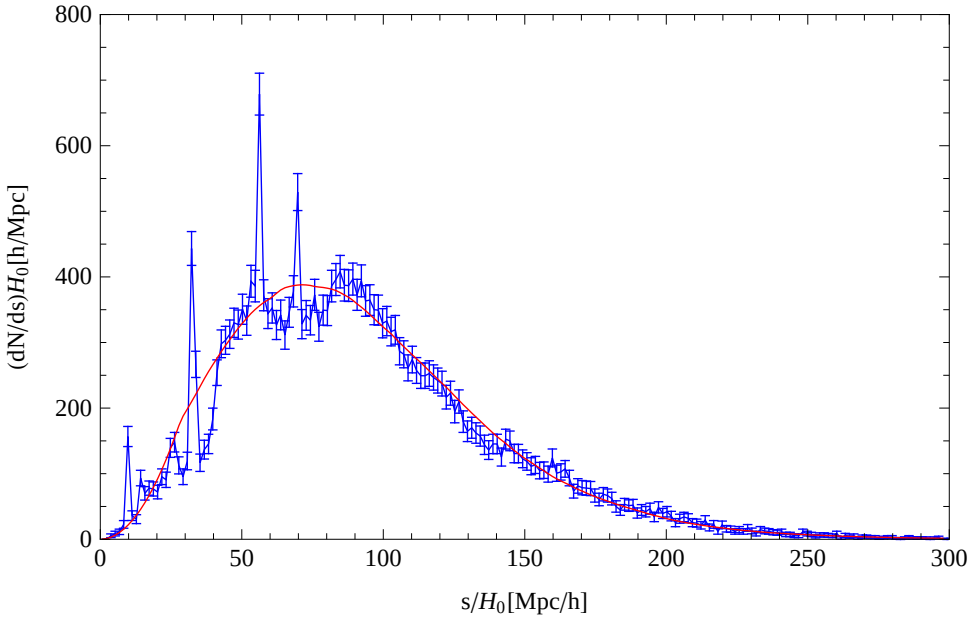
and, for an evolutionary correction, $e(z) = 3.04z$. This value was obtained by minimizing the difference between the dN/ds obtained from 4.3, including the k-correction and the evolutionary term, and the dN/ds of the catalog.

The total correction to the estimate of the absolute magnitudes of the 2MRS galaxies, adopted here is then:

$$M_{K_s} = m_{K_s} - 25 - 5log(d_L) - k(z) - e(z) \tag{4.5}$$

where m_{K_s} is the apparent K_s magnitude of the galaxy and d_L is its luminosity distance in Mpc h^{-1}. In this thesis we assume that the $k-$correction is the same for all galaxies. This is an approximation since galaxies of different types and colours have in general different $k-$corrections. However $k-$corrections are quite independent on galaxy type for infrared surveys at low redshifts (Cowie et al. 1994, Kochanek et al. 2001) so that this approximation can be considered valid for the 2MRS catalog. The redshift distribution of all 2MRS galaxies is shown in fig. 4.10 (blue histogram). The binning is $\Delta r = \Delta s/H_0 = 1.5 Mpc/h$ and the error bars are given by the Poisson noise of objects count in each bin. The red continous line is the mean dN/ds computed with F/T method. This second dN/ds is used both to estimate the mean number of objects at each redshift (we assume that no angular selection effect exist, above $|b| = 8$ deg) and used as a probability distribution function to generate the random catalog of the FKP estimator (see Section 2.3.2). The random catalog contains more than 1.5 million of objects, i.e more than 30 times more than the real one, to minimize shot noise.

Figure 4.10: Observed $(dN/ds)H_0$ of the 2MRS galaxy sample in bins Δs=150 Km s^{-1} denoted by the blue point with Poissonian $1-\sigma$ uncertainties. Mean $(dN/ds)H_0$ denoted by the solid red line computed with F/T method (see text for details).

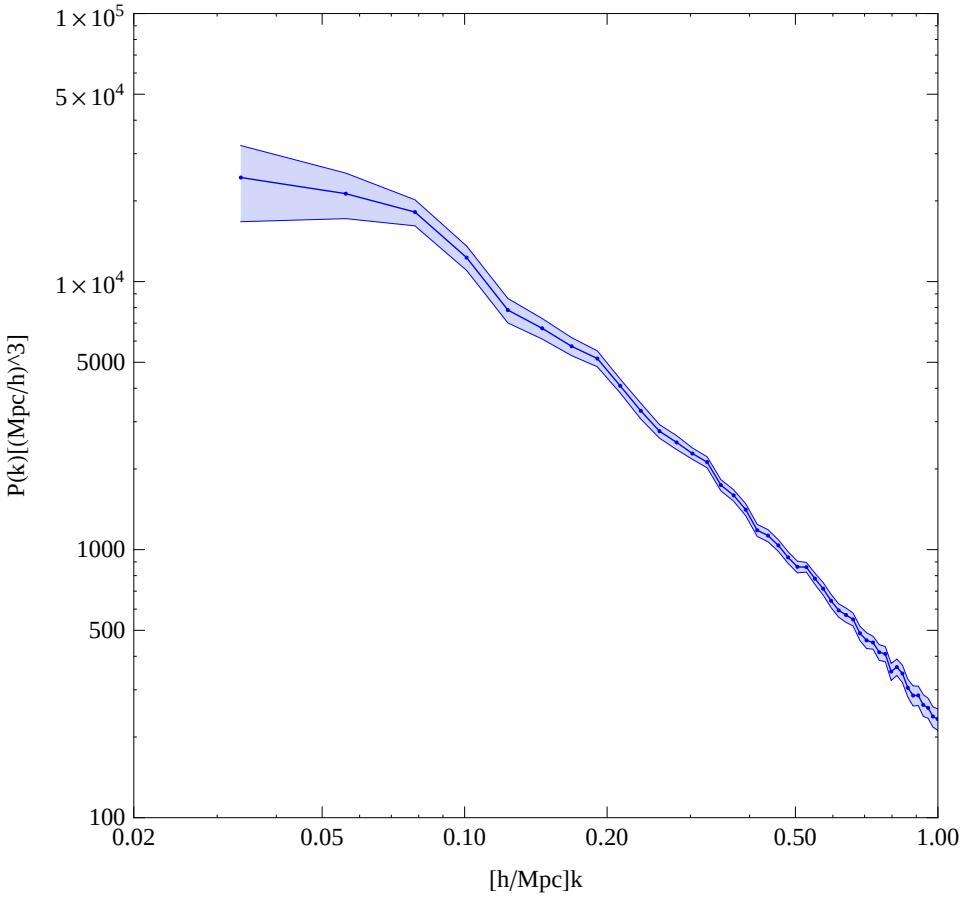


Finally we compute the power spectrum with the FKP estimator described previously. For this goal we include the catalog in a cube, i.e. with a size equal to 280 Mpc h^{-1}, which is the equal to the diameter of the spherical considered sample, zero padding the galaxy density field outside the surveyed volume. To apply the FKP estimator we need to set a representative value for the spectral power on the scale of interest, to set the statistical weight. We use P =8000 $(Mpc/h)^3$, which, as we verified a posteriori, is approximatively equal to the amplitude of the power spectrum in the range $0.1-0.15$ h Mpc^{-1}. We have verified that changing this value does not significantly affect the result. The measured power spectrum has been computed in k-bins $\Delta k = 2\pi/280h$ Mpc^{-1} and it is plotted in fig. 4.11 with 1σ errors obtained from the scatter among the 95 2MRS mocks. In order to compute the error bars we take the 95 mock realizations of the 2MRS, described in the in Section 4.3.1. They match the real catalog geometry, luminosity cut and selection function. In the plot 4.11 the measured power spectrum is represented by the small dots connected with a continous line. The shaded area represents the uncertainty strip.

This power spectrum is convolved with the window function of the catalog. To infer the true galaxy power spectrum we need to estimate the

Fourier transform of the latter. This was done by computing the power spectrum of the random catalog on 1024^3 grid with size $L = 1400$ Mpc h^{-1} to match a resolution of $\Delta k = 2\pi/1400h$ Mpc^{-1}. The result is shown in Fig. 4.12. The geometry of the 2MRS catalog is close to spherical. This is evident in the same figure when we overplotted the Fourier transform of a spherical top hat (blue curve) to the actual window function.

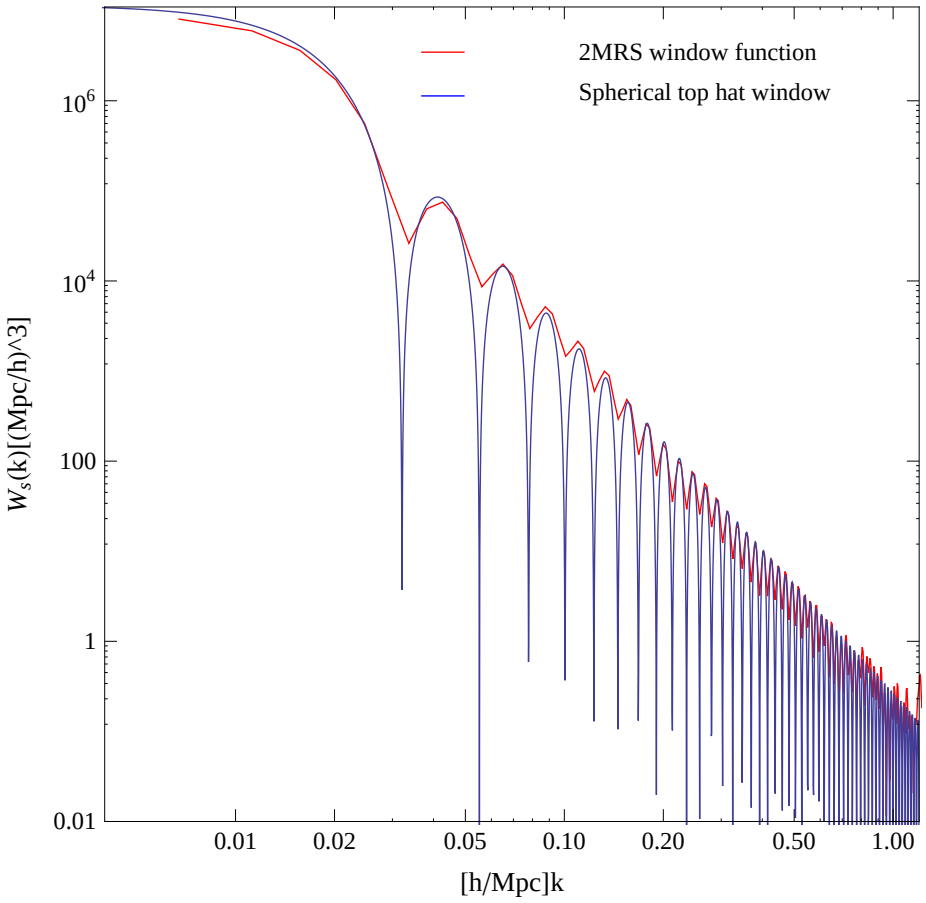
Figure 4.11: The power spectrum of the 2MRS galaxy sample is represented by blue dots connected with continous line. 1σ errors from the *rms* scatter among 95 mock catalogs are represented by the shaded area.



4.5 Relative bias estimates

As a first application of the $P(k)$ estimator we have established so far, we compute the relative bias of different galaxy populations within the 2MRS sample. The rationale behind this choice is that this estimate is model

Figure 4.12: Spherically averaged window function, $W_s(k)$, of the 2MRS galaxy sample, represented by the red line. It corresponds to the power spectrum of the random catalog, computed on 1024^3 grid with size $L = 1400$ Mpc h^{-1} to match a resolution of $\Delta k = 2\pi/1400h$ Mpc^{-1}. The blu line denotes the spherical top hat with radius R =140 Mpc h^{-1}. The plot evidences that the geometry of the 2MRS catalog is close to spherical.



independent: as long as we consider linear scales, and all objects trace the same velocity field (i.e. no velocity bias), the relative bias does not depend on the underlying cosmology. For this purpose we first split the main sample in luminosity bins and compare the power spectra of these subsamples, then we repeat the same exercise dividing the 2MRS sample in two populations with galaxies of different morphological types.

4.5.1 Bias as a function of luminosity

It is a well known fact that galaxy bias depends on luminosity. More precisely it increases with the luminosity of the objects, a fact that we interpret with the hypotesis that the more the luminous is the object the more massive is its host dark matter halo. This trend was clearly detected in several galaxy samples (e.g. Norberg et al. 2001, 2002 ; Zehavi et al. 2011). Estimates at $z \sim 0$ have been limited by the restricted number of objects in the sample. For this reason we tackle this problem here, exploiting the angular size and number of objects in the 2MRS catalog. For this purpose we divide the main sample, defined in Section 4.4.1, in 4 absolute magnitude bins, dubbed S1, S2, S3, S4. Each of them contains a similar number of objects to guarantee similar errors in the bias estimates. Table 4.4 summarize the main characteristics of these subsamples.

Table 4.4: Luminosity (not volume limited) subsamples in the 2MRS catalog. $m_{Ks,cut}$ = 11.75mag. All the magnitude are isophotal magnitudes. $< M_{Ks} >$ is the average magnitude of each bin. N denotes the number of objects in each sample, after rejecting all galaxies beyond the maximum distance of 140 Mpc h^{-1}.

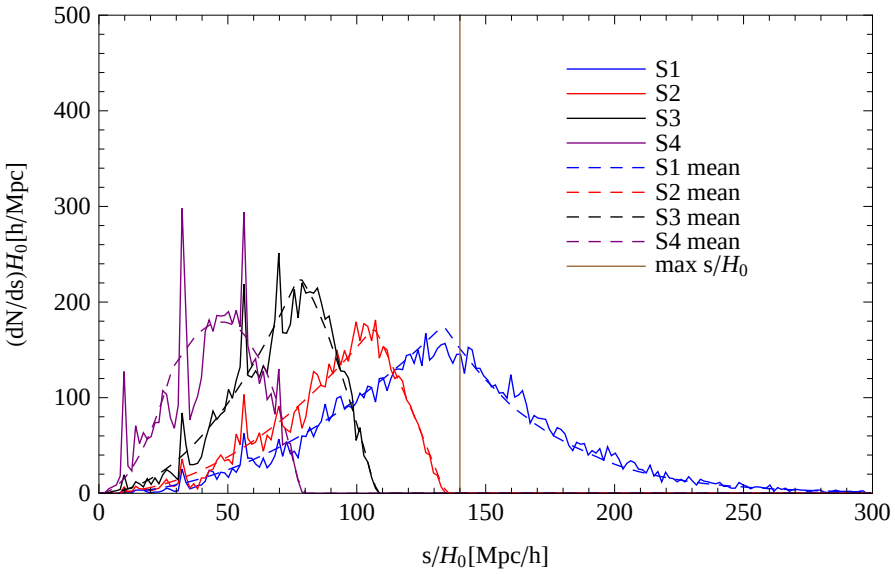
	$M_{Ks,range}[mag]$	$< M_{Ks} > [mag]$	N
S1	$M_{Ks} \leq -24$	-24.36	8402
S2	$-24 < M_{Ks} \leq -23.5$	-23.75	8678
S3	$-23.5 < M_{Ks} \leq -22.8$	-23.18	9749
S4	$M_{Ks} > -22.8$	-22.16	8075

We compute the mean number density and the dN/ds of each subsample through F/T method and build a random catalog matching the computed mean number density for each subsample. Fig. 4.13 shows the measured dN/ds (continous histogram) and the one obtained from the F/T method (dashed curves) for each subsample, both computed in bins Δs =150 Km s^{-1}. The samples are characterized by different median redshifts as a result of the luminosity cuts. The brightest objects have, on average, largest redshifts than the faint ones, as expected. The vertical line in fig. 4.13, drawn at 140 Mpc h^{-1}, corresponds to the maximum distance of the objects used to estimate the power spectrum. All the objects beyond this distance are not considered in the analysis.

For each subsample we compute the power spectrum using the FKP estimator in the same computational box with lenght 280 Mpc h^{-1} and 256^3 grid points centered at the objects' central position. The use of the same computational box for the luminosity bins makes easier the comparison between the power spectrum measures because in this case the $P(k)$ is computed in the same k bins for all the subsamples. The power spectra

of each subsample are shown in fig. 4.14.

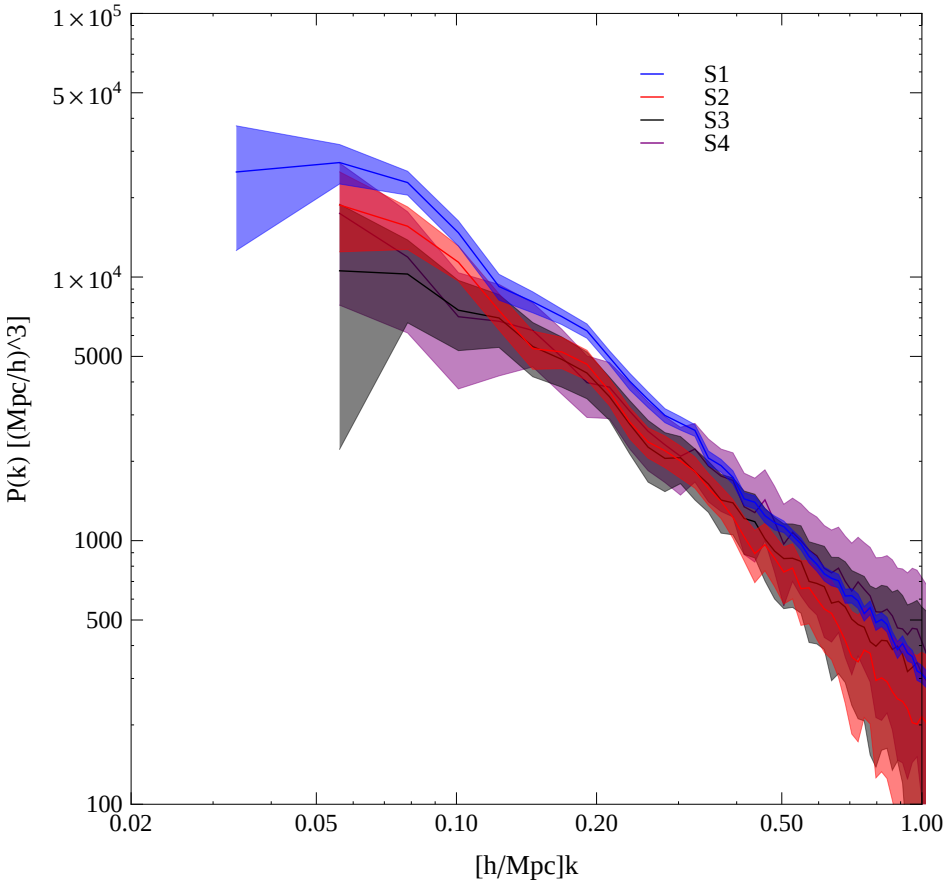
Figure 4.13: The solid lines correspond to the observed dN/ds of each 2MRS magnitude (not volume limited) subsample computed in bins Δs=150 Km s^{-1}. The absolute magnitude range and the number of objects in each subsample is indicated in Table 4.4. The dashed lines represent the mean dN/ds of each subsample computed with F/T method (see text for details). The vertical line denotes the maximum distance from the observer. All the objects beyond this distance are not considered in the analysis.



We notice that on large scales, except for the less luminous bin, the amplitude of the spectra increases with the luminosity, which is expected. This reflects the different bias of each sample. In the linear regime we can estimate the bias from the ratio of the power spectra: $\sqrt{PSi/PSj}$, with PS = power spectrum, i ranges from 1 to 4 and identifies each luminosity subsample and $j = 3$, i.e. relative biases are referred to objects in the luminosity range $-23.5 < M_{Ks} \leq -22.8$. The choice of S3 as reference sample allows a more quantitative comparison with the results of Westover (2007) who computed the relative bias of 2MRS galaxies using the 2-point correlation function. In that case the pivot at b value was that corresponding to M^* galaxies. In this case, M^* corresponds to -23.43, [1], a value that lies

[1] M^* is the characteristic luminosity of the sample and corresponds to characteristic lumnosity L^* of the Schecher-like luminosity function of 2MRS galaxies. Schecher function

Figure 4.14: Power spectrum of each magnitude subsample: S1, S2, S3 and S4. The absolute magnitude range and the number of objects in each subsample is indicated in Table 4.4. The shaded areas represent $1-\sigma$ errors from the scatter among the mocks. The minimum k is higher for the fainter samples, due to the smaller volume covered by them.



within the magnitude range of the S3 sample. Fig. 4.15 shows in blue the spectral ratios $\sqrt{PSi(k)/PS3(k)}$, for each subsample. In all cases, but S4, the spectral ratio drops at $k \simeq 0.1\ h$ Mpc^{-1}. The possible origin of the drop at large scales is discussed hereinafter in this Section. Since we are interested in the linear regime, where it make sense to compute relative biases from the spectral ratio, we shall compute the relative bias using large scales only.

To estimate errors we have repeated the same procedure as in the real

is given by: $\Phi(L)dL = \Phi^* \left(\frac{L}{L^*}\right)^{\alpha} e^{\left(-\frac{L}{L^*}\right)} \frac{dL}{L^*}$ where $\Phi(L)$ represents the luminosity function.

catalog on the 95 mocks. For each mock we extract four absolute magnitude subsamples. Each magnitude bin is defined as for the real catalog. Let's denote the 4 luminosity subsamples for the $j-$th mock as: $M1_j$, $M2_j$, $M3_j$, $M4_j$, where j is an integer index ranging from 1 to 95 that identifies the mock catalog. The mean redshift distribution of mock galaxies in the different subsamples is shown in fig.4.16. The agreement with those of the real galaxies is remarkable since mock catalogs were not built to match these properties. These are the dN/ds used to build the random catalogs and to estimate the power spectrum of the mock subsamples.

For each mock we have computed the ratios $\sqrt{PMi_j/PM3_j}$, their mean and scatter among the mocks. In particular the *rms* scatter used to estimate the 1-σ error is:

$$\sigma_{i3} = \sqrt{\frac{\sum_{j=1}^{N}(\sqrt{PMi_j/PM3_j} - <\sqrt{PMi_j/PM3_j}>)^2}{N-1}} \qquad (4.6)$$

where N is the total number of mocks and $<>$ is the average over all mocks.

To estimate the relative bias from the spectra ratios shown in fig.4.15 we fit a constant function to the data points, each one weighted by its error bar, over a limited range of wave numbers. In the linear regime this constant function is equal to the ratio of the linear biases of the two subsamples times the ratio of the Kaiser boost factors of the two populations. The Kaiser boost factor, eq. 1.58, also depends on galaxy bias. It turns out that the actual dependence is small and could be safely ignored, which we did in the following analysis. If one accounts for it then the estimated bias ratio decreases by ~ 4 % for the brightest sample and remains constant in the faintest. We fit the bias ratios in two different k-ranges $[k_{min}, k_{max}]$ for each subsample. For k_{max} we have considered two different values, 0.15 and 0.3 h Mpc^{-1}, to check the robustness of the results and the impact of nonlinearity, whereas we set $k_{min} = 0.067$ h Mpc^{-1}, a value significantly larger than the fundamental mode of the cube used to compute the power spectrum $2\pi/280 = 0.022$ h Mpc^{-1}. This is a conservative choice that reflects the need to consider the range of scales accessible to all the luminosity-subcatalogs. Since the faintest subsample is also the smallest one, the criterion used to set k_{min} is as follows. We consider the diameter of the smallest (faintest) subsample, D_{min}, and set $k_{min} = 2\pi/D_{min} + \Delta k = 0.067$ h Mpc^{-1}, where Δk is the binsize. The rationale of adding the extra Δk term is to reduce the impact of the window function. The advantage is that we try to eliminate any systematics induced by comparing samples with different window functions. The drawback, of course, is to erase information on large scales.

We estimate the bias ratios for the four different subsamples as a function of the luminosity. The bias ratios are measured with respect to the S3 sample: $b/b(S3)$. For this reason, in the S3 case the bias ratio is set equal to

unity and has no error. We set the luminosity L for each subsample equal to the mean of the different luminosity bins.

To compare our results with other works, we show, in fig. 4.17, the estimated bias ratios with respect to L^* (as already anticipated we set the corresponding absolute magnitude value at $M^* = -23.43$, computed in Westover, 2007, for the 2MRS galaxies). M^* doesn't coincide with the mean absolute magnitude of galaxies in the S3 magnitude bin (although M^* lies in that bin), for this reason L/L^* is different from unity for the S3 subsample. At the same time we shift the relative bias values $b/b(S3)$ to b/b^* assuming that, for the S3 sample, the relative bias b/b^* is equal to the result of an independent work by Norberg et al., 2001, (see below for details about this work). The variations in the b/b^* ratios due to this shift are around 3%, well within the error bars for the S1, S2 and S4 samples. Filled dots show the case of $k_{max} = 0.15$ h Mpc^{-1} and open dots the case $k_{max} = 0.3$ h Mpc^{-1}. The latter were slightly shifted for clarity. Error bars represent the *rms* scatter from the mocks. The smallness of the error bars on the bias ratio reflects the fact that, the cosmic variance is negligible in this case since the different subcatalogs sample the same structure in largely overlapping regions of the Universe.

We note a trend of more luminous objects to be more biased than the faint ones. This is expected. At low luminosity the relative bias is quite constant. Indeed the faintest point suggests that bias may be higher for the faintest sample. However, the significance of this trend is well within 1-σ. The fact that the two sets of points are in good agreement is very reassuring since it possibly indicates that our bias estimates are not significantly affected by nonlinear effects.

To compare these results with previous tests and theoretical predictions we overplot to the datapoints four different curves. The black one is the best fit proposed by Westover (2007) who performed a similar analysis, based however on the two-point correlation function, using an earlier version of the 2MRS catalog that contained nearly half of the objects, with $K_s < 11.5$. In that analysis six different luminosity subsamples roughly matching ours were selected and the redshift space correlation function was estimated for each of them. The bias ratio was then computed from the amplitude of the correlation function at a separation of 2.4 Mpc h^{-1}, i.e. well into the nonlinear regime. The blue curve is also from Westover (2007) and is an analogous fit that uses the *projected* correlation function at a separation of 5.9 Mpc h^{-1}, hence simultaneously eliminating redshift distortions and alleviating the impact of nonlinearity. These analytic fits have the form

$$b/b^* = A + C(L/L^*)\,,$$

where A and C are the free parameters of the model and L^* is the characteristic luminosity of the Schecher fit to the galaxy luminosity

function. In the two cases the best fit parameters are $A = 0.82 \pm 0.08\,, C = 0.11 \pm 0.05$ and $A = 0.73 \pm 0.07\,, C = 0.24 \pm 0.04$, respectively.

None of the two curves seem to fit our data. The black curve lies systematically below our data points, which is somewhat difficult to explain in terms of nonlinear effect. We notice that whereas the bias ratio measured by Westover (2007) at $L = L^*$ is indeed equal to unity, this value represents a local maximum in the measured b/b^* vs L relation. Since the bias ratio is set equal to unity at $L = L^*$, this seems more an issue of incorrect normalisation. Indeed, if the black curve is shifted upwards in the plot so that is passes through the pivot value at $L = L^*$ the quality of the fit improves considerably. The blue curve is significantly steeper and does not provide a good fit to our data points. Again, this is not surprising. That curve is a fit to a clustering analysis performed in real space and probing the nonlinear regime, whereas we made an effort to estimate the relative bias in the linear regime and in redshift space. We do not expect that linear redshift distortions may significantly affect the result of our analysis. This is certainly true when the bias ratio is equal to unity, i.e. for the faint samples, but is also true for the bright samples where $b/b^* \sim 1.2$, unless the absolute bias of bright 2MRS galaxies is very large, which is not the case, as we shall see below. As anticipated we estimate that the effect is of the order of a few per cent. Therefore, the mismatch must be attributed to nonlinear effects. Which is indeed plausible, since we are probing scales in the range $[30, 100]$ Mpc h^{-1}, much larger than those considered in Westover's analysis.

The red dashed curve is also a best fit and refers to 2dFGRS galaxies, Norberg et al. (2001). As for the blue curve, these values were estimated from the projected galaxy-galaxy correlation function at a separation of 4.9 Mpc h^{-1}. The fitted curve has the same shape as in the previous cases and parameters $A = 0.85\,, C = 0.15$ (no errors provided). The curve seems to provide an acceptable fit to our data as well, which is quite surprising given the fact that this analysis is performed in real space, non-linear regime and uses a different galaxy sample with different type composition and at slightly higher redshifts than 2MRS. Finally, the black shaded curve refers to the bias ratio of an even higher redshift sample, that of SDSS galaxies (Tegmark et al., (2004)). In this case the relative bias was estimated from the power spectrum, obtained from the galaxy power spectra, in real space, on scales that are similar or larger than ours. Therefore, in this case the main differences with our estimates arise from the different composition of the sample and its redshift. This curve, $b/b^* = A + C(L/L^*) + B(M - M^*)$, with parameters $A = 0.85$, $C = 0.15$, $B = -0.04$, provides a good fit to our data points except the faintest one.

In order to check the robusteness of our results, we repeat the same analysis as before with volume limited samples rather than flux limited. This would "clean" each subsample from any incompleteness or selection effects that we may have been underestimated. In order to extract volume limited

samples we consider the subcatalogs S1,S2 and S3 and reject, for each of them, all the objects beyond a maximum distance from the observer R_{max}. In the parent luminosity-subsamples the mean galaxy density decreases beyond R_{max} because of the apparent magnitude cut of the survey. R_{max} provides then the location of the maximum of the dN/ds curve (see fig.4.13). The volume limited subsamples, dubbed $S1_{VL}$,$S2_{VL}$ and $S3_{VL}$, are then characterised by a constant mean number density of objects. The number of objects, the absolute magnitude range and the maximum distance from the observer for each subsample are reported in Table 4.5. We exclude from the analysis the faintest sample ($S4_{VL}$) due to the limited size of its volume.

Table 4.5: Luminosity subsamples in the 2MRS catalog, volume limited. All the magnitude are isophotal magnitudes. N denotes the number of objects in each sample. R_{max} indicates the maximum distance from the observer for each subsample. The samples are obtained rejecting all the objects beyond R_{max} in the subsamples S1, S2 and S3 described in Table 4.4. The subsamples Si_{VL80} are characterized by the same volume ($R_{max} = 80$ Mpc h^{-1}).

	$M_{Ks,range}[mag]$	$R_{max}[h^{-1}Mpc]$	N
$S1_{VL}$	$M_{Ks} \leq -24$	130	6918
$S2_{VL}$	$-24 < M_{Ks} \leq -23.5$	110	6764
$S3_{VL}$	$-23.5 < M_{Ks} \leq -22.8$	80	6232
$S1_{VL80}$	$M_{Ks} \leq -24$	80	1496
$S2_{VL80}$	$-24 < M_{Ks} \leq -23.5$	80	2496
$S3_{VL80}$	$-23.5 < M_{Ks} \leq -22.8$	80	6232

As before, for each subsample we compute the power spectrum in the same computational box with lenght 280 Mpc h^{-1} and 256^3 grid points centered at the objects' central position. In fig.4.15 we show the spectral ratios computed in case of volume limited samples (red squares): $\sqrt{PSi_{VL}(k)/PS3_{VL}(k)}$, $i = 1, 2$. The error bars are from the scatter among 95 volume limited mock catalogs. We note that the results agree quite well within the error bars with the previous ones (in blue). Fig.4.18 shows the bias versus luminosity results, computed in the same k ranges used in the analysis of the parent luminosity-subsamples. The results are quite similar to the ones obtained with the parent luminosity-subsamples S1, S2 and S3 (shown in fig.4.17): they totally agree when we restrict the analysis below $k_{max} = 0.15$ h Mpc^{-1}. This implies that the effect of having excluded the external part of the original samples (the more luminous objects) does not change significantly the results and confirms their robustness. At the same time the drop in the power ratio at $k = 0.15$ h Mpc^{-1} is still evident in the volume limited case (fig.4.15), especially when comparing $PS1_{VL}$ and $PS3_{VL}$ and is at the origin of the discrepancies between the linear analysis

(restricted to $k_{max} = 0.15$ h Mpc^{-1}) and the nonlinear one (that extends out to $k_{max} = 0.3$ h Mpc^{-1}).

To corroborate the robustness of the luminosity dependence of the bias we performed an additional test in which we we further reduce the volume of the $S1_{VL}$, $S2_{VL}$ and $S3_{VL}$ by keeping only objects within $R_{max} = 80$ h Mpc^{-1}, i.e. by comparing volume limited samples of the same size but different number density. The main characteristics of the subsamples, dubbed $S1_{VL80}$, $S2_{VL80}$ and $S3_{VL80}$, are reported in Table 4.5. For each subsample we compute the power spectrum in the same computational box with lenght 160 Mpc h^{-1} (two times the maximum distance from the observer) and 256^3 grid points centered at the objects' central position. Fig.4.15 shows the spectral ratios (green triangles). The error bars are from the scatter among the mocks. We note that the drop in the power ratio bias has decreased, significantly when comparing $PS2_{VL80}$ and $PS3_{VL80}$.The results for the relative bias are shown in fig.4.19. We note that they are more stable than the previous two cases with respect to k_{max}.

In the light of these results we conclude that the drop in the power ratio at $k = 0.15$ h Mpc^{-1} is not an intrinsic feature but probably reflects the uncertainties in the estimate of the selection function at large distances, or possibly some evolutionary effect that was not modelled properly. Nonlinear effects may also play a role. Note, however, that because of the limited number of objects and the size of the survey the significance of this drop is barely 1σ. Table 4.6 summarizes the results of bias versus luminosity obtained.

To summarize, in this study we find a dependency of galaxy bias on luminosity. The qualitative trend is in agreement with the results of previous studies: a weak luminosity dependence at low luminosities and an increase above L^*. The increase is particularly sharp for the brightest galaxies in the sample. A more quantitative comparison, however, is hampered by several effects. Norberg et al. (2001) and Tegmark et al. (2004) use samples that are different from ours in terms of redshift and galaxy composition. In addition, our analysis is performed in redshift space whereas the other cases shown in the plot, apart from the continuous black curve, refer to real space analyses. In the linear regime the correction would be trivial, the Kaiser boost, and rather small. But one should also account for the fact that correction for nonlinear motions may reverberate to large scales when Fingers of God associated to massive clusters are compressed (see e.g. fig.55 in Tegmark et al., 2004). This of course depends on the details of the compression, the prominence of the Fingers of God features in the redshift maps and the type of objects in the catalog. In the 2MRS case, which preferentially contains late type galaxies and in which we have only compressed a handful of clusters, the effect should not be large. However, a dedicated analysis would be needed to quantify the effect. Finally, the different analyses have focused on different scales. Those of Westover (2007)

and Norberg et al. (2001) have considered small, highly nonlinear scales, in which both dynamical nonlinear effects and scale-dependent bias affect the result of the analysis. Again, quantifying the effect would require a proper analysis that can only be based on realistic mock catalogs of all samples considered in this comparisons.

As for the analysis presented here, a few improvements can be foreseen.The effect of the window function of each subsample should be explicitly accounted for, when considering different volume samples. However, the stability of our results to the choice of k_{min} indicates that the improvement in this sense would be marginal. Another improvement of the analysis could be to adjust the statistical weight of the FKP estimator according to the subsample considered rather than using a unique weight, calibrated on the full 2MRS catalog, for all sub-catalogs considered.

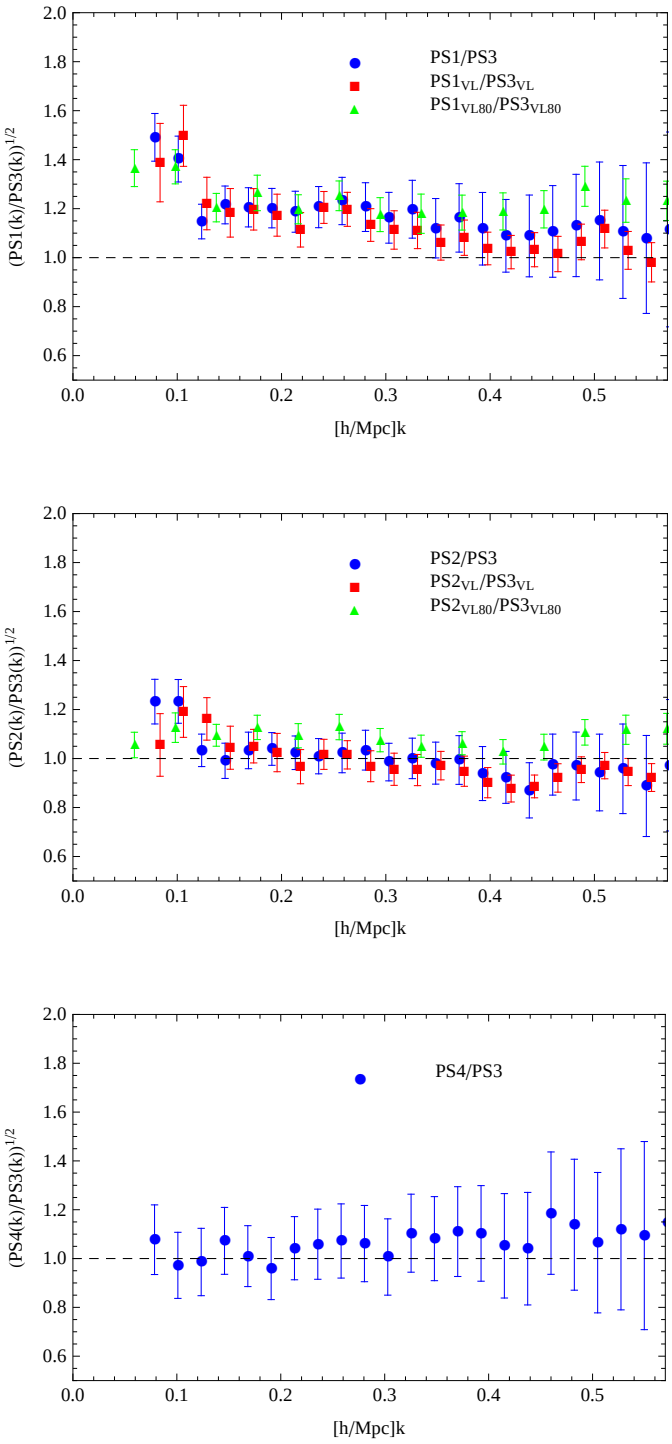


Figure 4.15: Power spectrum ratios vs. the wavenumber k: in blue $\sqrt{PSi(k)/PS3(k)}$, $i = 1, 2, 4$. Red: $\sqrt{PSi_{VL}(k)/PS3_{VL}(k)}$, $i = 1, 2$ (volume limited samples). Green: $\sqrt{PSi_{VL}(k)/PS3_{VL}(k)}$, $i = 1, 2$ (volume limited and same maximum distance (80 Mpc h^{-1})) Error bars are given by the scatter among the spectral ratios among the mocks. The red squares are slightly shifted to higher k values.

Figure 4.16: Mean dN/ds of mock subsamples computed in bins Δs=150 Km s^{-1}. As done for the real catalog, we divide each of the 95 2MRS mocks in four absolute magnitude subsamples (magnitude bins are indicated in Table 4.4). Each solid line corresponds to the mean dN/ds of the 95 mock subsamples relative to the same magnitude bin. The vertical line denotes the maximum distance of the subsamples. As for the real catalog, all objects beyond that distance are not considered in the analysis. The dN/ds is shown also for distances larger than that.

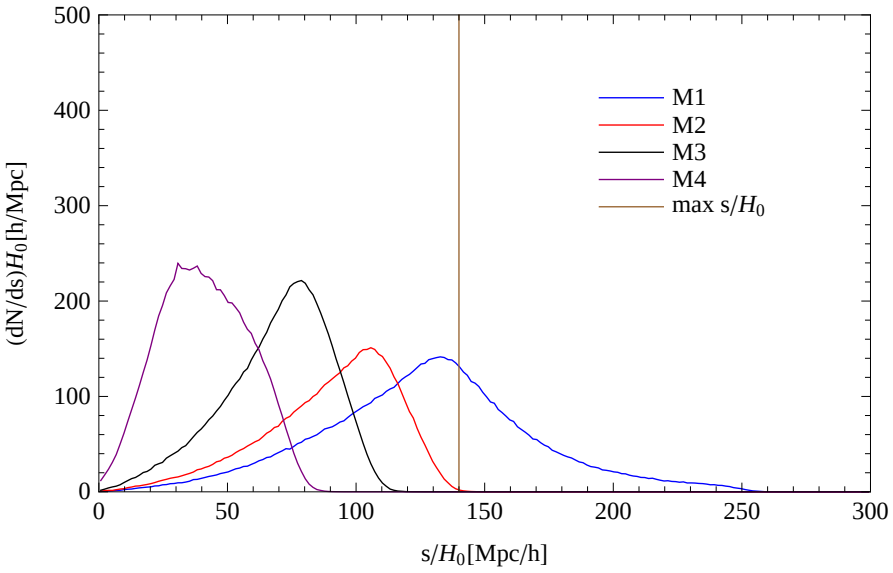


Figure 4.17: Bias versus luminosity obtained using the flux limited samples S1, S2, S3 and S4. The filled and open dots denote our results with the 2MRS galaxies for $k_{max} = 0.15$ and $k_{max} = 0.30$ h Mpc^{-1} respectively, both redshift space. The open dot circles are slightly shifted to higher L values. The black curve denotes the Westover (2007) best fit model to the bias vs. luminosity results obtained by measuring the ratio between the correlation functions of 2MRS galaxies in redshift space at $r = 2.4$ Mpc h^{-1}: $b/b^* = 0.82 + 0.11L/L^*$. The blue curve denotes the Westover (2007) best fit model to the bias vs. luminosity results obtained from the ratio between the projected correlation functions of 2MRS galaxies at $r = 5.9$ Mpc h^{-1}: $b/b^* = 0.73 + 0.24L/L^*$. The red dashed curve and the black dashed one correspond respectively to Norberg (2001) model for 2dF galaxies, with $b/b^* = 0.85 + 0.15L/L^*$, obtained using projected correlation functions at a separation of $r = 4.9$ Mpc h^{-1}, and Tegmark (2004) model for SDSS galaxies, with $b/b^* = 0.85 + 0.15L/L^* - 0.04(M - M^*)$, obtained using the real space power spectrum at scales similar or larger than ours.

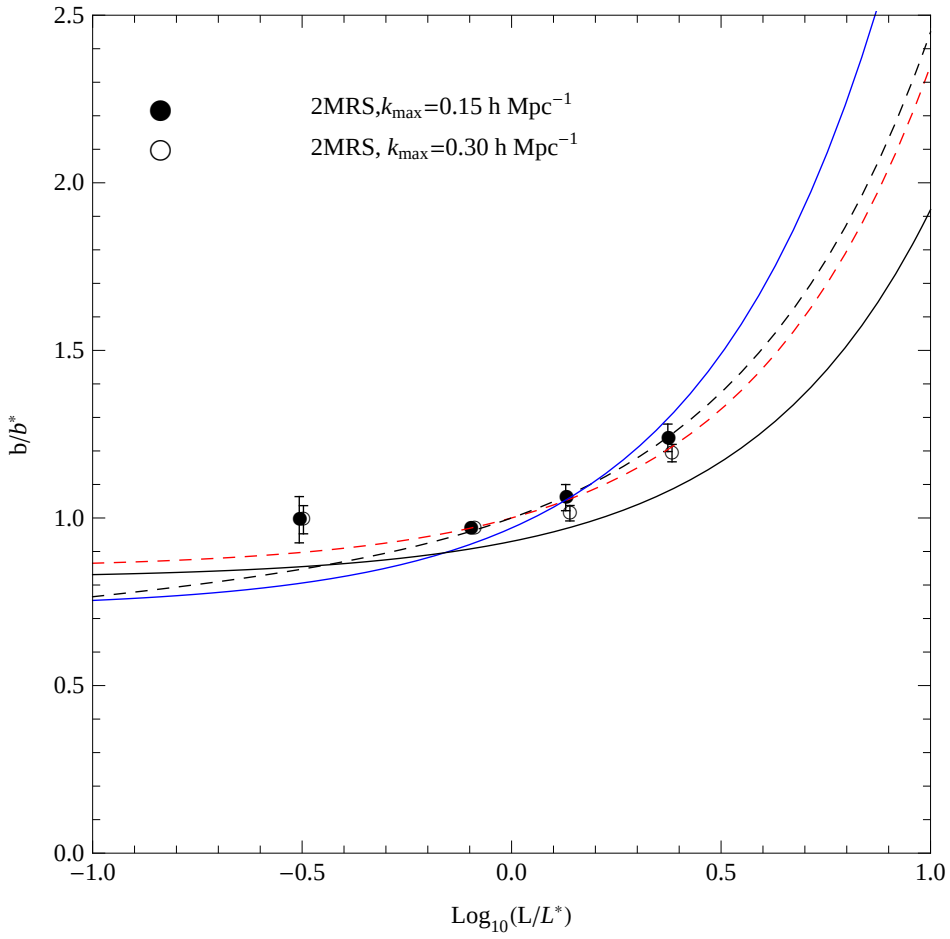


Figure 4.18: Bias versus luminosity with volume limited samples. The filled and open dots denote our results with the 2MRS galaxies for $k_{max} = 0.15$ and $k_{max} = 0.30$ h Mpc^{-1} respectively, both redshift space. The open dot circles are slightly shifted to higher L values. The black curve denotes the Westover (2007) best fit model to the bias vs. luminosity results obtained by measuring the ratio between the correlation functions of 2MRS galaxies in redshift space at $r = 2.4$ Mpc h^{-1}: $b/b^* = 0.82+0.11L/L^*$. The blue curve denotes the Westover (2007) best fit model to the bias vs. luminosity results obtained from the ratio between the projected correlation functions of 2MRS galaxies at $r = 5.9$ Mpc h^{-1}: $b/b^* = 0.73+0.24L/L^*$. The red dashed curve and the black dashed one correspond respectively to Norberg (2001) model for 2dF galaxies, with $b/b^* = 0.85 + 0.15L/L^*$, obtained using projected correlation functions at a separation of $r = 4.9$ Mpc h^{-1}, and Tegmark (2004) model for SDSS galaxies, with $b/b^* = 0.85 + 0.15L/L^* - 0.04(M - M^*)$, obtained using the real space power spectrum at scales similar or larger than ours.

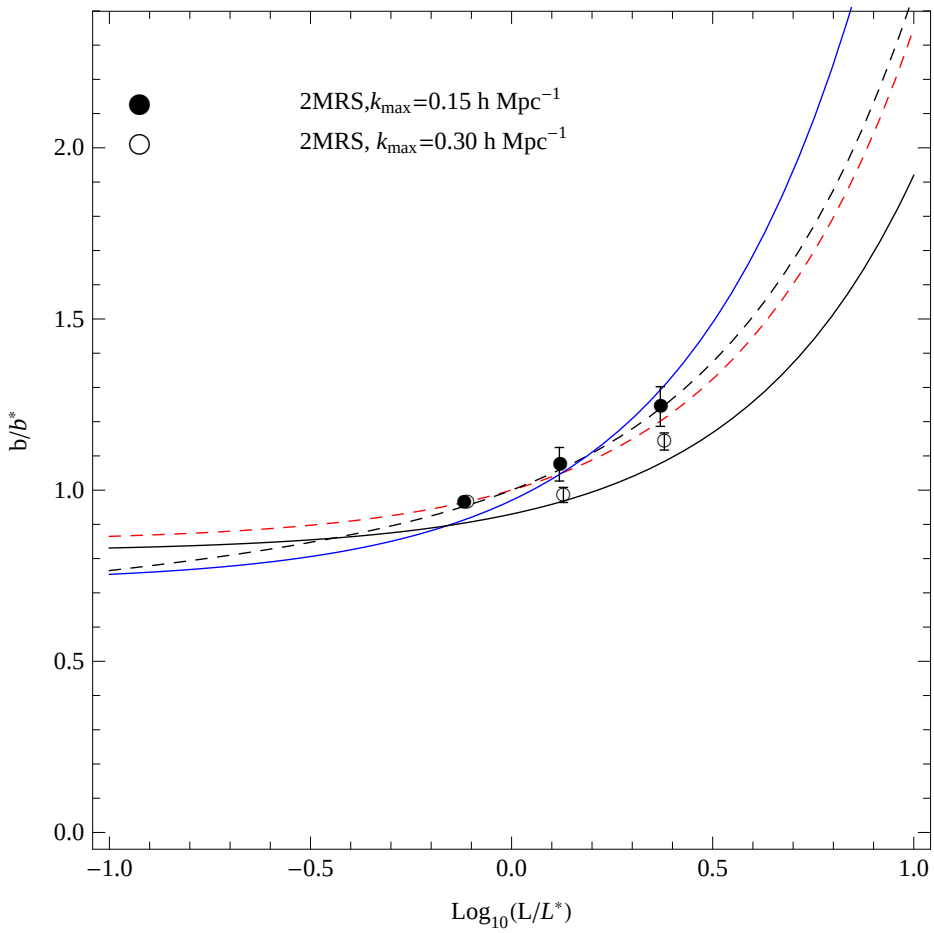


Figure 4.19: Bias versus luminosity with samples that trace exactly the same spatial region (maximum distance from the observer 80 Mpc h^{-1}). The filled and open dots denote our results with the 2MRS galaxies for $k_{max} = 0.15$ and $k_{max} = 0.30$ h Mpc^{-1} respectively, both redshift space. The open dot circles are slightly shifted to higher L values. The black curve denotes the Westover (2007) best fit model to the bias vs. luminosity results obtained by measuring the ratio between the correlation functions of 2MRS galaxies in redshift space at $r = 2.4$ Mpc h^{-1}: $b/b^* = 0.82 + 0.11L/L^*$. The blue curve denotes the Westover (2007) best fit model to the bias vs. luminosity results obtained from the ratio between the projected correlation functions of 2MRS galaxies at $r = 5.9$ Mpc h^{-1}: $b/b^* = 0.73 + 0.24L/L^*$. The red dashed curve and the black dashed one correspond respectively to Norberg (2001) model for 2dF galaxies, with $b/b^* = 0.85 + 0.15L/L^*$, obtained using projected correlation functions at a separation of $r = 4.9$ Mpc h^{-1}, and Tegmark (2004) model for SDSS galaxies, with $b/b^* = 0.85 + 0.15L/L^* - 0.04(M - M^*)$, obtained using the real space power spectrum at scales similar or larger than ours.

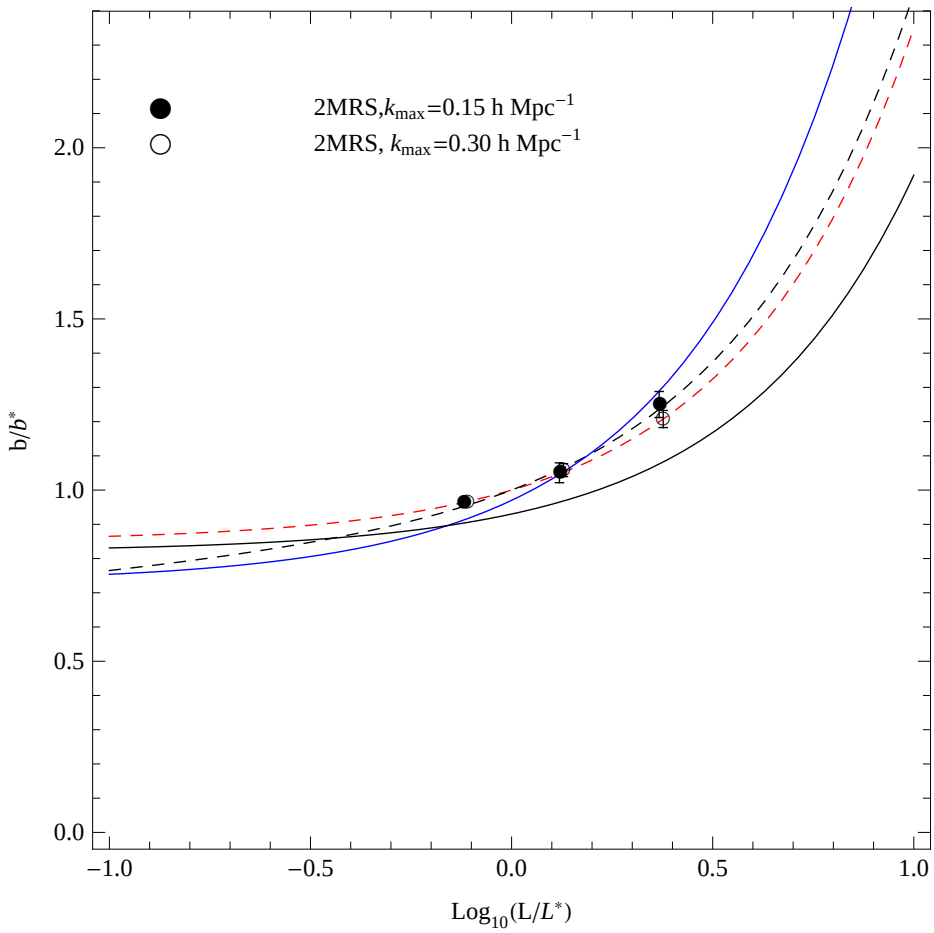


Table 4.6: Bias versus luminosity. FL indicates the original not volume limited subsamples extracted from the 2MRS main sample (S1,S2,S3,S4). VL indicates that the bias versus luminosity results are obtained from volume limited subsamples, obtained rejecting all objects beyond a maximum distance (R_{max}) from the original subsamples. VL80 indicates that we use samples which cover the same volume ($R_{max} = 80$ Mpc h^{-1} for all the subsamples). In the last two cases, VL and VL80 we exclude from the analysis the faintest luminosity bin. The results are obtained fitting the spectral ratios shown in fig. 4.15 with a constant function in the k range $[k_{min}, k_{max}]$. In all cases $k_{min} \sim 0.067$ Mpc^{-1} h. The value of L^* is assumed to correspond to the absolute magnitude value $M^* = -23.43$ computed in Westover (2007) for the 2MRS galaxies. The value of L for each luminosity bin corresponds to the average luminosity value inside that bin.

	$k_{max}[Mpc^{-1}h]$	L/L^*	b/b*
FL	0.15	2.36	1.239±0.041
FL	0.15	1.35	1.061±0.039
FL	0.15	0.80	0.970
FL	0.15	0.31	0.995±0.069
FL	0.30	2.36	1.194±0.026
FL	0.30	1.35	1.014±0.023
FL	0.30	0.80	0.970
FL	0.30	0.31	0.995±0.042
VL	0.15	2.34	1.244±0.058
VL	0.15	1.31	1.076±0.049
VL	0.15	0.76	0.964
VL	0.30	2.34	1.142±0.025
VL	0.30	1.31	0.986±0.022
VL	0.30	0.76	0.964
$VL80$	0.15	2.33	1.250±0.038
$VL80$	0.15	1.31	1.051±0.029
$VL80$	0.15	0.76	0.964
$VL80$	0.30	2.33	1.208±0.025
$VL80$	0.30	1.31	1.058±0.019
$VL80$	0.30	0.76	0.964

4.5.2 Bias dependence on galaxy morphological type

Different types of galaxies have different clustering properties. Indeed, morphological segregation is a long known effect (Dressler et al., 1980). Our goal here is to quantify the bias dependence on galaxy type in 2MRS. To do this we have repeated the same analysis presented in the previous Section dividing, however, the catalog in 2 subsamples characterised by different morphological types, instead of four subsamples of different luminosities.

For this purpose we have considered the sample of 20860 galaxies with $|b| > 10$ deg and $m_{K_s} \leq 11.254$ mag for which the morphological T-type is provided. The T-type ranges from -7 for unclassified elliptical galaxies to 0 for S0/a galaxies and +9 for Sm galaxies (see de Vaucouleurs et al. (1976) for details on the galaxy classification). Irregular and peculiar galaxies are also classified as separate classes.

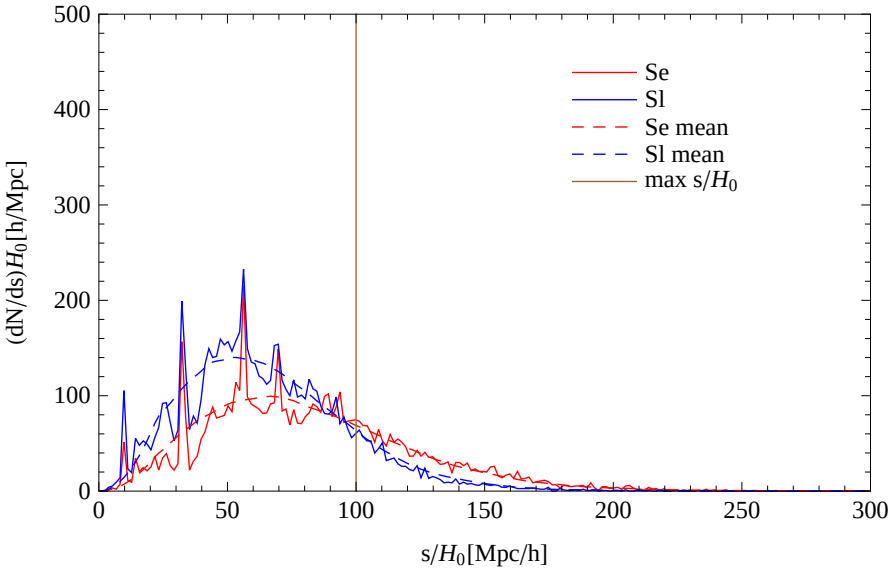
We divide the sample in two subsets: *Se* containing early type galaxies and *Sl* containing late type galaxies, using the following criteria: *Se* with T-type in the range [-9,-1] (elliptical and lenticular up to *S*0+ type) constitute the *early type* sample, galaxies with T-type in the range [0,+9] (SO/a and all spirals) are placed in the *Sl* subsample. Irregular galaxies and non classified spirals are excluded from the analysis. The two sets have different characteristics like their respective luminosity functions that, as shown in Kochanek et al. (2001), are quite robust to the selection criteria since shifting the T-type boundary between *early* and *late type* by +1 or -1 doesn't change the luminosity functions.

We extract a semi volume limited sample from both samples applying the same procedure described in Subsection 4.3.1 and consider all objects that do not belong to our Local Group and out a maximum distance of 100 Mpc h^{-1}. The radius of the samples is then smaller than for all sub-catalogs considered so far. The rational behind this choice is the brighter magnitude cut that we have used here (to guarantee a morphological classification to each object).

The two *Se* and *Sl* samples contain respectively 6538 and 9447 objects. Their dN/ds are plotted in fig. 4.20 (histograms) and compared with the same quantities estimated using the F/T method (dashed curve).

The power spectra of both subsamples were estimated using the same procedure described in the previous Subsection but using a smaller computational box of 200 Mpc h^{-1}. The result is shown in fig. 4.21 where the red and blue line refer to the early and late type time samples, respectively, as indicated in the plot. The shaded region represents the 1-σ uncertainty strip obtained from the *rms* scatter among the mocks. These mocks are not the same one used so far. The parent catalog extracted from the simulation does not contain any morphological information. Therefore, to build mocks mimicking the ones used in this analysis we decided to use color information to split the mock catalog into two subsamples with relative

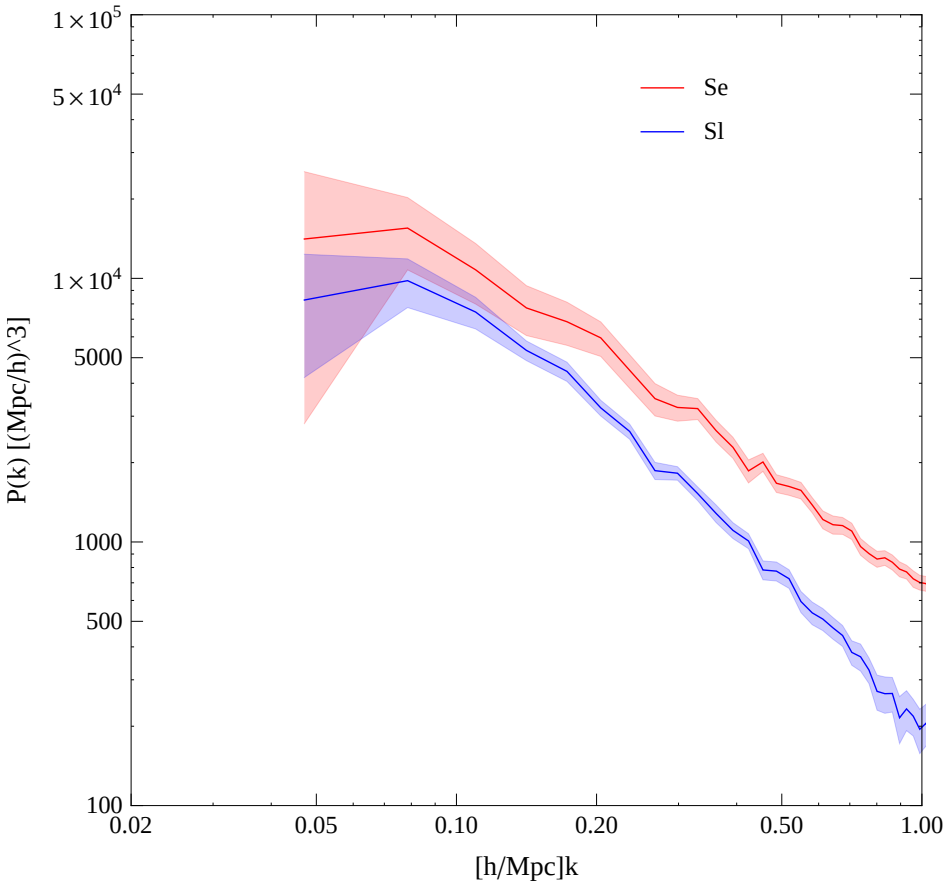
Figure 4.20: dN/ds of the two morphological subsamples: early type (Se) and late type (Sl) galaxies. The solid lines correspond to the observed dN/ds of each subsample computed in bins $\Delta s = 150$ Km s^{-1}. The dashed lines represent the mean dN/ds of each subsample computed with F/T method (see text for details). The vertical line denotes the maximum distance from the observer. All the objects beyond this distance are not considered in the analysis.



number of objects and characteristics similar to that of the real catalogs. The idea is to identify early and late type galaxies with red and blue galaxies, respectively. So all we had to do was to separate red and blue objects in the mocks where color information is available thanks to the fact that magnitudes are specified not only in the 2MRS near-infrared bands but also in the optical SDSS bands.

For this purpose we tried several different criteria, following different color-color or color-magnitude cuts proposed in the literature (Baldry et al. (2004), Zehavi et al. (2005), Zehavi et al. (2011), Loveday et al. (2012)). It turned out that the most effective in our case is the one proposed by Zehavi et al. (2005) that considers the relation between the $(g - r)$ colour and the M_r magnitude. The distribution of mock 2MRS galaxies in this plane (see fig. 4.22) shows a clear bimodality that can be exploited to separate the the two population. As shown in the plot, our only selection criterion was a color cut at $g - r$=0.75, slightly different from the one proposed by Zehavi

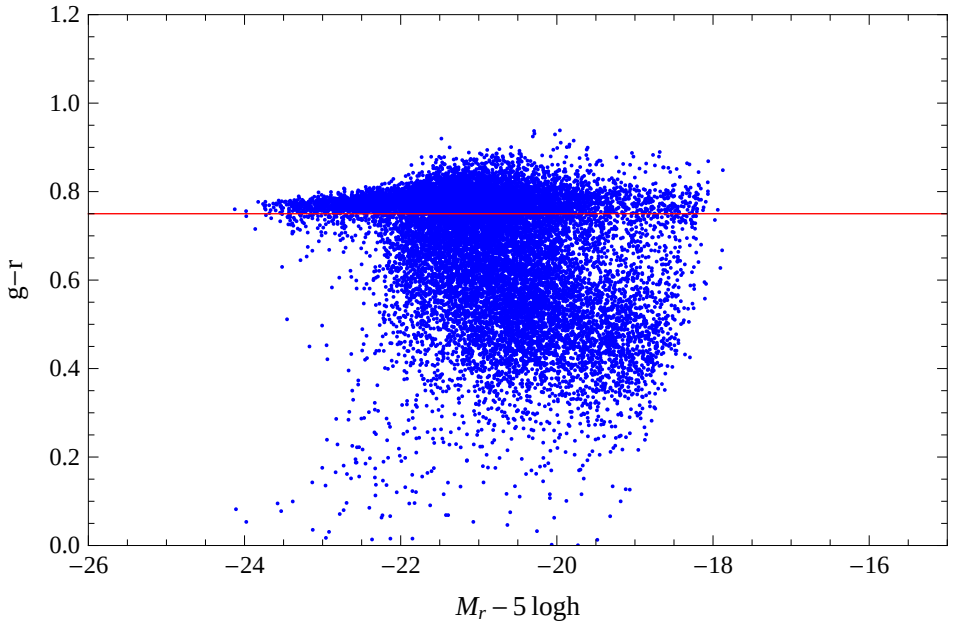
Figure 4.21: Power spectrum of early type (Se) and late type (Sl) galaxies. The shaded areas represent $1-\sigma$ errors from the scatter among the mocks.



et al. (2005) that, however, was calibrated for SDSS galaxies at significantly higher redshifts. The relative number of objects in the red vs. blue galaxies mock samples turned out to be close to that of the early vs. late galaxies in the real samples. A remarkable similarity was also found among the redshift distributions of the corresponding real and mock subsamples.

In fig. 4.23 we show the square root of the power spectrum ratio between the Se and Sl subsamples $\sqrt{PSe/PSl}$ as a function of the wavenumber. This plot, which basically corresponds to the bias ratio between the two samples, indicates that the ratio is rather constant across the k-range, i.e. it is rather insensitive to nonlinearities at least out to $k = 0.4$ Mpc^{-1} h, despite the fact that nonlinear effects are expected to be more prominent for the Se sample,

Figure 4.22: Colour-magnitude diagram of one 2MRS mock catalog. The red line shows the colour cut $g - r = 0.75$, applied to the mocks in order to separate red from blue galaxies.



since early type objects preferentially inhabit high density environments. Errorbars were computed from the mocks. For each mock we considered the power spectra of the blue and red subsamples and their ratio. Then we averaged over all ratios and considered the *rms* scatter that we used as a proxy for the 1-σ random error, i.e.

$$\sigma_{el} = \sqrt{\frac{\sum_{j=1}^{N}(\sqrt{Pr_j/Pb_j} - <\sqrt{Pr_j/Pb_j}>)^2}{N-1}} \tag{4.7}$$

where j runs over all N mocks and Pr and Pb indicate the power spectrum of the red and blue objects, respectively.

From this power ratio we derive an estimate for the relative bias of the early vs. late galaxies by determining the best fitting constant function in the range $[k_{min}, k_{max}]$, where $k_{min} = 0.04$ Mpc^{-1} h and k_{max} varies between 0.15 and 0.5 Mpc^{-1} h, i.e. well into the non-linear regime. The results, listed in Table 4.7, are quite robust to the choice of k_{max} out to 0.4 Mpc^{-1} h, as anticipated by the visual inspection of fig. 4.23. The value of the relative bias shows that early type galaxies are more clustered than late type objects, as expected.

Westover (2007) also performs a similar analysis but provides the values of the bias for early and late types in different luminosity bins. Averaging over the published values, the relative bias $b_e/b_l \simeq 1.32$ with a 1-σ error of about 15%. This value matches our estimate at $k_{max} = 0.5$ Mpc^{-1} h, which is not surprising since Westover (2007) estimates the bias ratio at 2.4 h^{-1} Mpc, i.e. in the nonlinear regime.

Figure 4.23: Power spectrum ratio vs. the wavenumber k: $\sqrt{PSe/PSl}$. Error bars are given by the scatter among the spectral ratios among the mocks.

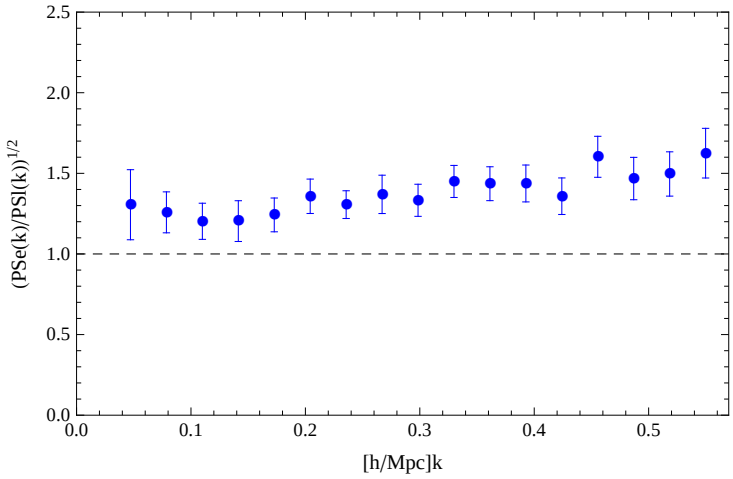


Table 4.7: Relative bias early−late type galaxies obtained fitting $(PSe(k)/PSl(k))^{1/2}$ up to different k_{max}. 1σ errors are from χ^2 minimization.

$k_{max}[h$ Mpc$^{-1}]$	$b \pm \sigma$
0.15	1.228 ± 0.067
0.20	1.232 ± 0.056
0.30	1.289 ± 0.037
0.50	1.353± 0.029

4.6 Estimating cosmological parameters

In this Section we fit a model power spectrum to that of all 2MRS galaxies shown in fig.4.11. In practice we repeat the analysis performed on the halo catalogs of the BASICC simulation, in Section 3.3 and on the realistic 2MRS catalogs earlier on in this Chapter. We shall fit different model power spectra to the measured 2MRS spectrum and estimate the best fitting parameters and their uncertainties by minimising the χ^2 differences between the two spectra. As anticipated, we assume uncorrelated power in different k-bins and use the results of the analysis presented in the first part of this Chapter to select k-ranges in which we compare the measured spectrum with the various models. The upper limits of the interval $[k_{min}, k_{max}]$ are listed in Table 4.9 together with the associated model. In all cases the value of k_{min} was set equal to 0.03 h Mpc^{-1}.

The models are those described is Section 4.3.3.

All of them are based on the ΛCDM model and depend on a number of common cosmological parameters that characterise the properties of the background cosmology and the properties of the primordial density fluctuations. Some of them are kept fixed at the values recently estimated by CMB analysis performed by the Planck collaboration (2013). They are listed in Table 4.8 and include the Hubble constant, H_0, the baryon density Ω_b, the spectral normalisation σ_8 and the primordial index n_s. The rationale behind keeping these parameters at their estimated values, ignoring their errors, is twofold. First, as we shall see, the uncertainty in the estimate of the free parameters from our 2MRS analysis is rather large, and these errors will not significant affect the error budget. The second reason is that some of the parameters in the Table, namely σ_8, are completely degenerate with some of the free parameters (the galaxy bias or the distortion parameter β). The problem could be avoided by considering combination of cosmological parameters like the product σ_8 times the growth rate of density fluctuations that carries a similar information content (Percival & White, 2009) but we prefer here to take a more traditional rule of considering the individual parameters and fix some of them at pre-determined values.

Table 4.8: Values of the fixed parameters. The values are the best fit values for the ΛCDM model from Planck collaboration, 2013.

Parameter	Value
$\Omega_b h^2$	0.022068
H_0	67.11
σ_8	0.8344
n_s	0.9624

In all cases considered the free parameters are the galaxy bias, b, and the mass density parameter, Ω_M, and our goal is to estimate their value by means of χ^2 minimisation. Since the magnitude of the Kaiser boost depends on their value, they both determine the amplitude of the power spectrum but only Ω_M affects its shape. Therefore, to estimate both of them we need to fit the measured spectrum over a sufficiently large k- range and the goodness of the result will be a compromise between the need to have a large leverage on the k-axis and that of reducing systematic errors induced by nonlinear effects on small scales. Some of the models explored require some additional parameter (the Q-value that quantifies mild nonliearities or the pairwise velocity dispersion σ in the small scale damping term). In these cases the additional parameters are free to vary but then we shall marginalise over it to focus on b and Ω_M only. To summarise, the parameters b, Ω_M and, if present, the additional parameter X where obtained by minimising the following quantity:

$$\chi^2(\Omega_M, b, X) = \frac{\sum_{i=1}^{N}(P_m(k_i) - P_{conv}(k_i, \Omega_M, b, X))^2}{\sigma^2(k_i)} \tag{4.8}$$

where $P_m(k_i)$ is the power spectrum estimated from the 2MRS sample, $P_{conv}(k_i)$ is the model power spectrum convolved with the window function, k_i is the modulus of the wavenumber in the i-th bin, $\sigma^2(k_i)$ is the variance from the *rms* dispersion among the mocks and the sum runs over all k-bins between k_{min} and k_{max}.

Let us now describe the outcome of the best fitting procedure for the different models considered.

- MODEL LW: $P_{conv}(k) = b^2 K(\Omega_M, b) P_{CAMB}(k) * WF$.

  In this expression $K(\Omega_M, b)$ is the Kaiser boost factor given by $(1 + \frac{2}{3}\beta + \frac{1}{5}\beta^2)$ with $\beta = \Omega_M^{0.55}/b$, $P_{CAMB}(k)$ is the ΛCDM linear matter power spectrum obtained from $CAMB$ and $*WF$ indicates the convolution with the window function computed from the mocks. Since this model relies on linear theory, we limit the fit to large scales $k \leq 0.15$ Mpc^{-1} h. The results are shown in fig.4.24 where we plotted the confidence contour levels in the $b - \Omega_M$ plane. The contours filled with different shades of blue represent 1, 2, 3 σ contours. The best fit (b, Ω_M) values together with their 1-σ uncertainty obtained after marginalising over the other parameters are listed in Table 4.9. The best fit values appear to be in the right ballpark. The accuracy of the matter density parameter is rather limited, as anticipated. Ω_M is estimated with a 1-σ accuracy of about 20% and the *absolute* bias with an accuracy of about 3%. The accuracy of the latter result could depend on the fact that we are fixing the normalization parameter σ_8, which in Planck is determined with an accuracy of 3% but not in agreement with estimates from clustering.

- $NLW3$: $P_{conv}(k) = [P_{Hf}(k)b^2K(\Omega_M, b)] * WF$

  This case is analogous to the previous one except for the fact that the model matter power spectrum is nonlinear, obtained using $Halofit$, dubbed $P_{Hf}(k)$. The goal here is to assess whether nonlinear effects exist at $k \leq 0.15$ h Mpc^{-1} that may have affected the previous result. Therefore we shall fit the $NLW3$ to the data in the same k-range used in the LW, i.e. out to $k = 0.15$ h Mpc^{-1}. The results are shown as black contours in fig.4.24, superimposed to the coloured contours obtained from models LW. The agreement between the two results is remarkable. The new contours are slightly shifted to lower values of Ω_M and b. The shift may indeed indicate that non-linear effects affect the analysis even on these scales. On the other hand, the mismatch may reflect the fact that this model couples nonlinear dynamics, responsible for the $Halofit$ spectrum, with linear theory, responsible for the Kaiser boost, in an inconsistent fashion since it ignores the role of nonlinear, incoherent velocities. The net effect is to predict too large power which, in the fit, is compensated by decreasing the values of b and Ω_M. Irrespective of its possible origin, the significance of the shift is well within 1-σ and the result of this test constrain the magnitude of nonlinear effects at $k \leq 0.15$ h Mpc^{-1} and their impact on the outcome of the fitting procedure.

- NLW3$_G$: $P_{conv}(k) = [P_{Hf}b^2K(\Omega_M, b)G(k, \sigma)] * WF$

  To push the analysis to higher k values one must then include the effect of non-linear motions, which we model by adding a Gaussian damping term $G(k, \sigma) = \frac{\pi^{1/2}}{2}\frac{erf(\sigma k)}{\sigma k}$ to the $NLW3$ model. This additional term is characterised by a single parameter, σ, that is treated as a free parameter together with b and Ω_M. The tests performed with the mocks have shown that this model provides a good fit to the measured spectrum out to $k = 0.4$ h Mpc^{-1} , a value that we used for k_{max}. To check the sensitivity of the results on this choice we also push the analysis out to $k_{max} = 0.5$ h Mpc^{-1}, The results are shown in fig. 4.25. in the form of coloured contours (for $k_{max} = 0.4$ Mpc^{-1} h) and black contours (for $k_{max} = 0.5$ h Mpc^{-1}) in the $b - \Omega_M$ plane. To minimise the degeneracy among the parameters (that, as we have verified, is very weak when $k_{max} > 0.3$ h Mpc^{-1}) we put a flat prior [0,5] Mpc h^{-1} on σ. The contours in the plot were obtained after marginalising over this additional parameter.

  The two sets of contours are in very good agreement with each other, showing that the results are quite robust to the choice of the nonlinear scale, and with those obtained from the LW model. As shown in Table 4.9 including a damping term has improved the accuracy in the estimate of Ω_M and b to ~ 15 % and ~ 2.5 %, respectively.

- $NLW1$ and $NLW2$: $P_{conv}(k) = [b^2 P_{CAMB} K(\Omega_M, b) F(k, Q)] * WF$

  In both models, non-linear effects are modelled with a k-dependent function $F(k, Q)$ which in the $NLW1$ has the form $(1 + Qk^2)$ and in the case $NLW2$ has the form $(1 + Qk^{3/2})$. The tests performed with the mocks have shown with these models that the accuracy in the estimates of b and Ω_M is similar to that obtained with model NLW3$_G$ but the value of k_{max} is smaller. Therefore we shall restrict the fit to $k \leq 0.4\ h$ Mpc^{-1}. Even in this case we rest Q as a free parameter with a flat prior of [0.5]. The significance contours in fig. 4.26 refer to the $NLW1$ model (coloured contours) and to the $NLW2$ model (black curve contours) and where we marginalised over the additional nonlinear parameter Q.

  The results are in agreement, and sometimes in very good agreement, with those obtained from some of the other models and altogether demonstrate the robustness of these results.

Table 4.9: Best fit results for the cosmological parameters and 1σ marginalized errors.

Model	$k_{max}[h$ Mpc$^{-1}]$	$\Omega_M \pm \sigma$	$b \pm \sigma$
LW	0.15	0.320±0.063	1.228±0.038
$NLW3$	0.15	0.289±0.061	1.185±0.040
$NLW3_G$	0.40	0.334±0.053	1.253±0.032
$NLW3_G$	0.50	0.354±0.048	1.265±0.029
$NLW1$	0.40	0.350±0.040	1.240±0.027
$NLW2$	0.40	0.333±0.045	1.218±0.034

Figure 4.24: $1-2-3\sigma$ confidence levels on the plane $b-\Omega_M$. Fit of the 2MRS power spectrum up to $k_{max} = 0.15$ h Mpc^{-1}. The filled contours correspond to model $LW : b^2K(\Omega_M, b)P_{CAMB}(k) * WF$, where b is the linear bias parameter, $K(\Omega_M, b)$ is the "Kaiser boost" factor, $P_{CAMB}(k)$ is the linear matter power spectrum from CAMB, WF is the window function of the survey. The black solid contours are obtained fitting data with model $NLW3 : [P_{Hf}(k)b^2K(\Omega_M, b)] * WF$, where $P_{Hf}(k)$ is the non linear matter power spectrum from *Halofit*.

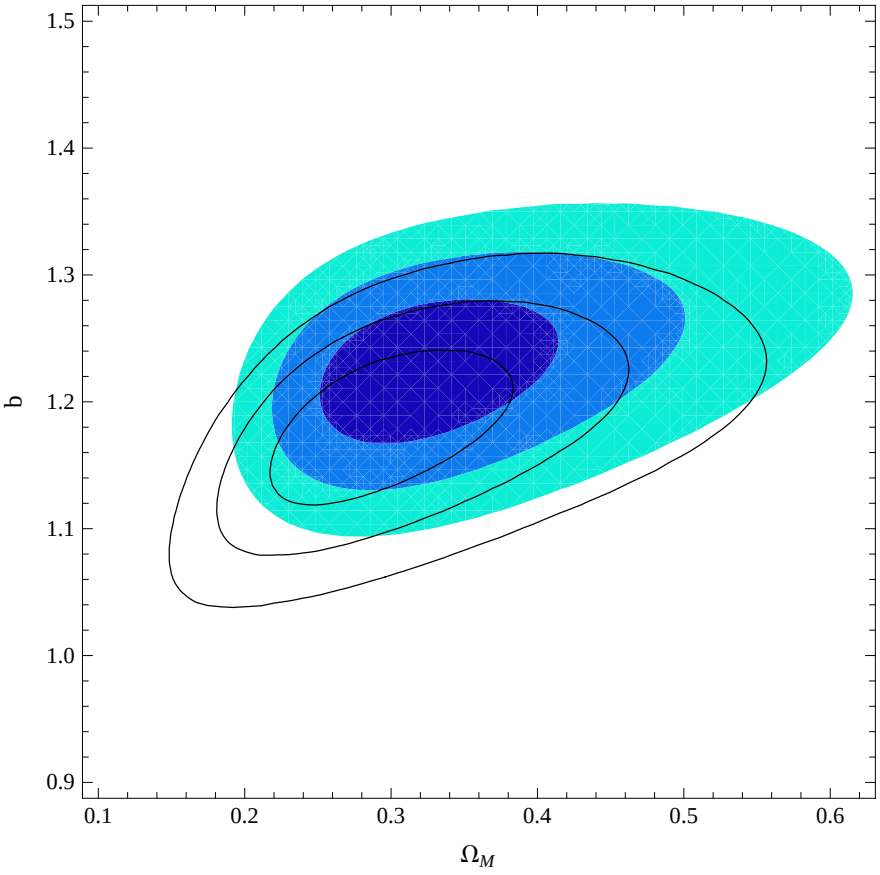


Figure 4.25: Marginalized constraints on the plane $b - \Omega_M$. Fit of the 2MRS power spectrum with *Halofit* plus Gaussian damping (model NLW3$_G$): $[P_{Hf} b^2 K(\Omega_M, b) G(k, \sigma)] * WF$, where $P_{Hf}(k)$ is the non linear matter power spectrum from *Halofit*, b is the linear bias parameter,$K(\Omega_M, b)$ is the "Kaiser boost" factor, $G(k, \sigma) = \frac{\pi^{1/2}}{2} \frac{erf(\sigma k)}{\sigma k}$ is the Gaussian damping term accounting for non linear motions, WF is the window function of the survey. The filled contours are obtained fitting data up to $k_{max} = 0.4\ h$ Mpc^{-1}. The black solid contours correspond to $k_{max} = 0.5\ h$ Mpc^{-1}.

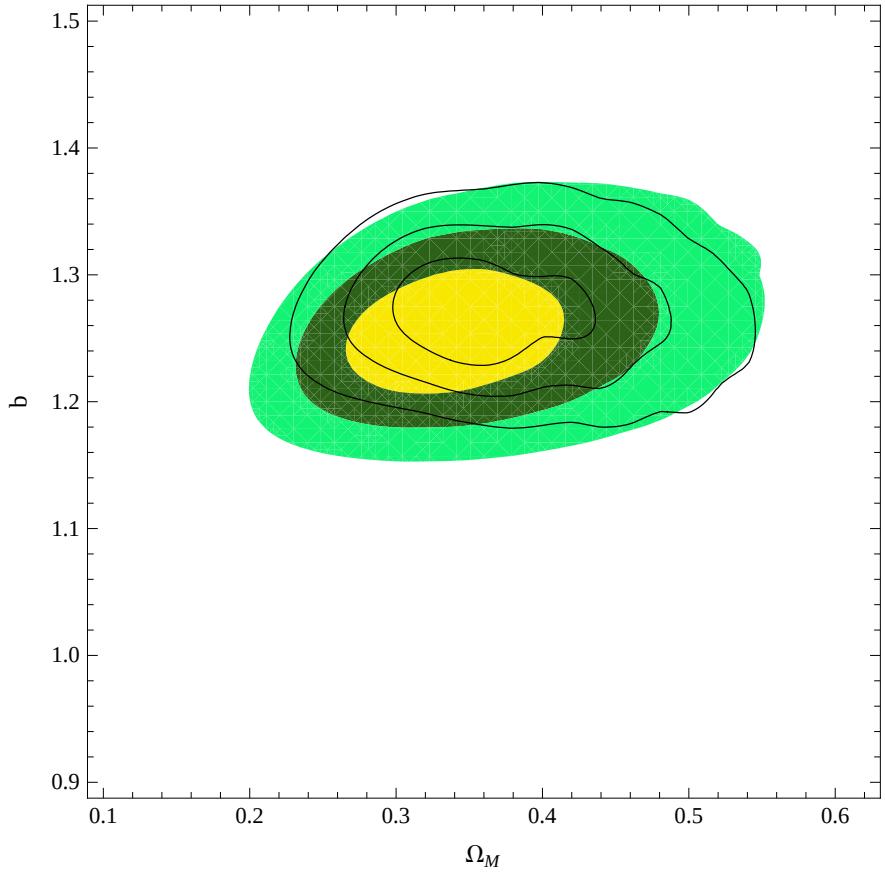
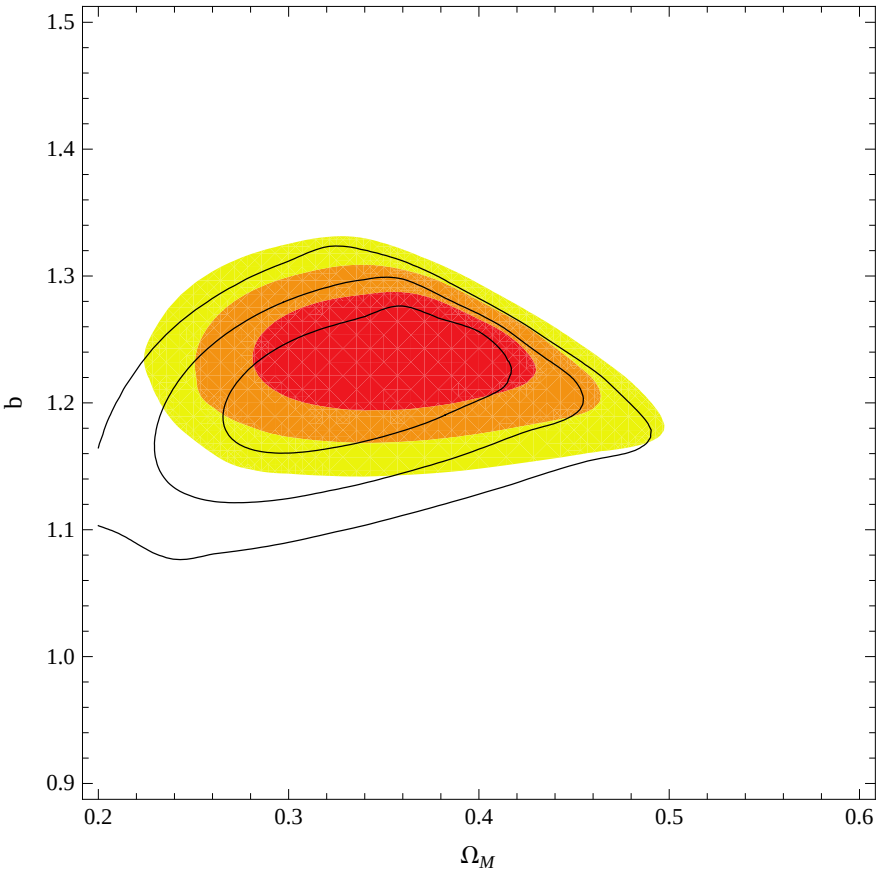


Figure 4.26: Marginalized constraints on the plane $b - \Omega_M$. Fit of the 2MRS power spectrum with $NLW1 : [b^2 P_{CAMB} K(\Omega_M, b)](1+Qk^2) * WF$ (filled contours) and $NLW2 : [b^2 P_{CAMB} K(\Omega_M, b)](1+Qk^{3/2})$ (black solid contours), where b is the linear bias parameter, $P_{CAMB}(k)$ is the linear matter power spectrum from CAMB, $K(\Omega_M, b)$ is the "Kaiser boost" factor, WF is the window function of the survey. Both contours are obtained fitting data up to $k_{max} = 0.4\ h$ Mpc^{-1}.



Chapter 5

Estimating scale-dependent galaxy bias

In the previous analyses we made the simplifying hypothesis that mass and galaxy overdensity field can be simply related through a linear bias relation. In fact, we made, except for some cases, the stronger hypothesis that linear theory holds, on the scales of the analysis, for both galaxy bias and for the density fluctuations. Disentangling the two effects is doable in principle since the type of nonlinearity, its time evolution and the possible associated scale-dependence are different in the two cases. Dynamical models for the evolution of density fluctuations are indeed available, either based on analytic approximation or on numerical N-body approach. In practice, however, these models have limitations. Their validity is usually limited to a restricted range of scales and depend on the assumed cosmological model. On the other hand, the state-of-the-art of theoretical models of galaxy evolution do not allow to predict galaxy bias with good accuracy. For all these reasons, a phenomenological approach in which deviations from the linear biasing prescriptions are measured directly from the data should be regarded as a valid alternative that we shall explore in this Chapter. The content of this Chapter is based on L. Amendola et al., *Constraints on a scale-dependent bias from galaxy clustering*, in preparation.

Di Porto et al. (2012), among others, have shown that future galaxy redshift surveys do contain enough information to break the degeneracy between the galaxy bias, clustering amplitude and the growth factor, effectively allowing to estimate galaxy bias from the data itself. Of course this comes at the price of making some assumptions on the biasing relation. We assume that the bias relation between the galaxy density field and the mass field is local, deterministic but not necessarily scale-independent. The locality and deterministic hypotheses are likely to break down on those scales in which processes involved in galaxy formation and evolution, like stellar feedback, are effective. These scales are typically small compared to the

typical size of current and future galaxy surveys and can be safely ignored by smoothing small scale fluctuations or by restricting the analysis of galaxy clustering on large scales.

From an observational viewpoint, the scale dependent bias has been advocated at low and at high redshifts to account for 2-point galaxy clustering statistics, higher order statistics, galaxy counts (e.g. Norberg et al. 2001, Verde et al. 2002, Branchini 2001, di Porto et al. 2014)and gravitational lensing (Hoekstra et al., 2002). In this Chapter we shall then assume that galaxy bias is scale dependent and we investigate the possibility of estimating it using next generation galaxy surveys. In particular we shall focus on a type of survey which has characteristics similar to that of the spectroscopic galaxy survey that will be carried out with the satellite *Euclid*, that will be launched by *ESA* in 2020 (Laureijs et al., 2011).

This survey is expected to target about 30 million galaxies whose redshift will be determined thanks to their H_α emission line over an area of 15,000 deg^2 in a redshift range that should be approximately equal to $z = [1.0, 2.0]$. We stress that the aim here is not to mimic the exact properties of the *Euclid* survey but to assess the possibility that a survey of this type could detect a scale dependency in galaxy bias.

Our primary theoretical tool here will be the Fisher Matrix analysis and we shall provide some details on its practical implementation in the next Sections. In addition, and with the aim of providing a realistic setup, we use mock *Euclid* catalogs to calibrate the reference scale-dependent bias model that we use in the analysis. The results are presented and discussed in the last two Sections.

In the Fisher matrix analysis we set a fiducial cosmological model. We choose a model with the following parameters: a Hubble parameter $h = 0.7$, mass density parameter $\Omega_M = 0.25$, baryon density $\Omega_b = 0.0445$, curvature parameter $\Omega_k = 0$, the primordial index $n_s = 1$, dark energy equation of state parameters $w_0 = -0.95$ and $w_1 = 0$ (see next Section for details), *rms* density fluctuation at 8 h^{-1} Mpc $\sigma_8 = 0.8$, and growth rate $f(\Omega_M) = \Omega_M(z)^\gamma$ with $\gamma = 0.545$.

In order to constrain the parameters, we use the Fisher matrix method (Fisher, R. A., 1935), (Tegmark, M. et al., 1997), that, following (Komatsu E., et al., 2011), we apply to the power spectrum analysis in redshift space.

5.1 Modelling the Redshift Space Power Spectrum

The analytic model of the power spectrum in redshift space is modelled as follows.

- We use $CMBFAST$ (Seljak, U., Zaldarriaga, M., 1996) to obtain the linear power spectrum of the matter in real space at the present epoch, $P_{0r}(k)$, where the subscript $_r$ characterises the reference model. This

power spectrum depends on most of the cosmological parameters listed above, namely: h, Ω_M, Ω_b, Ω_{DE}, Ω_k and σ_8.

- To convert the reference power spectrum model into a different cosmology we use the multiplicative term

$$R(z) = \frac{D_r^2(z)H(z)}{D^2(z)H_r(z)} \,, \tag{5.1}$$

where $D(z)$ is the angular diameter distance and $H(z)$ is the expansion history, i.e. the Hubble parameter as a function of epoch:

$$\begin{aligned} H(z)^2 &= H_0^2 \left[\Omega_M(1+z)^3 + \Omega_k(1+z)^2 + (1-\Omega_M-\Omega_k)\right] \\ &\quad \exp\left\{3\int\left(1+w_0+w_1\frac{z}{1+z}\right)\frac{dz}{1+z}\right\} \,, \end{aligned}$$

in which we allow for a more general dark energy model characterised by an equation of state $w(z) = w_0 + w_1\frac{z}{1+z}$.

- We model the linear redshift distortions using the Kaiser boost in which, unlike in the previous Chapters, we explicitly indicate the angular dependence instead of performing angular averages:

$$K(\Omega_m(z), b, \mu) = (1+\beta(z)\mu^2)^2 \,, \tag{5.2}$$

where $\beta(z) = \Omega_M(z)^\gamma/b$, μ is the cosine angle between the wavenumber vector $\vec{k}$ and the line of sight and b is the bias parameter.

- To account for the linear evolution of density fluctuations we further multiply the matter power spectrum at the present epoch by $D_1(z)$, the growth factor at the redshift z, properly normalized.

- To model nonlinear effects on large scales, where we shall focus our Fisher analysis, we follow Eisenstein et al. (2007) and Seo et al. (2007) and multiply $P_{0r}(k)$ by the factor

$$S(k,\mu) = \exp\left\{-k^2\left[\frac{(1-\mu^2)\Sigma_\perp^2}{2} + \frac{\mu^2\Sigma_\parallel^2}{2}\right]\right\} \tag{5.3}$$

where $\Sigma_\perp$ and $\Sigma_\parallel$ represent the Lagrangian displacement across and along the line of sight, respectively. They are related to the growth factor D_1 and to the growth rate $f(\Omega_M) = \Omega_M(z)^\gamma$ through $\Sigma_\perp = \Sigma_0 D_1$ and $\Sigma_\parallel = \Sigma_0 D_1(1+f)$. The value of Σ_0 is proportional to σ_8. For our reference cosmology, where $\sigma_8 = 0.8$, we have $\Sigma_0 = 11\,h^{-1}$Mpc.

- An additional issue, is the error in the measured redshift, δz, which propagates into an uncertainty in the estimated distances, $\sigma_r = \frac{\delta z}{H(z)}$.

Under the assumption that these errors do not correlate with the local galaxy density we can quantify its impact in the measured spectrum by an additional multiplicative term

$$W(k, \sigma_r, \mu) = e^{-k^2\mu^2\sigma_r^2} \,. \tag{5.4}$$

- We also include an additive, scale independent term to account for uncertainties in the shot noise removal, $P_s(z)$.
- Finally, to obtain a model for the galaxy power spectrum, we multiply that of the matter by a scale dependent bias $b(k, z)$ that we model in the next Section.

To summarise, the model power spectrum that we consider in this Chapter has the form as

$$P_{\rm g}(z, k, \mu) = P_{0r}(k)R(z)D_1^2(z)b(k, z)^2K(\Omega_m(z), b, \mu)W(k, \sigma_r, \mu)S(k, \mu)+P_{\rm s}(z) \,. \tag{5.5}$$

Observations will allow to estimate the power spectrum in finite redshift bins whose size should compromise between the need of large statistics and volume and the need to minimise evolution effects. To mimic the characteristics of a *Euclid* like survey we shall use bins of width $\Delta z = 0.2$ starting at $z = 0.6$ and with a number of objects similar to that expected for the real survey. The relevant numbers are provided in Table 5.1.

Table 5.1: Redshift bins used in our analysis. Col 1: central redshift of a z-bin of width $\Delta_z = 0.2$. Col 2: expected mean number density of objects within each bin (in Mpc^{-3} h^3) matching those expected in the Euclid survey according to Laureijs R., et al.(2011).

z	$\bar{n}$ [Mpc^{-3} h^3]
0.6	3.56×10^{-3}
0.8	2.42×10^{-3}
1.0	1.81×10^{-3}
1.2	1.44×10^{-3}
1.4	0.99×10^{-3}
1.6	0.55×10^{-3}
1.8	0.29×10^{-3}
2.0	0.15×10^{-3}

5.2 Modelling the scale dependent bias

Several authors have proposed phenomenological or theoretical models for the scale dependent bias $b(z, k)$ (see Section 2.4). Since here we are focusing

on the accuracy with which one can constrain galaxy bias rather than on the accuracy of the bias model, we shall consider the two rather simple analytic prescriptions given in eq. 2.59 and eq. 2.63, that we write here including dependence on redshift in the more general form:

$$\mathrm{P-model:} \quad b(z,k) = b_0(z) + b_1(z)\left(\frac{k}{k_1}\right)^n \tag{5.6}$$

$$\mathrm{Q-Model:} \quad b(z,k) = b_0(z)\left[\frac{1+Q(z)(k/k_1)^2}{1+A(z)(k/k_1)}\right]^{1/2}. \tag{5.7}$$

Furthermore here we use the subscript 0 to indicate the $k-$independent bias parameter.

Although other bias models have been proposed with stronger theoretical motivations, see Section 2.4, we have decided to adopt these two for two reasons. First, both models, especially the second one, have been used in the literature. This way our results can be directly compared with those of previous works, helping us to limit the allowed parameter space to be explored. Second, in this Chapter we focus on large scales encompassing the BAO scales (~ 100 Mpc h^{-1}) and consider a limited range of wavenumbers where the differences among the various models proposed in the literature are small. We then consider those that have a simple analytic form and are easier to handle within the Fisher Matrix formalism. Finally, we notice that neither model allows for non-Poisson shot noise term. This, however, could be easily accounted for by modifying the last term in eq. 5.5.

In our analysis we shall consider two types of reference models for both the P-model and the Q-model cases. The first is the unbiased reference model which corresponds to the case in which galaxies are unbiased tracers of the underlying mass density field. The bias parameters corresponding to this case are $b_0 = 1\,, b_1 = 0$ for the P-model and $b_0 = 1\,, A = 0\,, Q = 0$ for the Q-model. The second is a more realistic bias model calibrated using a mock *Euclid* catalog of galaxies, as detailed below. The parameters of the corresponding reference models for the P-model and Q-model, specified bin by bin are listed in Table 5.2.

5.2.1 Modelling galaxy bias from *Euclid* mock catalogs

Th *Euclid* spectroscopic redshift survey will target Hα emission-line galaxies and is expected to provide a highly complete redshift catalog of objects with a line flux H$\alpha \geq 3 \cdot 10^{-16}$ erg cm^{-2} s^{-1} over a large range of redshift. These objects are not expected to trace the mass field in an unbiased way. Current samples of Hα galaxies at the relevant redshifts are too small to constrain the bias of these objects. Therefore, in an attempt to obtain a realistic reference model we rely on numerical and semi-analytic techniques.

More specifically, we consider the so-called "100 deg^2" mock galaxy catalogs described in Chapter 3, Section 3.1. We have extracted a subsample

Table 5.2: Table of reference values for the P-model and Q-model parameters, for the realistic bias case. For the P-model (parameters b_0 and b_1) we set $k_1 = 1$ and list the values for three different choice of the power law index $n = 1\,, 1.28$ and 2. For the Q-model we set $k_1 = 1$ and A=1.7.

	P-Model						Q-Model	
	$n = 1$		$n = 1.28$		$n = 2$			
z	b_0	b_1	b_0	b_1	b_0	b_1	b_0	Q
0.6	1.052	0.669	1.060	0.707	1.159	0.668	1.239	4.70
0.8	1.038	0.673	1.088	0.663	1.169	0.682	1.257	4.54
1.0	1.133	0.747	1.191	0.750	1.277	0.786	1.359	4.92
1.2	1.217	0.988	1.298	0.971	1.410	1.017	1.494	5.50
1.4	1.355	1.093	1.441	1.063	1.550	1.123	1.635	5.70
1.6	1.487	1.220	1.583	1.192	1.706	1.251	1.752	6.62
1.8	1.614	1.401	1.719	1.399	1.879	1.436	1.920	6.99
2.0	1.754	1.439	1.853	1.495	2.011	1.650	2.084	6.47

consisting of all objects in the range $z = [0.6, 2.0]$ and Hα line flux larger than H$\alpha_{\rm min}$($= 3 \cdot 10^{-16}$ erg cm^{-2} s^{-1}).

To model galaxy bias we measure the power spectrum of the mock Hα galaxies in all redshift bins and compare the results with the linear predictions obtained from CAMB. In principle, we could use this as a reference bias model in the form of a stepwise function of k and z. However, this bias estimate is noisy both on small scales, due to the finite number of mock galaxies, and to large scales due to cosmic variance. Moreover, the results would be in the form of a 2D, z-k table which is cumbersome to implement in the Fisher Matrix procedure. For this reason we decide to regularise the bias estimate by fitting both the P-model and the Q-model to the measured bias.

The procedure to estimate the galaxy bias is as follows:

- We consider eight, partially overlapping cubic boxes of increasing redshifts and align them along the line of sight, centring each of them at the redshifts indicated in Table 5.1. Each of them is the largest box fully contained in the lightcone at the given redshift. As a consequence, the box size increases with the redshift.

- We select mock galaxies brighter than H$\alpha_{\rm min}$ from the simulation and apply appropriate statistical weights given by the inverse of the selection function, as described in Section 2.1.3, to account for the flux limit selection. The radial dependence of these weights turned out to be very mild within each box.

- We measure the power spectrum of the mock galaxies in real space within each cube. We limit our analysis between the minimum

wavenumber that we can reliably sample and that depends on the redshift and $k_{\rm MAX} = 1h\,{\rm Mpc}^{-1}$. This choice allows to reduce the impact of shot noise and nonlinear effects.

- We estimate the scale dependent galaxy bias as $b(k,z) = \sqrt{\frac{P_{\rm m}(k,z)}{P_{CAMB}(k,z)}}$, where $P_{\rm m}(k,z)$ is the measured power spectrum of the mock galaxies at redshift z and $P_{CAMB}(k,z)$ is the linear power spectrum of the mass obtained from CAMB using the same cosmological parameters as the parent N-body simulation.

- We fit the P-model and the Q-model to the estimated $b(k,z)$ and find the best fit parameters by minimising the χ^2 of the residuals. In the procedure we use errors obtained from the FKP estimate of the power spectrum, eq. 2.47. In the process, to improve the robustness of the minimisation we fix some of the free parameters. For the P-model we set $n = 1$, 1.28 and 2 ($n = 1.28$ is the best fit value of a three-parameters minimization) while for the Q-Model we set $A = 1.7$, following Cole et al. 2005. The minimisation is first carried out over a conservative range of wave numbers, $[k_{min}, k_{max}] = [0.03, 0.3]$. We then gradually expand the range to both large and small scales and stop when $\chi^2/{\rm d.o.f.} \sim 1$.

The best fit parameters are listed in Table 5.2.

5.3 Fisher Matrix Analysis

The observed power spectrum in a given redshift bin depends therefore on a number of parameters, denoted collectively as p_i. To estimate the Fisher matrix elements, given the model power spectrum, we calculate numerically or, whenever possible, analytically, the derivatives of the spectrum with respect to the free parameters p_i:

$$\left(\frac{\partial \ln P_{\rm g}}{\partial p_i}\right)_{\rm r}\,, \tag{5.8}$$

evaluated at the parameter values of the fiducial model. The 1-σ error for each parameter, p_i, is $\sigma_{p_i} = \sqrt{(F^{-1})_{ii}}$, where F^{-1} is the inverse Fisher matrix (for more details on Fisher matrix method see Kitching and Amara, 2009).

The elements of the Fisher matrix are evaluated at the parameter values of the fiducial model in each of the redshift bins through Tegmark et al. (1997):

$$F_{ij} = \frac{1}{8\pi^2}\int_{-1}^{+1} d\mu \int_{k_{\rm min}}^{k_{\rm max}} dk\; k^2 \left(\frac{\partial \ln P_{\rm g}}{\partial p_i}\frac{\partial \ln P_{\rm g}}{\partial p_j}\right)_{\rm r} V_{\rm eff}(k,\mu)\,, \tag{5.9}$$

where

$$V_{\rm eff}(k,\mu) = \left[\frac{\bar{n}P(k,\mu)}{\bar{n}P(k,\mu)+1}\right]^2 V_{\rm survey}\,, \tag{5.10}$$

is the effective volume of the survey sampled at the scale k along the direction μ. $V_{\rm survey}$ and $\bar{n}$ represents the volume of the survey and the mean number density of galaxies in each redshift bin.

Focusing on the free parameters of the biasing function we have for the P-model:

$$\frac{d\ln P}{db_0}|_r = \frac{2}{b_{0,r}} - \frac{2f\mu^2}{b_{0,r}(b_{0,r}+f\mu^2)} \tag{5.11}$$

$$\frac{d\ln P}{db_1}|_r = \frac{2}{b_{0,r}}(k/k_1)^n - \frac{2f\mu^2(k/k_1)^n}{b_{0,r}(b_{0,r}+f\mu^2)}\,, \tag{5.12}$$

and for the Q-model :

$$\frac{d\ln P}{db_0}|_r = \frac{2}{b_{0,r}} - \frac{2f\mu^2}{b_{0,r}\left[T(k,\mu)+f\mu^2\right]} \tag{5.13}$$

$$\frac{d\ln P}{dQ}|_r = \frac{(k/k_1)^2}{1+Q_r(k/k_1)^2} - \frac{f\mu^2(k/k_1)^2}{1+Q_r(k/k_1)^2}\frac{1}{f\mu^2+T(k,\mu)} \tag{5.14}$$

$$\frac{d\ln P}{dA}|_r = -\frac{k/k_1}{1+A_r(k/k_1)} + \frac{f\mu^2(k/k_1)b_{0,r}}{\left[1+A_r(k/k_1)\right]\left[f\mu^2+T(k,\mu)\right]} \tag{5.15}$$

where $T(k,\mu) = b_{0,r}\left[\frac{1+Q_r(k/k_1)^2}{1+A_r(k/k_1)}\right]^{1/2}$

The subscript $_r$ refers to the fiducial reference model. In addition to these, we also compute the uncertainties for the following cosmological parameters: $h\,,\Omega_M\,,\Omega_b\,,n_s\,,\sigma_8$ and γ. Their reference values were listed at the beginning of this Chapter.

5.4 Unbiased reference model: results

Let us consider first the reference case of a population of galaxies that are unbiased tracers of the mass distribution. And, in this framework, let us consider both the P- and the Q-models. For the P-model we explored two different reference cases: one characterised by a power-law index $n = 1$, corresponding to a bias with a mild dependence on the scale and one with a stronger scale dependence characterised by $n = 2$. The 1-σ uncertainties in the cosmological parameters obtained by considering all available redshift bins and after marginalising over all other parameters are listed in Table 5.3.

These results confirm the fact that, as already pointed out by di Porto et al. (2012), *Euclid* will constrain the cosmological parameters with % level accuracy. Those results, reported in Table 5.4, however, assumed that bias could depend on the redshift but not on the scale. The interesting result

Table 5.3: 1σ errors on cosmological parameters for the P- and Q- models for the reference case of an unbiased population of mass tracers.

	P-model		Q-model
	$n = 1$	$n = 2$	
σ_h	0.038	0.037	0.039
σ_{Ω_M}	0.016	0.015	0.016
σ_{Ω_b}	0.0036	0.0034	0.0036
σ_{n_s}	0.042	0.036	0.044
σ_γ	0.025	0.028	0.029
σ_{σ_8}	0.0044	0.0045	0.0047

Table 5.4: 1σ errors on cosmological parameters for scale independent biased population of mass tracers, from di Porto et al. 2012. The fiducial reference model is quite similar to ours: $h = 0.703$, $\Omega_M = 0.271$, $\Omega_b = 0.045$, $\Omega_k = 0$, $n_s = 0.966$, equation of state parameters for the dark energy $w_0 = -0.95$ and $w_1 = 0$, $\sigma_8 = 0.809$, and growth rate $f(\Omega_M) = \Omega_M(z)^\gamma$ with $\gamma = 0.545$, redshift dependent bias from Orsi et al., 2010.

	di Porto 2012 model
σ_h	0.008
σ_{Ω_M}	0.004
σ_{Ω_b}	0.0008
σ_{n_s}	0.02
σ_γ	0.03
σ_{σ_8}	0.007

here, is that a similar accuracy is also achieved when one allows for a scale dependent bias.

The expected errors on the bias parameters in each redshift bin are listed in Table 5.5. Errors for b_1 are larger than those for b_0 and increase with the index n. This is not surprising since the leading term of the expansion of the mass density contrast, b_0, is expected to be more accurately determined than the first term of the expansion, b_1, $n = 1$, or the second, b_1, $n = 2$. The constant bias term b_0 is expected to be determined with very high accuracy (i.e. at the % level) similar to that of the scale-dependent term b_1 as long as the scale dependence is close to linear. The accuracy degrades when the scale dependence becomes stronger.

In the Q-model, absolute errors on A and Q appear to be large. However, what really matter is the relative error that we can gauge by comparing the values in Table 5.5 with the corresponding bias parameters values for the reference case in Table 5.2. With this respect, the accuracy in the estimate of the constant bias term b_0 is similar to that of the P-model, as expected. Errors on the nonlinear bias parameters are much larger. This is not entirely surprising since weaker constraints result from having one more parameter to marginalize over. Uncertainties on A, corresponding to the first order of the expansion, are of the order of 15% and the ones on the second term Q are about 50% of the signal.

Errors increase with the redshift. Again, this is not surprising as the effective volume of the survey associated to the generic redshift shell decreases when moving to high redshifts.

Table 5.5: $1-\sigma$ errors on the bias parameters for the P-model with $n = 1, = 2$ and the Q-model, for the reference case of an unbiased population of mass tracers. We have set $b_0 = 1$, $b_1 = 0$, $A = 0$, $Q = 0$ and $k_1 = 1.0$.

	P-model				Q-model		
	n=1		n=2				
z	σ_{b_0}	σ_{b_1}	σ_{b_0}	σ_{b_1}	σ_{b_0}	σ_Q	σ_A
0.6	0.013	0.14	0.0081	1.2	0.017	3.0	0.35
0.8	0.013	0.13	0.0093	0.97	0.017	2.5	0.32
1.0	0.013	0.12	0.011	0.86	0.017	2.2	0.31
1.2	0.014	0.12	0.012	0.82	0.018	2.2	0.31
1.4	0.014	0.13	0.013	0.91	0.020	2.5	0.34
1.6	0.016	0.16	0.014	1.2	0.023	3.4	0.42
1.8	0.019	0.22	0.016	1.9	0.027	5.4	0.60
2.0	0.026	0.34	0.019	3.3	0.037	9.4	0.97

To investigate the covariance among parameters we show their 68% contours. Figure 5.1 refers to the b_0 and b_1 parameters in the P-model case with $n = 1$. The different ellipses with increasing areas refer to bins of increasing redshift. They are all centred on the same reference model

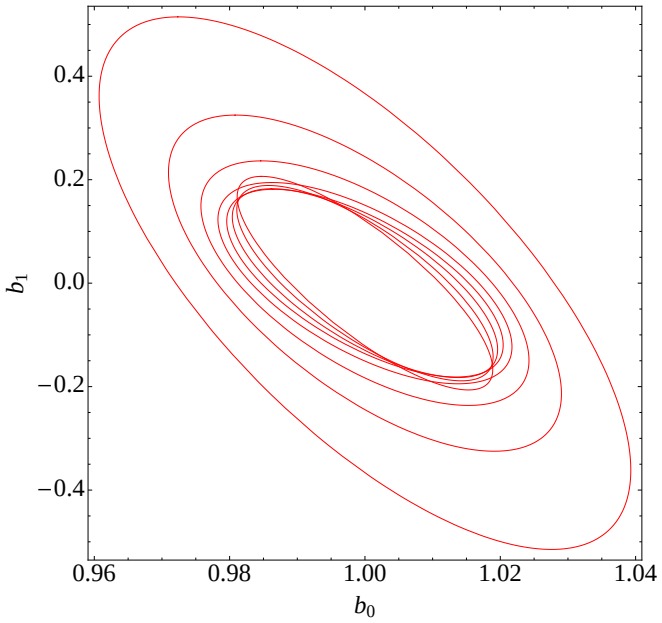


Figure 5.1: Contours plot at 68% probability for b_0 and b_1 estimated in different redshift bins. The higher the redshift the larger the contour. We show the case P-model with $n = 1$, for the reference case of an unbiased population of mass tracers. We set $k_1 = 1.0$.

that is the same at all redshifts. There is a clear anti-correlation between b_1 and b_0 since in a limited $k-$range on large scales an increase in b_0 can be compensated by a decrease in b_1. This is less true when the scale dependence is steep. Indeed the probability contours in the case $n = 2$ (not shown here) are much more circular.

Similar consideration apply to the Q-model. The correlation between A and Q indicates that these parameters can compensate each other to give the same spectral amplitude. The fact that in this case the covariance is positive simply reflects the fact that in the bias functional form the two terms appear at the nominator and denominator. The corresponding probability contours are shown in figure 5.2.

5.5 Realistic bias reference model: results

We have repeated the same Fisher Matrix analysis presented in the previous Section using, however, a more realistic reference bias model obtained from the mock *Euclid* catalog and fitted with both a P- and a Q-model. The only difference is that we now consider a smaller redshift range $z = [0.8, 1.8]$ which is likely to match the one that will be effectively probed by *Euclid*.

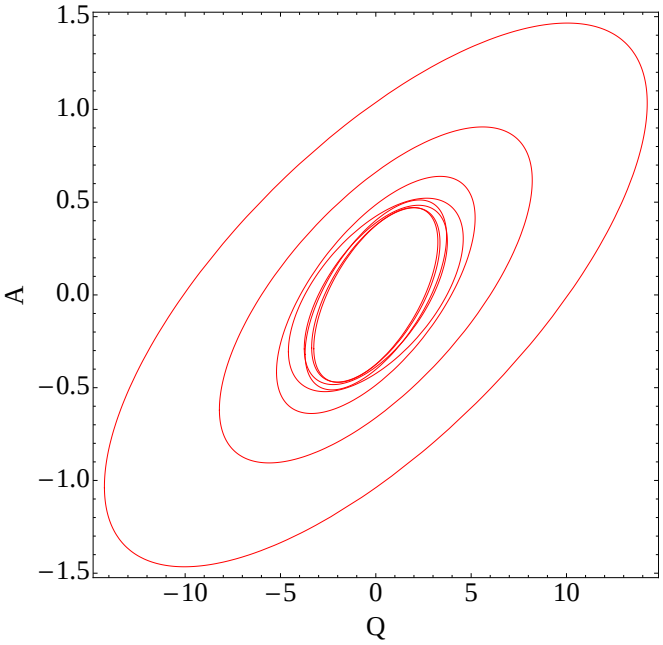


Figure 5.2: Contours plot at 68% probability for A and Q, Q-model parameters, estimated in different redshift bins, for the reference case of an unbiased population of mass tracers. We set $k_1 = 1.0$.

We first present the constraints on the cosmological parameters obtained with the various models. Those shown in Table 5.6 are marginalised over all other parameters.

We notice that the uncertainties on various parameters decrease by $\sim 10-20\%$ with respect to the case of the previous Section with the unbiased model as a reference. This is probably due to the change in the fiducial model and the resulting increase of the fiducial power spectrum due to the presence of a bias different from zero. An increase in the power should correspond to an increase in the value of the Fisher matrix elements (see eq. 5.9 and eq. 5.10), so that the errors decrease. Anyway we can say that the uncertainties on the cosmological parameters are quite similar among the different bias models and match quite well those obtained in the previous case with the unbiased model as a reference. This is reassuring and confirms the fact that the accuracy with which a survey like Euclid will be able to measure the cosmological parameters is at the % level, even when we drop the theoretical prejudice that galaxy bias is linear on the BAO scales. The only exception is the growth index γ for which the random errors in the realistic case are significantly larger than in the unbiased fiducial model, especially when one considers the Q-model, for which the relative random error can be as large as 10 %. This probably reflects the fact that, for this analysis, we have restricted the redshift range available, effectively reducing the information on the rate at which fluctuations grow. The increase of the uncertainty on γ (and the σ_8 parameter too) respect to the unbiased case can also probably reflect the fact that the bias in this fiducial model changes with the redshift, effectively increasing the number of degrees of freedom in the model and that these two parameters are the cosmological parameters mostly affected by the bias since the power spectrum is sensitive to the combinations $\sigma_8 b(k,z)$ and $f/b(k,z)$.

Table 5.6: Table of errors on cosmological parameters for the realistic biased P-model (with $n = 1$, $n = 1.28$ and $n = 2$) and Q-model, when one considers six, rather than 8, redshift bins and $k_1 = 1.0$.

		P-model		Q-model
σ_{cosmo}	$n = 1$	$n = 1.28$	$n = 2$	
σ_h	0.035	0.035	0.029	0.031
σ_{Ω_M}	0.015	0.014	0.012	0.013
σ_{Ω_b}	0.0033	0.032	0.0028	0.0027
σ_{n_s}	0.040	0.038	0.030	0.034
σ_γ	0.036	0.035	0.036	0.052
σ_{σ_8}	0.0049	0.0048	0.0046	0.0073

Table 5.7 shows the correlation parameters for all the combinations of cosmological parameters for the P-model (with $n = 1$). For each generic pair

of cosmological parameters (X, Y), the correlation parameters are computed as: $cov(X, Y)/(\sigma_X \sigma_Y)$, where $cov(X, Y)$ is the covariance between the two, and (σ_X, σ_Y) are their 1σ uncertainties. Of course some degeneracy among these cosmological parameters is to be expected. We note that the power spectrum shape parameters h, Ω_M, Ω_b and n_s are highly correlated, while γ is very poorly correlated with the other parameters, except for the normalization parameter σ_8. Let us focus on the covariance between σ_8 and γ, which is illustrated in fig. 5.3. In the plot we show the 68% probability contours in the σ_8-γ plane for the P-model case. The two parameters are anti-correlated. This is not surprising, since both parameters contribute to the amplitude of the power spectrum in redshift space so that an increase in the *rms* amplitude of density fluctuations can be compensated by a decrease in the index γ which, for a given value of Ω_M, also affects the spectral amplitude through the Kaiser boost. The Q-model case for the same couple of parameters is similar to the P-model case, the anti-correlation is confirmed, but the error size increases, due to higher number of free parameters in the model. Let us now focus on two additional cases in which we also expect covariance among parameters: Ω_M-h and n_s-Ω_b. Ω_M and h are expected to be highly degenerate since the spectral shape constrains the parameter combination $\Omega_M h^2$. This is clearly seen in figure 5.4 in which we show their 68 % probability contour. The degeneracy is broken by the additional information on the spectral amplitude and its evolution. This plot, like the next one, is obtained from the P-model (with $n = 1$). However, the very same trends are seen when the Q-model is considered.

Table 5.7: Correlations between the various cosmological parameters, obtained when assuming the realistic biased P-model (with $n = 1$) as reference model. The cosmological parameters are indicated in the first row and first column.

Parameters	h	Ω_M	Ω_b	n_s	γ
Ω_M	0.99				
Ω_b	1.00	0.98			
n_s	-0.94	-0.94	-0.92		
γ	0.0078	0.01	0.0089	-0.024	
σ_8	-0.32	-0.33	0.32	0.43	-0.55

In figure 5.5 we show the analogous plot for n_s and Ω_b. In this case the degeneracy reflects the fact that, in the $k-$range explored, a stronger tilt in the power spectrum can be mimicked by a larger baryon fraction. In this case the degeneracy can be broken by the BAO features that constitute a unique baryon signature.

The uncertainties on the bias parameters are listed in Table 5.8.

The results are very similar to those found with the reference unbiased model and listed in Table 5.5. Also in this case the Fisher matrix analysis shows that with a *Euclid*-like survey it will be possible to estimate the scale-

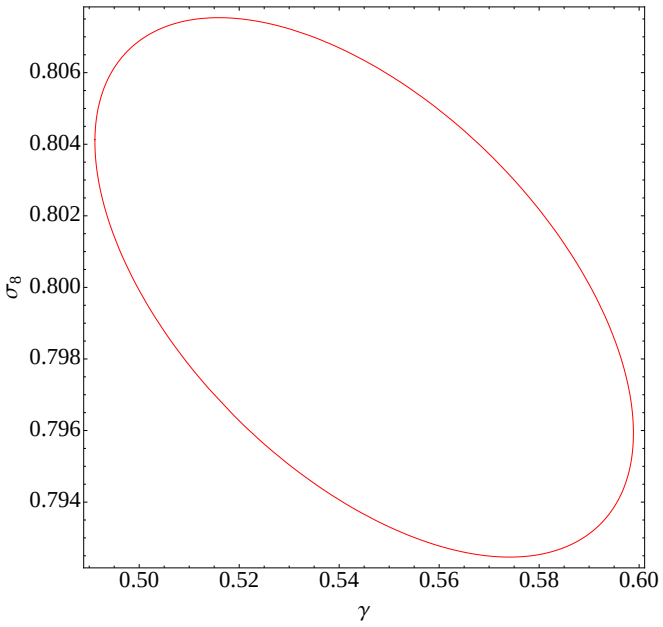


Figure 5.3: Errors on σ_8 parameter and γ in the case of realistic biased P-model with $n = 1$.

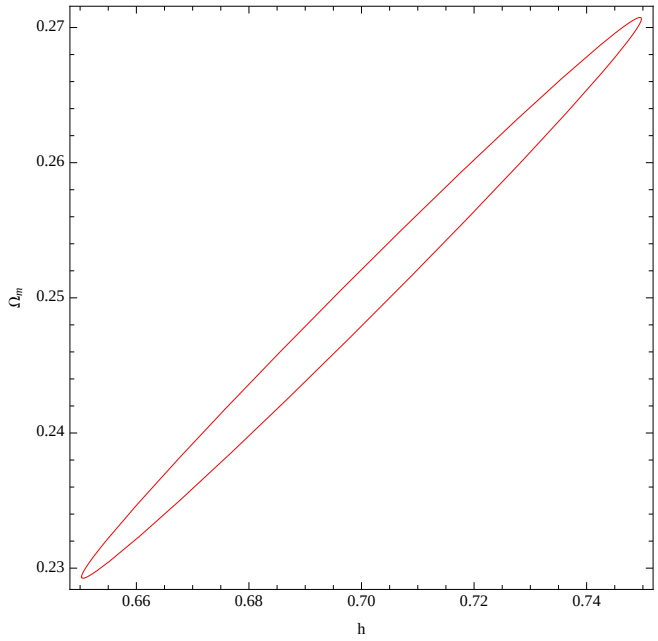


Figure 5.4: Errors on the Hubble constant parameter and Ω_M in the case of realistic biased P-model with $n = 1$.

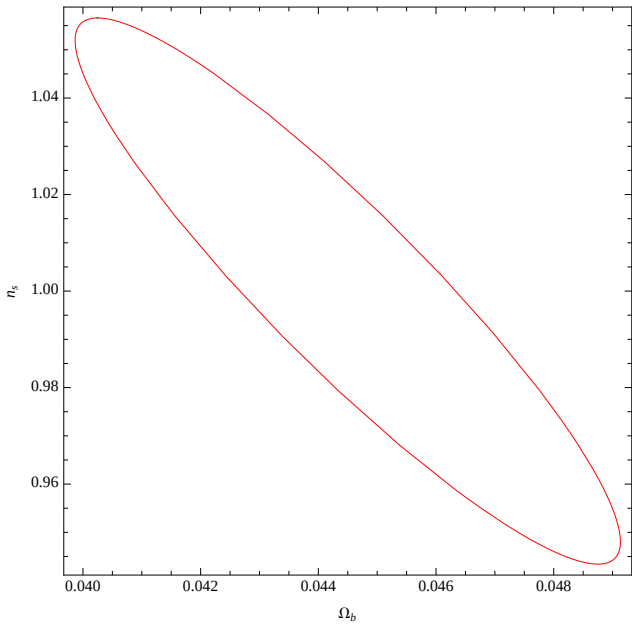


Figure 5.5: Errors on the primordial spectral index n_s and Ω_b in the case of realistic biased P-model, $n = 1$.

Table 5.8: $1-\sigma$ errors on bias parameters for the P- and Q-model for $k_1 = 1.0$ and for different $n-$ values, indicated in the header. The reference values are those listed in Table 5.2.

	P-model						Q-model		
	n=1		n=1.28		n=2				
z	σ_{b_0}	σ_{b_1}	σ_{b_0}	σ_{b_1}	σ_{b_0}	σ_{b_1}	σ_{b_0}	σ_Q	σ_A
0.8	0.014	0.12	0.012	0.19	0.011	0.91	0.029	2.6	0.46
1.0	0.016	0.12	0.014	0.18	0.013	0.76	0.033	1.9	0.39
1.2	0.018	0.12	0.016	0.18	0.016	0.66	0.039	1.6	0.36
1.4	0.021	0.13	0.019	0.20	0.019	0.69	0.045	1.6	0.34
1.6	0.024	0.16	0.022	0.24	0.023	0.89	0.049	1.8	0.37
1.8	0.028	0.21	0.026	0.32	0.027	1.3	0.056	2.1	0.40

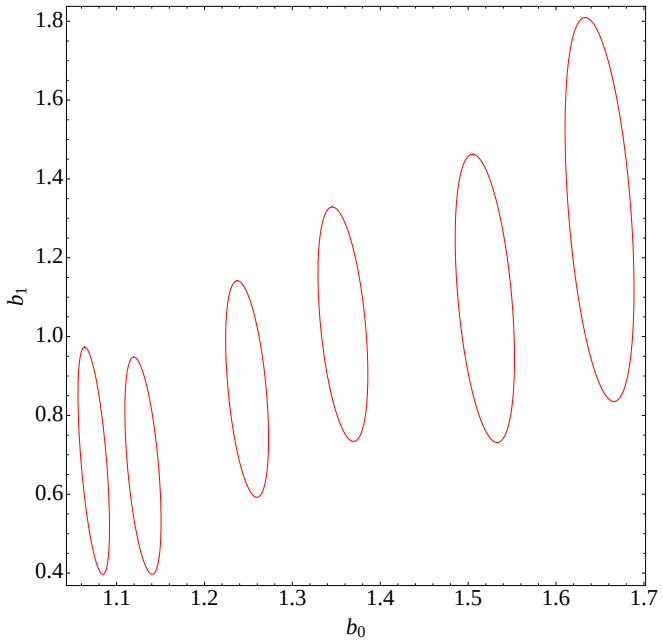


Figure 5.6: Constraints on b_0 and b_1 for six redshift values for the realistic biased P-model with $n = 1.28$ as reference model and $k_1 = 1$.

independent part of the galaxy bias parameters at the per-cent accuracy in all redshift bins. The sensitivity to the scale-dependence is lower and depends on its exact form. If the scale dependence is steep, i.e. if it is more evident on small scales, then it is hardly surprising that estimating the power spectrum on BAO scales will provide poor constraints. On the contrary, if scale dependence is mild, then the additional bias parameter b_1 in the P-model can be measured with a 10-20 % accuracy.

To illustrate the covariance between b_0 and b_1 we consider the representative case of $n = 1.28$. The corresponding 68 % probability contours are shown in figure 5.6. The different ellipses refer to the constraints from the six redshift bins. Each one is centred on its own reference model. No significant covariance is seen in the parameters. As it is apparent from the size of the ellipse, uncertainties increase with the redshift. The probability contours for the A and Q parameters, not shown here, are entirely consistent with those shown in figure 5.2.

Conclusions

In the recent years galaxy clustering has emerged as one of the most powerful probes to investigate the properties of the Universe and, in particular, the nature of its dark components: the dark matter and the dark energy. Quantitative estimates of galaxy clustering can be obtained from spectroscopic redshift surveys that can map the three dimensional distribution of galaxies.

Not surprisingly, large scientific collaborations are carrying out ambitious research programs aimed at performing large galaxy redshift surveys over ever increasing volumes and out to large redshift. At the end of the next decade, next generation surveys like *DESI* and *Euclid* will be complete and will have probed about 2 % of the observable Universe.

These ambitious programs aim at precision cosmology, i.e. at estimating cosmological parameters at the % accuracy, which is a pre-requisite to discriminate among alternative cosmological models and shed some light in the nature of the dark components. With such small random uncertainties, the risk is that the error budget could be dominated by systematic errors that may arise from many sources: nonlinear effects, imperfect estimator for galaxy clustering, the ill-constrained bias relation between galaxy and matter. Local studies, can, in this respect, play an important role to investigate the origin of some of these systematic errors and to set a sort of 'zero point' at $z = 0$, for the quantities that one aims to constrain, such as the growth rate of density fluctuations or galaxy bias.

In this thesis we have focused on the single largest local redshift survey to date, the 2MRS catalog, of about 40000 objects with measured redshift over the full sky. Indeed, large sky coverage is another desirable feature of this sample, since next generation surveys will also aim at wide sky coverage and investigating galaxy clustering over large angular separations is potentially affected by specific systematic errors. Here, we have focused on one of the most common statistical tools to galaxy clustering: the power spectrum. In particular, we have considered the FKP estimator with its many advantages (reliability, speed, flexibility) and few drawbacks (it is not designed for wide angle surveys).

With this general purpose as a guideline, we have addressed the specific issues detailed below.

- We first focus on the implementation of our own FKP estimator. While rather standard, there are subtleties in this estimator (like the choice of weights) that can be best dealt with by the developer of the code rather than its user. The making of the code and its validation have been performed by comparing the outputs of different, similar codes applied to the same mock catalog. Differences among the codes were below $10^{-5}\%$. This code-contest was performed in the framework of the Euclid Consortium activity to select a prototype code candidate for further development and implementation in the data analysis pipeline.

- The use of the Fast Fourier Transform is the main feature of the code and determines its speed. However, it also sets a preferential (cubic) geometry that does not match the intrinsic (quasi-spherical) 2MRS geometry. In particular, our specific implementation of embedding the surveyed volume in a cubic box, prevents us from performing a full analysis of the power spectrum and its anisotropies and extract from them useful cosmological information. For this reason we have focused on the power spectrum monopole only. The non-cubic geometry of the sample induces systematic differences between the estimated and the true power spectrum. This needs to be accounted for through an accurate modelling of the window function of the survey which we estimated numerically through the random galaxy catalog. In the case of a homogeneous, almost all-sky survey the window function is simple enough to be accurately estimated (in fact is quite close to an analytic top hat) and does not significantly affect the error budget. This has been assessed by means of a suite of simulated catalogs of extragalactic objects. Since the goal here was to assess geometry effect rather than astrophysical issues, we have considered catalogs of dark matter halos extracted from the large computational box of the BASICC simulation, rather than aiming for realistic mock galaxy catalogs that would have been smaller in size and number of objects.

  More specifically, we have extracted 125 samples from the parent halo catalog with size, geometry, radial selection and number of objects similar to that of 2MRS and computed their power spectra over a typical range [k_{min} $\sim$0.03,k_{NL} $\sim$0.5-0.6] h Mpc^{-1}. After accounting for the effect of the window function, which we do by means of a convolution integral that involves the true halo power spectrum measured from the full sample, we find no evidence for systematic errors on scales larger than $k < 0.7$ h Mpc^{-1}. This result is a good news for next generation surveys with large angular coverage on both hemispheres, like *Euclid*, that are also expected to be characterised by relatively simple window function that one could model with high precision.

- The ability to measure the redshift space distortions and the accuracy with which one can estimate the so called distortion parameter β closely related to the growth rate of density fluctuations is the second issue that we have addressed. A first assessment was done by using the very same halo catalogs extracted from the BASICC box. The goal there was to search for the typical scale at which, at $z = 0$, non linear effects become non-negligible and need to be modelled to obtain an unbiased estimate of β. This exercise shows that $k < 0.2\ h\ \mathrm{Mpc}^{-1}$ is a safe zone where to apply linear theory. However, this is probably an optimistic estimate since halos were extracted by ignoring their substructures, hence underestimating the impact of nonlinear motion on small scales.

- For this reason we have repeated the same analysis using, however, a suite of realistic mock 2MRS galaxy catalogs extracted from the Millennium Simulation. Mock galaxies were modelled from the halo and sub-halo populations using semi-analytic techniques. This new analysis demonstrated that non linear effects need to be modelled if one wants to analyse the galaxy power spectrum beyond $k \sim 0.1 - 0.15$ $h\ \mathrm{Mpc}^{-1}$.

  We found that a model redshift space power spectrum for galaxies obtained by combining the nonlinear matter power spectrum generated by *Halofit*, a small scale Gaussian damping term, a linear galaxy bias and a Kaiser boost is adequate to match the power spectrum of the mock 2MRS galaxies down to $k \sim 0.5\ h\ \mathrm{Mpc}^{-1}$ and to extract an unbiased estimate of the β parameter. An alternative model, which provides a similarly good fit, is one in which the matter power spectrum is described by linear theory using *CAMB* and nonlinear effects are modelled by a scale-dependent term similar to the one proposed by Cole *et al.* (2005).

- Realistic mocks were then used to pave the way to the analysis of the real 2MRS dataset. For that, we needed to estimate random errors which we have obtained from the clustering analysis of all the realistic mocks. The limited number of mocks didn't allow us to build the full covariance matrix. However, the fact that we have considered a power spectrum statistics and that we focused on linear or quasi-linear scales should limit the amplitude of the off-diagonal terms. Therefore, in the present analysis we have assumed that they are all negligible and that the errors can be obtained from the *rms* scatter among the mocks.

- Having calibrated the power spectrum analysis using the 2MRS mocks, we then turned to the real sample to estimate the physically relevant parameters. We first focused on the bias parameter and its dependence on galaxy properties. The analysis we have performed here is similar

to that of Westover (2007) in his PhD thesis with three important differences: *1)* we used a larger and deeper sample of galaxies with $K_s \leq 11.75$, *2)* we investigated galaxy clustering using the power spectrum instead of the spatial 2-point correlation function, *3)* the focus on large, quasi-linear scales instead of small, nonlinear scales. With this respect our results complement and improve those of Westover (2007) and, in absolute terms, they significantly improve current estimates of galaxy bias at $z \simeq 0$.

- Our results on the luminosity dependence of galaxy bias confirm the qualitative behaviour found by Westover (2007) on a dataset similar to ours and by Norberg et al. (2001) and Tegmark et al. (2004) using the 2dF and SDSS galaxy catalogs, respectively. The bias is confirmed to be a steep function of the luminosity for galaxies brighter than L^*, whereas for fainter objects the dependence is milder or absent altogether. The form of the bias vs. luminosity relation is very similar to that obtained by Westover (2007) when the bias is estimated in redshift space, like in our case. There is an offset difference, however, that would disappear when changing the normalization adopted by Westover (2007), which is based on a single, anomalously high, data point. Westover (2007) also provides an estimate of the bias in real space. In that case the normalisation problem disappears but bias vs. luminosity relation is found to be significantly steeper, in disagreement with our results on both the bright and faint end of the relation. Indeed, there is a better agreement with the results of Norberg et al. (2001) and Tegmark et al. (2004), the second of which where obtained from power spectra measured on large scales. All these discrepancies can be accounted for by considering the specific differences among these analyses. What is reassuring, and of astrophysical interest, is the fact that the dependence of the bias on luminosity at low redshift is confirmed, with a very similar trend, all the way down to $z = 0$.

- The dependence of the bias on galaxy morphological type has been explored as well. The issue is a long standing one and has been verified in several analyses. Here we have repeated the exercise with the goal of increasing the statistical evidence of the phenomenon at $z \sim 0$ where the available samples have been either small, like in the Westover (2007) analysis, or dominated by one single galaxy type, as in the IRAS 1.2 Jy or PSCz (Saunders et al., 2000) catalogs that contained almost exclusively late-type galaxies. 2MRS also contains a large fraction of late-type objects but in our analysis we have been able to divide the sample into two sub-catalogs, one dominated by early type galaxies and the other by $S0/a$ and spiral galaxies, with a similar number of members. The result has confirmed that early type

objects are more clustered than the late type ones, with a relative bias of 1.228 ± 0.067, in agreement with the previous findings of Westover (2007).

- The second issue we focused on was the estimate of cosmological parameters by comparing the measured monopole of the 2MRS power spectrum with model predictions. Ideally one would like to infer the values of all cosmological parameters that characterise the underlying mass power spectrum and its redshift distortions. Unfortunately, the limited k-range probed by the survey and the fact that we have restricted the analysis to the monopole only does not allow to remove the degeneracy among the parameters. For this reason we have decided to rely on external priors and fix all parameters except the bias and the mass density to the values estimated by the Planck team (2013). This is in principle sufficient to remove any degeneracy and to provide independent estimates of the two parameters. Second, under the hypothesis that General Relativity holds on cosmological scales, these two values can be used to determine the growth rate of density fluctuations $f = \Omega_M^{0.55}$ from the distortion parameter $\beta = f/b$ that determines the strength of the Kaiser boost-type redshift space distortion.

  This analysis allowed us to constrain Ω_M and b with an accuracy $\sim 20\%$ and $\sim 3\%$, respectively with best fitting values $\Omega_M \simeq 0.33$ and $b \simeq 1.23$ in agreement with previous results and theoretical prejudices. To assess the robustness of these results to nonlinear effects, that on the scale of the analysis are expected to be rather small but not necessarily negligible, we have used several different models for the redshift space galaxy power spectrum, in which we have used different receipts to model nonlinear effects. The results from the different models, each one obtained in its own k-range, turned out to be in very good agreement, indicating that nonlinear effects are well under control. In so doing we have considered models that depend on some additional parameter whose physical interpretation may not be straightforward. In those cases the additional parameter was kept free and marginalised over in the final results.

- The previous results showed that non linear effects do affect the analysis of the galaxy power spectrum even at relatively large scales. Their typical footprint is to induce additional scale dependence in the linear power spectrum; a fact that can be exploited to detect and characterise them. One of such effect is scale dependent bias, which is certainly expected on small scales affected by stellar feedback effects but can also affect much larger scales like those in which the BAO feature is observed, especially in the analysis of the power spectrum.

Scale dependent bias is known to be a relatively small effect, though, and would not affect the conclusion of our 2MRS analysis. However, it potentially affects the results of the clustering analysis of next generation surveys like *Euclid* from which, as we have seen, one expects to estimate the cosmological parameters with a per-cent accuracy. For this purpose we have investigated the impact of possible scale dependent bias in the power spectrum analysis of a *Euclid*-like survey and the possibility to use the analysis itself to detect and quantify this scale dependence.

We have therefore investigated this issue. In this case, instead of using mock galaxy catalogs as in the 2MRS case, we adopted a Fisher matrix approach. Mock *Euclid* galaxy catalogs were indeed used but only to obtain a realistic bias model as a reference. The Fisher Matrix analysis was performing in different spherical, not-overlapping shells over a large redshift range $z = [0.6, 2.0]$ that we have subsequently reduced to $z = [0.8, 1.8]$ to mimic the more realistic case of the real *Euclid* survey and to explore the impact of reducing the interval. The number of objects in each shell was set to match the expected values. The analysis involved all type of free parameters, the ones that describe the background cosmology, these that describe the density fluctuations and their growth and the bias parameters. The results show that adding an extra degree of freedom in the model power spectrum to allow for a possible scale dependent bias does not significantly affect the quality of the fit. All relevant parameters are estimated with a few per-cent accuracy except the growth index γ. The error on this parameter was found to depend on the specific bias model and in the range 5-10 % when the reduced redshift interval is considered.

The accuracy on galaxy bias depends on its scale dependence. The scale independent part of the bias, i.e. the 0-th order of the Taylor expansion of the overdensity field, can be determined with high (~ 2 %) accuracy. The errors on the parameters that quantify the amplitude of the scale dependence are significantly larger. Their magnitude depends on the exact type of scale dependence. If the dependence is linear in k, which corresponds to a first order term in the Taylor expansion, then errors increase to a level of $\sim$ 10 %. In case of quadratic dependence the errors are of the same size as the signal and the scale dependence cannot be detected, at least if the power spectrum analysis is focused on BAO scales, as we have assumed here.

The common aspect of the various analyses presented in this thesis is the analysis of galaxy clustering in wide angle, almost all-sky surveys, its specific issues and possible solutions. With this respect, the work presented here can be expanded along several directions. A crucial one, also considering

the upcoming galaxy surveys, is the choice of the best estimator for the power spectrum. With an almost spherical geometry, the natural way of characterising galaxy clustering would be to consider some spherical harmonics rather than Fourier-based estimator. The so-called Fourier-Bessel estimator (see e.g. Tadros et al. 1999) in which the galaxy density field is expanded in spherical harmonics on the 2=sphere and in Bessel functions along the radial direction would be the natural choice. However, this estimator is computationally too costly to deal with the large number of objects in the next generation surveys.

The question is then whether the FKP estimator can still be used for the spectral analysis and how. Embedding the surveyed volume in a box, as we did here with the 2MRS, is not an option both because of the overwhelmingly large fraction of volume that would remain empty and also because, has we have seen, this choice does not allow to probe the anisotropy in galaxy clustering and the estimate of the fluctuations' growth rate. A possibility would be to "pixelize" the survey with large, possibly overlapping boxes with their axis aligned along the line of sight and then combine the various power spectra in a self consistent fashion. Or to project the celestial sphere onto a flat coordinate system, using methods that are common in cartography, and pixelize the flat sky in the new coordinate system.

One of the future goals is to asses the performance of these methods using dedicated mock catalogs or real surveys. The problem with this second option is that 2MRS is the best all-sky catalog available so far with no plan for further extension, while a crucial feature for the study described above is depth. With this respect, a significant improvement can be obtained by using photometric-redshift catalogs instead of spectroscopic catalogs. The all sky 2MASS Photometric Redshift catalog (2MPZ) (Bilicky, 2014, now publicly available at http://surveys.roe.ac.uk/ssa/TWOMPZ) is probably the best example. This catalog contains about 900000 galaxies out to $z = 0.3$ with well calibrated photometric redshifts. Photometric redshifts are intrinsically less precise of the spectroscopic ones, which effectively dilutes the clustering signal, and have non-Gaussian errors that however can be effectively characterised if the spectroscopic control sample is large, representative and reliable. If the goal is to optimise the use of the FKP estimator and not to obtain accurate estimates of cosmological parameters, then this catalog is probably the best way to calibrate an FKP-based analysis on a realistic all-sky catalog. An effective relative calibration can be obtained by comparing results with those obtained by alternative estimators currently available. Examples in this direction are the standard estimator of the angular power spectrum applied, however, to the galaxy distribution in different redshift shells. Or the estimator proposed by Beutler et al. (2014) aimed at minimising large angle distortions. All this we plan to investigate in the next future.

Acknowledgments

First of all, I would like to express my gratitude to my supervisor Professor Enzo Branchini, for his careful guidance and support for this thesis work.

I would like also to thank all collaborators who have taken part in this work. Thanks Andres Balaguera Antolìnez, for code comparison and 2MRS analysis and for his help and support.

I am very grateful to Prof. Luca Amendola, Eloisa Menegoni and Cinzia di Porto, for the work on scale-dependent bias, for the results and all the plots of Chapter 5.

This thesis work uses data by the 2MRS team (2.4 version of the 2MRS catalog), 2MRS mock catalogs extracted from the Millennium Simulation (Springel et al., 2005) and generated from the model of De Lucia & Blaizot (2007), the BASICC simulation data by Raul Angulo et al., the Euclid "100deg^2" mock galaxy catalogs provided by Alex Merson.

Thanks Prof. Adi Nusser for providing me the FoG collapse code and the F/T method code.

I am grateful to Prof. Matteo Viel, who provided me dark matter cosmolgical simulations on which we tested our code, Federico Marulli, for his technical support, Prof. Lado Samushia and Prof. Will Percival for code comparison, Mauro Roncarelli for code implementation and Maciej Bilicki for his introduction to 2MPZ.

I thank Prof. Lauro Moscardini for his comments.

I thank Roma Tre University.

Last, I would like to thank all the PhD students in the Roma Tre Physics Department, my friends and my family.

Bibliography

[1] Abazajian K. N. et al., The Seventh Data Release of the Sloan Digital Sky Survey *ApJS, 182, 543, 2009*, [arXiv:0812.0649]

[2] Ahn C. P. et al., The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment *ApJS, 211, 17, 2014*, [arXiv:1307.7735]

[3] Angulo R. E. et al., The detectability of baryonic acoustic oscillations in future galaxy surveys *Mon.Not.Roy.Astron.Soc.383:755,2008*, [arXiv:astro-ph/0702543]

[4] Balaguera-Antolinez A. et al., The REFLEX II galaxy cluster survey: power spectrum analysis *Mon.Not.Roy.Astron.Soc.413:386B,2011*, [arXiv:astro-ph/1012.1322]

[5] Baldry I.K. et al., Quantifying the bimodal color-magnitude distribution of galaxies *Astrophys.J.600:681-694,2004*, [arXiv:astro-ph/0309710v1]

[6] Baugh C. M. et al., Can the faint submillimetre galaxies be explained in the Λ cold dark matter model? *Mon.Not.Roy.Astron.Soc., 356, 1191, 2005*, [arXiv:astro-ph/0406069]

[7] Bernardeau F. et al., Large-Scale Structure of the Universe and Cosmological Perturbation Theory *Phys.Rept.367:1-248,2002*, [arXiv:astro-ph/0112551]

[8] Beutler F. et al., The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Testing gravity with redshift-space distortions using the power spectrum multipoles *Mon.Not.Roy.Astron.Soc.443.1065B,2014* [arXiv:astro-ph/1312.4611v2]

[9] Bianchi D. et al., Statistical and systematic errors in redshift-space distortion measurements from large surveys *Mon.Not.Roy.Astron.Soc.427, 2420, 2012* [arXiv:1203.1545]

[10] Bilicki M. et al, Mapping the Cosmic Web with the largest all-sky surveys [arXiv:astro-ph/1408.0799]

[11] Blake C. et al, The WiggleZ Dark Energy Survey: the growth rate of cosmic structure since redshift z = 0.9 *Mon.Not.Roy.Astron.Soc.415.2876B,2011*, [arXiv:astro-ph/1104.2948]

[12] Bower R. G. et al, Breaking the hierarchy of galaxy formation *Mon.Not.Roy.Astron.Soc.370:645B,2006*, [arXiv:astro-ph/0511338]

[13] Branchini E. et al, The velocity field of 2MRS, Ks=11.75 galaxies: constraints on β and bulk flow from the luminosity function *Mon.Not.Roy.Astron.Soc.424:472B,2012*, [arXiv:astro-ph/1202.5206]

[14] Branchini E. et al, Probing the Mass Distribution with IRAS Galaxies, [arXiv:astro-ph/0110611]

[15] Coe D., Fisher matrices and confidence ellipses: a quick-start guide and software *arXiv:0906.4123v1 [astro-ph.IM]*

[16] Cole S. et al, Fourier Analysis of Redshift Space Distortions and the Determination of Ω. *Mon.Not.Roy.Astron.Soc., 267: p.785,1994*, [arXiv:astro-ph/9308003]

[17] Cole S. et al, Hierarchical Galaxy Formation *Mon.Not.Roy.Astron.Soc.319, 168, 2000*, [arXiv:astro-ph/0007281]

[18] Cole S. et al, The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications *Mon.Not.Roy.Astron.Soc.362:505C,2005*, [arXiv:astro-ph/0501174]

[19] Coles P., Lucchin F., Cosmology. The origin and evoloution of cosmic structrures. *Wiley, 2002*

[20] Colless M. et al, The 2dF Galaxy Redshift Survey: Final Data Release, 2003 [arXiv:astro-ph/0306581]

[21] Cooray A., Non-linear galaxy power spectrum and cosmological parameters *Mon.Not.Roy.Astron.Soc.348, 250C,2004*, [arXiv:astro-ph/0311515]

[22] Cooray A. and Sheth R., Halo models of large scale structure *PhR,372,1C,2002*, [arXiv:astro-ph/0206508]

[23] Cowie L.L. et al., The Hawaii K-band galaxy survey: I. Deep K-band imaging. *ApJ, 434, 114, 1994.*

[24] Cui W. et al. An ideal mass assignment scheme for measuring the Power Spectrum with FFTs *ApJ, 687:p.738,2008*, [arXiv:0804.0070]

[25] D'Agostini G., Le basi del metodo sperimentale *Dispense*, Ottobre 2001

[26] Davis M. and Hucra J., A survey of galaxy redshifts. III - The density field and the induced gravity field, *ApJ, 254, 437, 1982.*

[27] Davis M. et al., Local Gravity versus Local Velocity: Solutions for β and nonlinear bias *Mon.Not.Roy.Astron.Soc.413 ,2906D,2011*, [arXiv:1011.3114]

[28] De Lucia G. and Blaizot J., The hierarchical formation of the brightest cluster galaxies *Mon.Not.Roy.Astron.Soc.375 , 2, 2007.*

[29] de Putter R. et al., Thinking Outside the Box: Effects of Modes Larger than the Survey on Matter Power Spectrum Covariance *JCAP, 04, 019D,2012*, [arXiv:1111.6596]

[30] Vaucouleurs, G. de, Vaucouleurs, A. de, Corwin, H.G., Second Reference Catalogue of Bright Galaxies *University of Texas Press, Austin, 1976*

[31] di Porto, C. et al., Growth factor and galaxy bias from future redshift surveys: a study on parametrizations *Mon.Not.Roy.Astron.Soc.419 , 2, 2012.* [arXiv:1101.2453]

[32] di Porto, C. et al., Simultaneous constraints on bias, normalization and growth index through power spectrum measurements *Mon.Not.Roy.Astron.Soc. Letters, 423 , L97, 2012.* [arXiv:1201.2455]

[33] di Porto, C. et al., The VIMOS Public Extragalactic Redshift Survey (VIPERS). Measuring nonlinear galaxy bias at z$\sim$0.8 [arXiv:1406.6692]

[34] Dodelson S., Modern Cosmology, *Academic Press*, 2003

[35] Dressler et al., Galaxy morphology in rich clusters − Implications for the formation and evolution of galaxies *ApJ, 236, 351, 1980*

[36] Einstein A., Die Feldgleichungen der Gravitation, *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin, 844–847, 1915*

[37] Einstein A., Die Grundlage der allgemeinen Relativit*ä*tstheorie, *Annalen der Physik 49, 1916*

[38] Einstein A., Kosmologische Betrachtungen zur allgemeinen Relativit*ä*tstheorie, *Sitzungsberichte der Preußischen Akademie der Wissenschaften: 142, 1917*

[39] Eisenstein D. and Hu W., Baryonic Features in the Matter Transfer Function *ApJ,496,605E,1998*, [arXiv:astro-ph/9709112]

[40] Eisenstein D. J. et al., Detection of the Baryon Acoustic Peak in the Large−Scale Correlation Function of SDSS Luminous Red Galaxies *ApJ,633 (2), p.560,2005* [arXiv:astro-ph/0501171]

[41] Eisenstein D. J. et al., On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter *ApJ, 664, 660, 2007* [arXiv:astro-ph/0604361]

[42] Feldman H. , Kaiser N., Peacock J., Power spectrum analysis of three-dimensional redshift surveys, *ApJ,426:23-37,1994*, [arXiv:astro-ph/9304022]

[43] Fisher, R. A. Fisher J., *Roy. Stat. Soc. 98, 39-54, 1935*

[44] Freedman W.L., Madore B.F., The Hubble constant, *arXiv:1004.1856v1 [astro-ph.CO]*

[45] Freedman W.L., et al., Final results from the Hubble Space Telescope Key Project to misure the Hubble Constant, *arXiv:astro-ph/0012376*, 2000

Frigo M., Johnson S., 2005, Proceedings of the IEEE,93, 216

[46] Frigo M. and Johnson S., , The Design and Implementation of FFTW3, *Proceedings of the IEEE,93, 216, 2005*

[47] J. N. Fry and E. Gaztanaga, *ApJ, 413, 447, 1993* [arXiv:astro-ph/9302009]

[48] Garilli B., The VIMOS Public Extragalactic Survey (VIPERS). First Data Release of 57 204 spectroscopic measurements *Astronomy & Astrophysics, 562, A23, 2014* [arXiv:1310.1008]

[49] Giannone P., Elementi di astronomia, *Pitagora editrice Bologna*

[50] Gil-Marìn H. et al, The power spectrum and bispectrum of SDSS DR11 BOSS galaxies I: bias and gravity [arXiv:astro-ph/1407.5668]

[51] Hamilton A.J.S., Linear redshift distortions: a review *ASSL, 231,185, 1997* [arXiv:astro-ph/9708102]

[52] Hamilton A.J.S. et al., Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies *ApJ. Lett., 374, L1, 1991*

[53] Hamilton A.J.S. and Tegmark M., The Real Space Power Spectrum of the PSCz Survey from 0.01 to 300 h Mpc^{-1} *Mon.Not.Roy.Astron.Soc.,330, 506, 2002* [arXiv:astro-ph/0008392]

[54] Hockney, R. W. and Eastwood, J. W., Computer Simulation Using Particles *Taylor & Francis, 1989.*

[55] H. Hoekstra et al., *ApJS, 577, 604, 2002* [arXiv:astro-ph/0206103]

[56] Hucra J. et al., The 2MASS Redshift Survey- Description and data release *ApJS, 199, 26,2012* [arXiv:1108.0669]

[57] Jarrett T.H. et al., 2MASS Extended Source Catalog: Overwiew and Algorithms *The Astronomical Journal, 119:2498-2531, 2000* [arXiv:astro-ph/0004318v1]

[58] Jeong, D. 2010, PhD thesis, The University of Texas

[59] Jeong, D. and Komatsu, E., Perturbation Theory Reloaded: Analytical Calculation of Nonlinearity in Baryonic Oscillations in the RealSpace Matter Power Spectrum *ApJ, 651, 619, 2006* [arXiv:astro-ph/0604075]

[60] Jeong D. and Komatsu E. Perturbation Theory Reloaded. II. Nonlinear Bias, Baryon Acoustic Oscillations, and Millennium Simulation in Real Space *ApJ, 691, 569, 2009* [arXiv:0805.2632]

[61] Jing Y.P., Correcting for the alias effect when measuring the power spectrum using FFT *ApJ, 620: 559J,2005* [arXiv:astro-ph/0409240]

[62] Jing Y.P. et al, Spatial Correlation Function and Pairwise Velocity Dispersion of Galaxies: Cold Dark Matter Models versus the Las Campanas Survey *ApJ, 494, 1, 1998* [arXiv:astro-ph/9707106]

[63] Kaiser N., On the spatial correlations of Abell clusters *ApJ„ 284L, 9, 1984*

[64] Kaiser N., Clustering in real space and in redshift space *Mon. Not. R. Astron. Soc., 227, 1, 1987*

[65] T. D. Kitching and A. Amara, *Mon. Not. R. Astron. Soc, 398, 2134, 2009*

[66] Kochanek C.S. et al., The K-Band Galaxy Luminosity Function *ApJ, 560: p.566,2001* [arXiv:astro-ph/0011456]

[67] Komatsu E. et al., Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation *ApJS, 192, 18, 2011* [arXiv:1001.4538]

[68] Lahav O. et al., Measuring our universe from galaxy redshift surveys *Living Reviews in Relativity, vol. 7, no. 8,2004* [arXiv:astro-ph/0310642]

[69] Laurejis R. et al., Euclid Definition Study Report, 2011 [arXiv:1110.3193]

[70] Laurejis R. et al., Euclid: ESA's mission to map the geometry of the dark universe *SPIE, 8442E, 0TL, 2012* [arXiv:astro-ph/0310642]

[71] Lewis A. et al., Efficient Computation of CMB anisotropies in closed FRW models, *ApJ., 538, 473, 2000* [arXiv:astro-ph/9911177]

[72] Loveday et al., Galaxy and Mass Assembly (GAMA): ugriz galaxy luminosity functions *Mon. Not. R. Astron. Soc, 420, 1239, 2012*

[73] Magira H. et al., Cosmological Redshift-Space distortion on clustering of high-Redshift objects: correction for nonlinear effects in the power spectrum and tests with $N-$body simulations *ApJ:528, p.30,2000* [arXiv:astro-ph/9907438]

[74] Marulli F. et al., Cosmology with clustering anisotropies: disentangling dynamic and geometric distortions in galaxy redshift surveys *Mon. Not. R. Astron. Soc.,426 2566,2012* [arXiv:1203.1002]

[75] Merson A. et al., Lightcone mock catalogues from semi-analytic models of galaxy formation - I. Construction and application to the BzK colour selection *Mon. Not. R. Astron. Soc.,429, 556,2013* [arXiv:1206.4049]

[76] Montesano, F. 2011, PhD thesis

[77] Norberg P. et al., The 2dF Galaxy Redshift Survey: Luminosity dependence of galaxy clustering *Mon. Not. R. Astron. Soc, 328, 64, 2001* [arXiv:astro-ph/0105500]

[78] Norberg P. et al., The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type *Mon. Not. R. Astron. Soc, 332, 827, 2002* [arXiv:astro-ph/0112043]

[79] Orsi A. et al., Probing dark energy with future redshift surveys: A comparison of emission line and broad band selection in the near infrared *Mon. Not. R. Astron. Soc, 405, 1006, 2010* [arXiv:0911.0669]

[80] Peacock J.A. and the 2dFGRS team, Measuring Large-Scale Structure with the 2dF Galaxy Redshift Survey (Plenary Talk) *AIP Conf. Proc. 586, 245 (2001),*[arXiv:astro-ph/0105450v1]

[81] Parkinson D. et al., The WiggleZ Dark Energy Survey: Final data release and cosmological results *Physical Review D, 86, 103518, 2012* [arXiv:1210.2130]

[82] Peacock, J. A. and Dodds, S. J. Non-linear evolution of cosmological power spectra *Mon. Not. R. Astron. Soc., 280, L19, 1996* [arXiv:astro-ph/9603031]

[83] Percival W. et al., The 2dF Galaxy Redshift Survey: The power spectrum and the matter content of the universe *Mon. Not. R. Astron. Soc,327: p.1297,2001* [arXiv:astro-ph/0105252]

[84] Percival W. et al., Fourier analysis of luminosity-dependent galaxy clustering *Mon. Not. R. Astron. Soc,347, 654,2004* [arXiv:astro-ph/0306511]

[85] Percival W. and White M., Testing cosmological structure formation using redshift-space distortions, *Mon. Not. R. Astron. Soc, 393, 297, 2009*

[86] Perlmutter, S. et al., Measurements of Omega and Lambda from 42 High-Redshift Supernovae *ApJ, 517: p.565, 1999* [arXiv:astro-ph/9812133]

[87] Planck Collaboration, Planck 2013 results. XVI. Cosmological parameters [arXiv:1303.5076]

[88] Press W. H. and Schechter P., Formation of Galaxies and Clusters of Galaxies by Self- Similar Gravitational Condensation *ApJ, 187, 425, 1974*

[89] Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P., Numerical recipes in FORTRAN. The art of scientific computing, *Cambridge University Press, 1992*

[90] Riess, A. G., et al., Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant *ApJ, 116: p.1009,1998* [arXiv: astro-ph/9805201]

[91] Ross A. et al., The Clustering of Galaxies in SDSS-III DR9 Baryon Oscillation Spectroscopic Survey: Constraints on Primordial Non-Gaussianity *Mon. Not. R. Astron. Soc,.428: p.1116,2013* [arXiv:1208.1491]

[92] Rubin V. et al., 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, Part 1, vol 238: p. 471, 1980*

[93] Ryden B., Introduction to cosmology, *Addison Wesley*

[94] Saunders W., The PSCz catalogue *Mon. Not. R. Astron. Soc., 317, 55, 2000*

[95] Schmidt M., Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources *ApJ, 151, 394, 1968*

[96] Scoccimarro R. et al., Power Spectrum Correlations Induced by Non-Linear Clustering *ApJ, 527, 1, 1999* [arXiv:astro-ph/9901099]

[97] Seljak, U., Zaldarriaga, M., *ApJ, 469, 437 1996*

[98] Seo H-J, Eisenstein D.J., Probing dark energy with barionic acoustic oscillations from future large galaxy redshift surveys, [arXiv:astro-ph/0307460v1]

[99] H.-J. Seo and D. J. Eisenstein, *ApJ, 633, 575, 2005* [arXiv:astro-ph/0507338]

[100] Seo H-J, Eisenstein D.J., Improved Forecasts for the Baryon Acoustic Oscillations and Cosmological Distance Scale *ApJ, 665, 14, 2007* [arXiv:astro-ph/0701079]

[101] Sheth R. K. and Diaferio A., Peculiar velocities of galaxies and clusters *Mon. Not. R. Astron. Soc., 322, 901, 2001* [arXiv:astro-ph/0009166]

[102] Sheth R. K. and Tormen G., Large-scale bias and the peak background split *Mon. Not. R. Astron. Soc., 308, 119, 1999* [arXiv:astro-ph/9901122]

[103] Smith C. et al., Linearizing the Observed Power Spectrum. *Mon. Not. R. Astron. Soc., 297, 910S,1998* [arXiv:astro-ph/9702099]

[104] Smith R.E. et al., Stable clustering, the halo model and nonlinear cosmological power spectra *Mon. Not. R. Astron. Soc.,341:1311S,2003* [arXiv:astro-ph/0207664]

[105] Smith R.E. et al., Scale dependence of halo and galaxy bias: Effects in real space *Phys. Rev. D, 75, 063512, 2007* [arXiv:astro-ph/0609547]

[106] Smoot G.F. et al., Structure in the COBE differential microwave radiometer first-year maps *ApJ, 396L: 1, 1992*

[107] Springel V. et al., Simulating the joint evolution of quasars, galaxies and their large-scale distribution *Nature, 435: p. 629 ,2005* [arXiv:astro-ph/0504097]

[108] Tadros et al., Spherical harmonic analysis of the PSCz galaxy catalogue: redshift distortions and the real-space power spectrum *Mon. Not. R. Astron. Soc.,305, 527, 1999* [arXiv:astro-ph/9901351]

[109] Takahashi et al., Simulations of barionic acoustic oscillations II: covariance matrix of the matter power spectrum. *ApJ,700: p.479,2009* [arXiv:0902.0371]

[110] Tegmark M., Measuring Cosmological Parameters with Galaxy Surveys *Phys. Rev. Lett., 79, 3806, 1997* [arXiv:astro-ph/9706198]

[111] Tegmark M. et al., *ApJ., 480, 22, 1997*

[112] Tegmark M. et al., The three-dimensional power spectrum of galaxies from the Sloan Digital Sky Survey *ApJ, 606: p.702,2004* [arXiv:astro-ph/0310725]

[113] Tinker J.L. et al. On the Mass-to-Light Ratio of Large-Scale Structure *ApJ, 631, 41, 2005* [arXiv:astro-ph/0411777]

[114] Verde L., Statistical methods in cosmology, *arXiv:0911.3105v1[astro-ph.CO]*

[115] Verde L. et al., *Mon. Not. R. Astron. Soc., 335, 432, 2002* [arXiv:astro-ph/0112161]

[116] Westover, M. 2007, PhD thesis, Harvard University

[117] White M. et al., Forecasting cosmological constraints from redshift surveys *Mon. Not. R. Astron. Soc., 397: p.1348,2009* [arXiv:0810.1518]

[118] Yahil The Local Group − The solar motion relative to its centroid *ApJ, 217, 903, 1977*

[119] Yamamoto K. et al., A Measurement of the Quadrupole Power Spectrum in the Clustering of the 2dF QSO Survey *Publ. Astron. Soc. Jap. 58, 93, 2006* [arXiv:astro-ph/0505115]

[120] Zehavi et al., The Luminosity and Color Dependence of the Galaxy Correlation Function *ApJ, 630, 1,2005* [arXiv:astro-ph/0408569]

[121] Zehavi et al., Galaxy Clustering in the completed SDSS Redshift Survey: the dependence on colour and luminosity. *ApJ, 736, 59, 2011* [arXiv:1005.2413]

[122] Zeldovich Y.B., Gravitational instability: An approximate theory for large density perturbations. *Astr. Astrophys., 5, 84, 1970*

[123] Zheng et al., Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies *ApJ, 633, 791,2005* [arXiv:astro-ph/0408564]

[124] Zwicky, F. , Die Rotverschiebung von extragalaktischen Nebeln, *Helvetica Physica Acta 6: 110–127 (1933)*

[125] http://tdc-www.cfa.harvard.edu/2mrs/

[126] http://astro.dur.ac.uk/∼40qra/lightcones/EUCLID/

[127] camb.info/

[128] http://www.roe.ac.uk/ jap/haloes/