AIX-MARSEILLE UNIVERSITE

ECOLE DOCTORALE PHYSIQUE ET SCIENCES DE LA MATIERE

T H E S E

en vue d'obtenir le titre de

Docteur en Sciences

de l'Université de Aix-Marseille

Mention : Physique Théorique, Physique Mathématique, Physique des Particules et Astrophysique

Spécialité : Cosmologie

Julien BEL

Les moments cumulant d'ordre supérieur à deux points des champs cosmologiques: propriétés théoriques et applications

Thèse dirigée par Christian Marinoni

préparée au Centre de Physique Théorique (Marseille)

Président :	Pierre Taxil
Rapporteurs :	István Szapudi
	Enrique Gaztañaga
Directeur :	Christian Marinoni
Examinateurs :	Francis Bernardeau
	Luigi Guzzo
	Yannick Mellier

Abstract

Cosmology is a relatively young discipline. Originally a branch of astronomy, it is nowaday the cross road where fundamental physics, particle physics and astrophysics meet. This interdisciplinary character has contributed to an acceleration progress in the field. The past two decades have witnessed a great leap forward in our understanding of the universe as a whole and of the origin and evolution of the structures it contains. Some important questions have been settled such as the precise value of the present day expansion rate H_0 of the universe and the geometry of the spatial sections of the universe. Other questions, such as when and how cosmic structures formed are on the way of being definitively addressed. At the same time, new questions have been raised concerning the physical nature of inflation and of dark energy, the mysterious causes of the accelerated expansion of the universe. Many questions, however, still remain without any convincing answer. Among them, what is dark matter? Why is it escaping detection ?

Undoubtedly, what makes cosmology so exciting and challenging is that it is becoming a precise, data-driven field. Many independent cosmic probes seem to point toward a coherent, coarse grained picture of the universe, usually called the concordance model of cosmology. Over the next decade, a host of information from new telescopes and satellites (LOFAR, SKA, Planck, LSST, BigBoss, EUCLID, LIGO, LISA) will be available at all wavelengths. It is not improbable, therefore, that we will witness some unexpected observational discoveries triggering new shifts in theoretical paradigms.

At the same time, cosmology is becoming an increasingly non-linear field. Thanks to the increasing computing power, numerical simulations of astrophysical processes and of cosmic structure formations will become ever more sophisticated and will provide us with a better understanding of the dark matter physics and of also the physics of the dark ages. Furthermore, the analytical approaches of fundamental physics will help to work out more quantitative predictions about high energy phenomena in the early universe, or about the skeleton of gravitation at both micro (quantum) and macro (cosmic) scales.

At present, despite formidable progress in data collection from both ground and space, and despite the development of performant cosmic probes and of sophisticated error analysis schemes, observational constraints on fundamental cosmological quantities are still rather loose. As a result of the imprecision on the amplitude on the dark matter and dark energy densities or on the dark energy equation of state, we cannot yet discriminate between the (many) competing theories proposed to interpret the observations. The overarching questions are is it necessary to refine the standard description of the universe by adding new features, what can be said about them, and how they may be related to the actual view of the uinverse. Cosmologists are therefore in a state of theoretical agnosticism, or if one prefers, the standard model of cosmology is characterized by too many

degrees of freedoms.

The main line of this thesis is that our best chances of finding and characterizing the essential ingredients of a well grounded cosmological model is by enlarging the arsenal of methods with which we can hunt for new physics. While it is of paramount importance to continue to refine, de-bias and improve, the testing strategies that contributed to establish the concordance model, it is also crucial to challenge, with new methods, all the sectors of the current cosmological paradigm. This thesis, therefore, adresses the challenge of developing new and performant cosmic probes that aim at optimizing the scientific output of future large redshift surveys. The goal is twofold. From the theoretical side, I aim at developing new testing strategies that are minimally (if not at all) affected by astrophysical uncertainties or by not fully motivated phenomenological models. This will make cosmological interpretations easier and safer. From the observational side, the goal is to gauge the performances of the proposed strategies using current, state of the art, redshift data, and to demonstrate their potential for the future large cosmological missions such as BigBOSS and EUCLID.

A more detailed description of the content of my work is the following. In chapter I, after reviewing the essential theoretical and observational bases of the concordance model of cosmology, I introduce the theoretical background necessary to characterize the clustering of matter on large cosmic scales. I introduce my notations and briefly overview the current understanding of the hierarchical clustering of cosmic matter, both from a dynamical and a statistical point of view. Then, I discuss the high-order statistical descriptors of the density perturbation field, and, in particular, the quantities that will be extensively studied in this thesis, i.e. the one- and two-point cumulants of the cosmic density fields. I finally conclude by discussing the biasing issue, the puzzling problem of relating the spatial distributions of dark matter and luminous baryons on large cosmic scales.

I present in Chapter 2 an original way for testing the reliability of a fundamental facet of the Cosmological Principle, i.e. that we are not privileged observers of the universe. An often explored approach for addressing some cosmological puzzles, in fact, is to question the validity of the Cosmological Principle, i.e. the statement that the universe looks homogeneous and isotropic to any comoving observers. The goal of my analysis is thus to test the coherence of this founding assumption by identifying the scale above which the universe appears isotropic to any comoving observers. I describe how I measured the scale of every-where isotropy using luminous red galaxies from the Sloan Digital Sky Survey. I finally show that the galaxy distribution can be considered as being everywhere isotropic on scales larger than $150h^{-1}$Mpc. The content of this section is presented in the paper "*The scale of cosmic Isotropy*", by **Marinoni, Bel & Buzzi**, 2012, JCAP, arXiv:1205.3309

I present, in Chapter 3, an original way of estimating the root mean square of matter fluctuations in spheres of radius R without any prior knowledge on the power spectrum of matter fluctuations. To this purpose I first show that, if the bias between galaxies and

matter can be described by a non-linear, local, transformation, then the 2-point reduced cumulants of the galaxy density field ($C_{g,nm}$) conserve the same hierarchical properties as the corresponding mass statistics C_{nm}. I use this result to show how one can recover the biasing function by computing the one-point and two-point reduced cumulants of the matter and galaxy density fields. Finally, I derive a formula that allows to estimate the real space amplitude of the matter fluctuation (on a linear scale R) directly from redshift space data. The robustness of the formalism as well as its performances are tested using two different numerical simulations of the large scale structure of the universe (LasDamas and Horizon). I show that it is possible to recover the *rms* amplitude of the matter fluctuations hidden in the simulations. Some of the material presented in this chapter is published in the paper "*Second-order matter fluctuations via higher order galaxy correlators*" by **Bel & Marinoni** 2012, MNRAS, 424, 971

I present, in Chapter 4, a simple (probably the simplest) cosmological test that one can think of in order to fix the values of fundamental cosmological parameters. Simplicity stems from the fact that this cosmic test, while exploiting the clustering properties of galaxies, is specifically engineered to be independent of linear bias, of redshift distortions and of the non-linear phenomenological modeling of the matter power spectrum. Moreover, it does not require the calibration of any standard rod and it is free from any look-back time (evolutionary) effects. This cosmic probe is constructed as the ratio η between the one and two-point, second-order moments of the smoothed galaxy density field. I present the cosmological constraints on the reduced matter density parameter Ω_m obtained by analyzing with this technique the luminous red galaxy sample of the Sloan Digital Sky Survey. By taking a strong gaussian prior for H_0 (from HST measurements) and for $\Omega_b h^2$ (from Big Bang nucleo-synthesis measurements) and a weak prior for the spectral index n_s, I constrain the value of Ω_m with a relative precision of 11%. This result is even more remarkable if one considers that it has been obtained without any CMB prior, and without assuming neither the flatness of the universe nor the hypothesis that dark energy is essentially the Einstein cosmological constant (that is without assuming that the dark energy equation of state is $w = -1$). The material of this section is published in the paper "*Determination of the abundance of cosmic matter via the cell count moments of the galaxy distribution*", by **Bel & Marinoni**, 2012, submitted.

Contents

Remerciements

Chapter 1

Introduction

1.1 A quick overview of the standard model of cosmology

In this introductory chapter I will briefly review our current understanding about the structure of the cosmic space-time and the evolution of its matter and energy content. I will introduce the essential elements of the standard theory which models the kinematical and dynamical properties of the cosmic metric and I will discuss how perturbations of this smooth background grow, evolve and eventually turn into the large-scale structures that populate the universe. I will also present the observational evidences that support and confirm theoretical predictions, and I will discuss the major challenging issues that the model is facing at present.

1.1.1 Standard cosmological model

The description of the universe is built on two independent major assumptions:

- the distribution of cosmic matter and energy appears homogeneous and isotropic to a class of fundamental cosmic observers, called *comoving* observers. These are defined as those particular set of observers, freely falling within the cosmic gravitational field, that move with the average velocity of matter in their respective neighborhoods. According to this cosmological principle (CP), there are no privileged positions in the universe.

- the general relativistic theory of gravitation (GR) still holds and applies on large cosmic scale.

As a consequence, the universe is modeled as a 4D manifold (with three spatial and one temporal dimensions), it is possible to define a universal time which ticks the same way for every comoving observer, and the separation between 2 infinitesimally closed space-time events is given by the Robertson & Walker (Robertson, 1929; Walker, 1936) line element

$$\mathrm{d}s^2 = (c\mathrm{d}t)^2 - a(t)\left\{\mathrm{d}\chi^2 + S_k^2(\chi)\left(\mathrm{d}\theta^2 + \sin\theta\mathrm{d}\phi^2\right)\right\}, \tag{1.1}$$

where c is the speed of light in vacuum, and t is the cosmic time. Note that spatial points (or equivalently comoving matter particles) are labeled using spherical comoving coordinates (χ, θ, ϕ): χ represents the radial geodesics while θ and ϕ are the two degrees of freedom on sky. The spatial configurations compatible with the symmetries implied by the PC are only 3 and they are labeled by the index k. A flat universe ($k = 0$), in which case $S_0(\chi) = \chi$, a positively(/negatively) curved universe ($k = +1(/-1)$), in which case $S_k(\chi) = sin(\chi)(/\sinh(\chi))$. Note, also, that comoving (i.e. time independent) distances are transformed into physical distance by the scale factor $a(t)$. That is, the physical separation $\vec{r}$ at the same cosmic instant t, between two space positions of comoving separation $\vec{x}$ at the same cosmic instant t, is

$$\vec{r} = a\vec{x}.$$

By taking its time derivative we deduce that the relative velocity between comoving observers is

$$\vec{V} = \dot{a}\vec{x},$$

where the dot indicates a derivative with respect to cosmic time. From the two expressions above it is straightforward to deduce the Hubble (1929) law

$$\vec{V} = H\vec{r},$$

where $H \equiv \frac{\dot{a}}{a}$ is a time dependent parameter (Hubble parameter) which contains information about the cosmic expansion rate. Two comoving observers separated by the spatial distance $|\vec{r}|$ have a relative velocity of $H|\vec{r}|$, and their kinematics (expansion/contraction) depends on the sign of the parameter H (positive/negative). Since we cannot measure cosmic separations via a ruler, distances are not cosmological observables and can only be defined via theoretical models. Among the various possible operational definitions (Hogg, 1999) , of particular importance is the so called radial geodesic distance between two comoving coordinates. In fact, it can be demonstrated that other different operational definitions can be simply deduced from the knowledge of this fundamental one. The comoving separation χ is computed along the photon path that connects the emitter and the receiver. By imposing that $ds = 0$ in 1.1, and by setting $\mathrm{d}\theta = \mathrm{d}\phi = 0$, one obtains

$$\chi \equiv c\int_{t_e}^{t_r} \frac{\mathrm{d}t}{a(t)}, \tag{1.2}$$

where t_e and t_r are the emission and reception time respectively. An immediate consequence of this relation is that the wavelength of the traveling photon is stretched during the propagation in an expanding universe. The cosmological shift in wavelength between emission λ_e and reception λ_0, parameterized via the redshift $z \equiv (\lambda_r - \lambda_e)/\lambda_e$, is in fact

linked to the cosmic scale factor at the emission t_e and reception time t_0 via the relation

$$\frac{a(t_0)}{a(t)} = 1 + z. \tag{1.3}$$

where a_0 is the scale factor today, $a_0 \equiv a(t_0)$.

It has been shown by Slipher (1912) that galaxies appear to show a systematic redshift ($z > 0$) while Hubble (1929) found that the galactic redshift correlates linearly with the distance of the emitter $z \propto d$. This last remarkable observation confirms a prediction of equation 1.3. If $t \sim t_0$, i.e. for small separations between the source and the receiver,

$$d = \frac{cz}{H(t_0)}. \tag{1.4}$$

Within the standard model of cosmology, the redshift is therefore a proxy for distance. We stress that since the Hubble constant H_0 was poorly constrained in the past, with values ranging in the interval $(50, 100)$ km s^{-1}Mpc^{-1}, even in the era of 'precision cosmology, it is still common practice to express the value of this constant via the a-dimensional Hubble parameter h, defined as $H_0 = 100h$ km s^{-1}Mpc^{-1}. As a consequence, cosmological distances are expressed in units of h^{-1}Mpc.

Since redshift, as opposite to photon travel time, is a directly measurable quantity, equation 1.2 can be re-expressed in a more observationally friendly way as

$$a_0\chi \equiv \frac{c}{H_0}\int_0^z \frac{\mathrm{d}x}{E(x)}, \tag{1.5}$$

where, by definition, $H_0 \equiv H(t_0)$ and where $E(z) \equiv H(z)/H_0$ is called the evolution parameter. The distance-redshift relation (equ. 1.5) shows that if we know the Hubble rate H at all the intermediate redshift between t_0 (today) and t_e we can infer the distance to any specific extragalactic object. Despite it is possible to try to constrain the scaling of $H(z)$ directly via observations (Percival et al., 2010), it is standard practice to let theory to predict how H evolves as a function of redshift. This second possibility follows from assuming that the dynamics of the universe as a whole is ruled by gravity, and that the strength of this field is correctly described by the Einstein's field equations (hereafter EFE, Einstein, 1915)

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{1.6}$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, $g_{\mu\nu}$ is the metric tensor, Λ is the cosmological constant, G is Newton's gravitational constant, and, finally, $T_{\mu\nu}$ is the stress energy tensor of the various cosmic species. The equations (1.1.1) are used to link the space-time properties (LHS) to a given distribution of sources(RHS). Originally, the cosmological constant was introduced by Einstein in order to allow a static universe ($H = 0$) to contain matter. This term, however, was dropped out from as soon as Hubble found

convincing evidences about the expansion of the universe. Now days, this parameter has found again its place in the EFE. It is the simplest ingredient that allows explaining the fact that the late time expansion of the universe is accelerated.

By exploiting the CP and the EFE, and by assuming as well that the cosmic species behave effectively as perfect fluids, Friedmann (1922) (originally) and Lemaitre (1931) (independently) were able to show that the scale factor of the universe evolves according to the following system of equations

$$\begin{cases} \left(\frac{\dot{a}}{a}\right)^2 + k\frac{c^2}{a^2} = \frac{8\pi G}{3}\sum_i \rho_i \\ \\ 2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + k\frac{c^2}{a^2} = -8\pi G\sum_i \frac{P_i}{c^2} \end{cases}, \tag{1.7}$$

where ρ_i and P_i are respectively the mass density and pressure of the different cosmic components (e.g. radiation, relativistic particles, cold matter...).

It is standard practice, to recast the first Friedmann equation into a more telling form

$$1 - \Omega_k(t) = \sum_i \Omega_i(t). \tag{1.8}$$

where $\Omega_i \equiv 8\pi G\rho_i/(3H^2)$ are the reduced density parameters associated to the specie i and $\Omega_k \equiv -kc^2/(aH)^2$ is the effective reduced density associated to the curvature of space. The meaning of this equation can then be easily spelled out: *"let me know about the energy content of the Universe, and I will tell you about its geometry"*

Combining the first and second Friedmann & Lemaitre (FL) equations, and assuming that different species do not interact nor exchange energy, one finds two more equations, that is the relativistic analog of a continuity equation for each cosmic component

$$\dot{\rho}_i + 3H(t)\left(1 + \frac{P_i}{\rho_i c^2}\right)\rho_i = 0. \tag{1.9}$$

and the acceleration or Raychaudhuri equation

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_i + 3\frac{P_i}{c^2}). \tag{1.10}$$

By parameterizing the equation of state of a given fluid as $w_i \equiv P_i/(\rho_i c^2)$, the differential equation (1.9) can be solved to give the evolution of density as a function of the redshift

$$\rho_i(z) = \rho_{i,0} e^{3\int_0^z (1+w_i)\mathrm{d}\ln(1+z)}, \tag{1.11}$$

where $\rho_{i,0}$ refers to the present day density. If the equation of state parameter w_i can be assumed to be constant (as in the case of cold matter, relativistic particles or radiation) then the time evolution of the reduced density parameters is simply

$$\Omega_i = \Omega_{i,0}\,(1+z)^{3(1+w_i)}\,. \tag{1.12}$$

Since matter on cosmological scales can be safely considered as being collision-less, $w_m = 0$. On the other hand, a photon fluid has a non negligible pressure and it is characterized by a constant equation of state $w_r = 1/3$. Note that the cosmological constant can be effectively interpreted as being the energy density associated to a specie (the vacuum) with a peculiar equation of state, i.e. $w = -1$. To this purpose it is enough to change place (from the LHS to the RHS) to the Λ term in eq. 1.1.1 and define $\rho_\Lambda \equiv \Lambda c^2/(8\pi G)$. This simple algebraic operation is however rich in physical consequences, the implications of which will be soon addressed.

We are now in the position to write down the expected expansion rate of the universe as function of its matter-energy content. This expression reads

$$H(z) = H_0\sqrt{\Omega_{k,0}(1+z)^2 + \Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4 + \Omega_{\Lambda,0}}, \tag{1.13}$$

and its is completely determined once the abundance of the different cosmic species Ω_i is known. Performant cosmological probes (that will be described in detail in the next session) gives us precise information about the value of these fundamental cosmological parameters and help us to draw important physical conclusions about the structure and dynamics of the universe. In particular, current observations indicate that

- the present day value of the Hubble constant is positive and small ($H_0 = 73.8 \pm 2.4$ km/s/Mpc), that is the universe is expanding and fairly old (~ 13.5 Gyrs).

- ordinary baryonic matter, i.e. massive standard model particles, is a minority ($\sim 1/6$) of all forms of matter ($\Omega_b = (0.021 \pm 0.001)h^{-2}$). This means that the dynamics of the universe is essentially driven by dark matter, that is particles that interact with the rest of the world only gravitationally. Although the first evidences about this non-standard cosmic component date back to 1933 (Zwicky, 1933), this cosmological component has yet to be directly detected in laboratory.

- matter is a minority of all the form of energy $\Omega_m = 0.247 \pm 0.026$ (Komatsu et al., 2011). This unexpected result is the consequence of recent observations indicating that the expansion of the universe, instead of being decelerated, as expected if gravity is sourced only by matter, is in fact accelerating. A negative equation of state (and more precisely an equation of state parameter $w < -1/3$) is what is needed to counteract gravitational attraction and force the universe to accelerate (see equation 1.10). We will simply call 'dark energy' any eventual physical component that is

characterized by such a peculiar equation of state parameter. Note that the cosmological constant is a serious candidate to explain current (although low resolution) observations. Most recent estimates indicate that $\Omega_\Lambda = 0.716 \pm 0.055$ (Spergel et al., 2007).

- the geometry of the universe is flat. Observations of the cosmic microwave background (CMB) indicate that the universe, though not eternal since it originated from a singularity in the finite past, is spatially infinite. We now know that even if we wait for an infinite time, we cannot observe the whole of the universe. In other terms, the universe has a finite event horizon.

The ensemble of these facts constitutes what is called the concordance model of cosmology. The adjective concordance emphasized that this picture is not the outcome of a single cosmic probe, however precise might it be, but rather the coherent result of independent and complementary exploration strategies.

1.1.2 Observational bases of the standard cosmological model

Nearly 40 years ago Cosmology turned into a ' respectable' science because cosmologists where able to measure the amplitude of Ω_r via the discovery of the Cosmic Microwave Background (CMB). Ten years ago, it turned into a ' precise' science, because of the effective possibility of using the universe as a giant laboratory where to constrain accurately the value of H_0, Ω_m, and Ω_Λ. Since the goal of this thesis is to present and apply new techniques to constrain these parameters with even further precision, in this section I will give a brief overview of some performant observational probes that helped us to establish the concordance model.

Measuring H_0 using the Hubble diagram of Cepheids and SNIa

Cepheids are pulsating stars whose absolute luminosity oscillates with a proper periodicity. It has been found by Leavitt & Pickering (1912) that the pulsation period is related to the intrinsic luminosity of the star. The process of finding characteristic observables (in this case time intervals) with a functional dependence on an intrinsic property of an object (in this case the absolute luminosity) is called standardization. Cepheids, in particular, are standard candles: once the intrinsic luminosity (or in astronomical term the absolute magnitude M) of an object is known, it is possible to estimate via the Period-Luminosity diagram the distance of any other Cepheids by simply measuring the apparent magnitude m via the relation

$$m - M = 5\log(d_L) + 25.$$

This relation between apparent magnitudes and distances is called Hubble diagram. The label L makes clear that this particular operational definition of distance rests on

energetic arguments, and loses its elementary geometrical meaning if the property of space-time, as it is the case in cosmology, are non trivial.

Nearby Cepheids were immediately recognized (see Hubble, 1929) as the ideal tool to fix the amplitude of the present day value of the Hubble parameter H_0 via the relation (1.4). Cepheids however are not visible at large distances from Earth. To probe distant cosmic regions a different kind of standard candles are used: the Supernovae Ia. A supernova is the results of the explosion of a star. As a consequence it is a transient and extremely bright celestial phenomenon. For instance, a type Ia supernova has a typical magnitude $M \sim -19$ which corresponds to the typical luminosity emited by all the stars trapped in a spiral galaxy.

There are two types of supernovae, Type I and Type II. This classification identifies two different explosion mechanisms. Type II supernovae are generated by the core-collapse of white dwarf stars whose mass is larger than a given threshold, the Chandrasekhar's mass ($\sim 1.4 M_\odot$). Type Ia supernovae, on the contrary, are possible only in binary star systems, in which a white dwarf accretes matter from the companion. The explosion, which takes place even before the white dwarf reaches the Chandrasekhar's mass, is typically brighter than in Supernovae II.

Supernovae Ia explode when they have nearly the same mass and chemical composition, a theoretical hypothesis which is well verified in objects with redshift lower then 0.1. They should therefore be characterized by the same maximum luminosity. As a consequence they can be used as standard candles, and because of their brightness, they allow measuring distances of hosting galaxies up to redshifts as high as $z = 2$. Since the light emitted from far away sources propagates in a curved and expanding background, the relation between the luminosity distance and redshift is not as simple as that given by relation 1.4. A more careful modeling of the energetic exchange between a source and a receiver is needed in order to obtain an appropriate definition of the luminosity distance.

The flux ϕ measured by a detector represents the energy $\mathrm{d}E_0$ collected per unit surface $\mathrm{d}^2 s_0$, and per unit exposure time $\mathrm{d}t_0$ at the observer position

$$\phi_{\nu_0} = \frac{\mathrm{d}E_0}{\mathrm{d}^2 s_0 \mathrm{d}t_0 \mathrm{d}\nu_0}.$$

By considering how time intervals and energy transform between the source and the receiver, one obtains that the previous relation can be recast in an Euclidean form ($\phi_{\nu_0} = L/(4\pi d_L^2)$ only if one defines the luminosity distance as

$$d_L = (1+z) a_0 S_k(\chi).$$

In order to obtain an accurate measurement of the Hubble constant, Freedman et al. (2001) used Cepheids to calibrate the luminosity of near supernovae Ia (hosted in galaxies where Cepheids are visible). By collecting a large sample of Cepheids and supernovae

Ia (with $z < 0.1$) using data from the Hubble Space Telescope (HST), they were able to provide a measurement of H_0 with a relative precision of 11% ($H_0 = 72 \pm 8$ km s^{-1}Mpc^{-1}). Today, the most accurate and precise estimate is $H_0 = 73.8 \pm 2.4$ km s^{-1}Mpc^{-1} (Riess et al., 2011) obtained by exploiting a total of 253 supernovae (calibrated using 600 Cepheids located in 8 galaxies hosting type Ia supernovae).

Measuring Ω_m using the baryon acoustic oscillations imprinted in the CMB and in the large-scale distribution of galaxies

The hot Big Bang theory assumes that the early universe was dominated by radiation and was sufficiently dense and hot to force free electrons and photons to couple via the Thomson scattering mechanism. As a result the baryonic component was characterized by a non-negligible pressure ($w \neq 0$). The antagonist effect of radiation pressure and gravity caused baryon perturbation to oscillate, generating the so called baryon acoustic oscillations (BAO). The resulting acoustic waves could propagate in the primordial plasma only up to the recombination epoch. At this epoch, indeed, the radiation density was diluted enough to allow ionized atoms to recombine and photons to free-stream. Once the pressure in the baryon fluid drops, the overdensity region, that is the sound wave, cannot propagate anymore. The overdensity excess is frozen in the supporting medium (i.e. the large scale distribution of baryonic matter). The comoving radius of this overdense spherical cell is called the comoving sound horizon, because it corresponds to the comoving distance a sound wave could travel from the end of inflation to the recombination epoch. This characteristic scale imprinted in the CMB map and in the distribution of galaxies provides a cosmic standard ruler. In principle, we can exploit it to fix the value of fundamental cosmological parameters.

In all generality a sufficiently small redshift difference Δz is related to a radial comoving size $s_{//}$ (along the line of sight) via the relation

$$s_{//} \simeq \frac{c}{H}\Delta z.$$

On the other hand, an angular separation $\Delta\theta$ can be related to a comoving transversal size $s_\perp$ (perpendicular to the line of sight) via

$$s_\perp \simeq \Delta\theta a_0 S_k(\chi),$$

where $\Delta\theta$ is assumed to be small. The mapping between comoving size and observed size of an object positioned at a given redshift is illustrated in figure (1.1). Since the comoving sound horizon at the decoupling epoch (r_s) can be measured with the CMB, one can deduce the characteristic scale $s_{//}$ and $s_\perp$ by using the previous equations. Since $r_{//} = r_\perp = r_s$ and since Δz and $\Delta\theta$ can be measured separately using the temperature map of the CMB and the spatial distribution of galaxies, it follows that the amplitude of

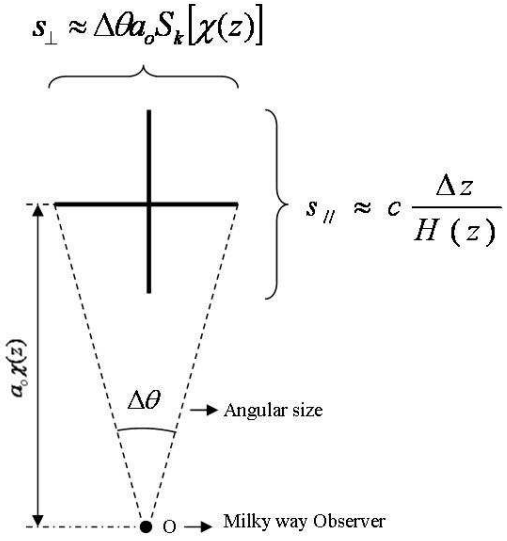

Figure 1.1: Using the BAO scale as a standard ruler. The thick cross represents the radial and transversal comoving size of spherical shell centered at a redshift z.

H and $a_0 S_k(\chi)$ can be estimated at different epochs.

In practice, the over-dense spherical shell caused by BAO is detectable via the statistical analysis of the large-scale distribution of galaxies. As a matter of fact, we expect to detect an excess of probability of finding galaxies separated by a distance corresponding to the BAO scale. In a general way, this excess probability (with respect to a random distribution) can be estimated using the two-point correlation function $\xi(r)$. Since the clustering of galaxies is driven by gravity, the overall shape of the two-point correlation function can be predicted by modeling the evolution of density perturbations in a homogeneous universe. This way, once a cosmological background is assumed, the overall shape of the two-point function of the galaxy distribution can be compared to the theoretical predictions of perturbation theory (or of N-body numerical experiments). Results obtained by implementing this approach (Eisenstein et al., 2005) show that the BAO probe is mostly sensitive to $\Omega_m h^2$ and $\Omega_b h^2$. However by combining the BAO and the CMB probes one can resolve degeneracies and put strong constraints on the cosmological parameters Ω_m, H_0 and Ω_Λ of a ΛCDM universe (Komatsu et al., 2011)

Measuring Ω_Λ using the Hubble diagram of SN Ia

While local SNIa are useful to fix the present day value of the cosmic expansion rate, at high redshift, they can be used to gain insight into other cosmological parameters, and in particular they can give us information about the abundance and nature of the cosmic components. At high redshift the Hubble diagram is in fact sensitive to all the fundamental

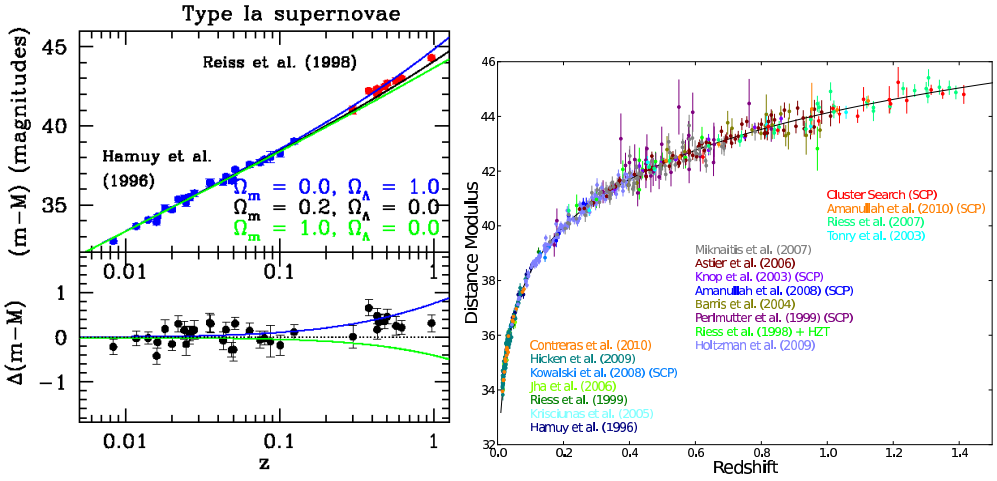

Figure 1.2: Collection of supernovae data. *left:* The upper panel shows the distance modulus as a function of redshift. Data from Riess et al. (1998) are compared to different theoretical predictions corresponding to different cosmological models. The lower panel shows the residual between the models and the measurements. *right:* One of the most recent collection of supernovae Ia data, as well as the corresponding luminosity distance (solid line) corresponding to a flat universe with $\Omega_\Lambda = 0.7$.

cosmological parameters that enters into the expression of the Hubble parameter (see equation 1.13), and not only to its normalization H_0. In particular the distance modulus $\mu = m - M$ in the Hubble diagram is

$$\mu = 25 + 5\log\left((1+z)a_0 S_k[\chi_{\vec{\Omega}}(z)]\right).$$

Once the redshift and the distance modulus of high z SNIa is known, one can then use this relation to constrain the amplitude of relevant parameters such as Ω_m, Ω_Λ, and test if the equation of state parameter w_Λ is compatible with being equal to -1. Riess et al. (1998) and Perlmutter et al. (1999) won the Nobel Prize in 2011 for the discovery of the acceleration of the expansion of the universe. In particular, the Hubble diagram of Riess et al. (1998), as well as a more recent version obtained with a larger set of data, are displayed in figure (1.2) and compared to the predictions of different cosmological models. These measurements, when combined with results from other probes (BAO and CMB) clearly indicate that the best fitting model is a cosmology dominated by dark energy. This is clearly seen in figure (1.3) where the joint constraints on the parameters (Ω_m,Ω_Λ) are plotted. Data are best fitted by models characterized by a positive value of the cosmological constant and a low (i.e. < 1) matter density. If one assumes that the universe is flat ($\Omega_k = 0$) but relaxes the assumption that the dark energy component is the cosmological constant, i.e. no prior is taken of w, then again data confirms that the most likely explanation for the dark energy phenomenon is a cosmological constant.

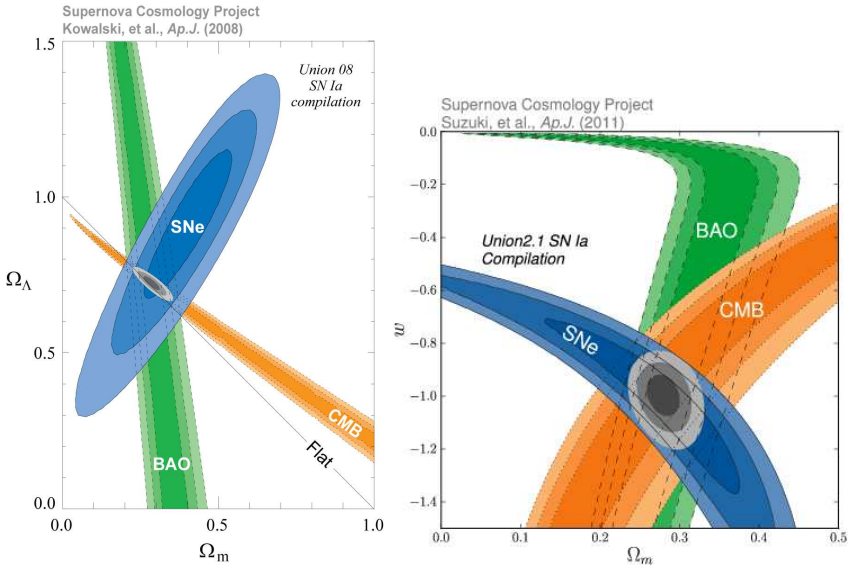

Figure 1.3: *left:* The concordance model of cosmology, joint constraint on the dark energy and matter density parameters. The joint analysis of CMB, BAO and Supernovae data prefer a flat universe with a dominant component of dark energy. *right:* The joint constraint on the matter density parameter Ω_m and the equation of state of dark energy when it is assumed that universe is flat ($\Omega_k = 0$)

1.1.3 Beyond the standard model

Although supported by a large variety of independent observations, the concordance model of cosmology appears to be poorly motivated from the point of view of fundamental physics. From one side, nearly 1/5 of the energy of the universe is contributed by dark matter particles for which we still lack any secure microscopic characterization and any direct laboratory evidences. From the other side, the simple explanation of the late time cosmic acceleration in terms of the Einstein's cosmological constant Λ appears to be fundamentally problematic.

The cosmological constant Λ appearing in the Einstein's field equations is not, as we would usually expect, a coupling constant. Moreover, its value almost coincides with the amplitude of the dark matter density parameter. This is not expected at all since theory predicts different time evolution histories for these two cosmological parameters.

An alternative explanation of the cosmological constant comes from quantum physics. It has been shown by Zel'dovich (1968) that a contribution of the vacuum energy to the stress energy tensor induces an accelerated expansion identical to that resulting from the cosmological constant. Quantum physics shows that the lowest energy level of a quantum system is not null, and it gives a contribution to the stress energy tensor of the form $-Vg_{\mu\nu}$ where V is the potential associated to the vacuum energy and $g_{\mu\nu}$ is the metric tensor. As the classical gravity description is supposed to break dawn at the Planck

energy density level ρ_{pl}, we can reasonably conclude that the vacuum energy density ρ_V is expected to be of the same order. On the other hand, the observational evidence about the accelerated expansion of the universe indicates that $\Omega_\Lambda \equiv (8\pi G/H_0^2)\rho_\Lambda$ is roughly of order one. This means that the energy density associated to the cosmological constant ρ_Λ is comparable with the critical density $\rho_{c,0} \equiv 3H_0^2/(8\pi G) = 2.77 \times 10^{11} h^2 M_\odot \mathrm{Mpc}^{-3}$. Since $\rho_{pl} = 10^{123}\rho_{c,0}$, the theoretical expectation for the vacuum energy density exceeds by 123 order of magnitudes the effective energy density required to explain the actual accelerated expansion of the universe.

Although the cosmological constant offers a phenomenologically viable description of the universe, these interpretational problems forced cosmologists to explore other possible causes of the late time acceleration of the universe. A minimally coupled scalar field in the primordial universe (e.g. Ratra & Peebles, 1988; Wetterich, 1988; Caldwell, Dave & Steinhardt, 1998; Ferreira & Joyce, 1998a,b; Zlatev, Wang & Steinhardt, 1999), also called quintessence, might be an appealing alternative to the cosmological constant scenario. Contrary to the vacuum energy case, the scalar field ϕ is not stationary ($\dot{\phi} \neq 0$). By supposing that its dynamics is similar to that of a perfect fluid, we can effectively describe quintessence as a field with density ρ_ϕ and a pressure P_ϕ given by

$$\rho_\phi c^2 = \frac{\dot{\phi}}{2} + V(\phi)$$

and

$$P_\phi = \frac{\dot{\phi}}{2} - V(\phi),$$

where the potential $V(\phi)$ defines the scalar field. As a result, quintessence is characterized by a time evolving, negative, equation of state w_ϕ as soon as the scalar field is slowly rolling, i.e. its kinetic energy is subdominant with respect to the potential energy. Since the stress energy tensor is conserved (cfr. 1.9) the scalar field satisfy the Klein Gordon equation

$$\ddot{\phi} + 3H\dot{\phi} + \frac{\partial V}{\partial \phi} = 0. \tag{1.14}$$

Interestingly, one can choose quintessence potentials so to avoid the coincidence problem (tracker models). For such models, Zlatev, Wang & Steinhardt (1999) have shown that, even if the energy density ρ_ϕ can vary by several order of magnitudes at the final stage of inflation, the overall evolution of ρ_ϕ converges (in a relatively short time) toward a constant asymptote.

More radical approaches consist in testing whether the Einstein Field Equations offers a satisfactory description of gravity on large cosmic scales. Such studies explore the viability of alternative gravitational laws or non minimal modifications of the Einstein's field equations. In this context, a class of hypothesis extensively analyzed are the so called

$f(R)$ models, constructed by substituting the Ricci scalar with a more general function $f(R)$ in the EFE. To recover the EFE on small, non cosmological scales, one requires that $f(R)$ goes as $R/2$ for small separations. Alternatively, massive gravity models are also proposed as possible solutions to the dark energy problem. In such models, which result from describing the full theory of gravity in more than 4 dimensions, the graviton has a mass (and therefore the gravitational interaction is not not anymore scale free). An often analyzed representative of this category is the DGP (Dvali, Gabadadze & Porrati, 2000) model, which modifies the equation for cosmic acceleration (1.10) as follows

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho_i - 3\frac{P_i}{c^2}) + H\frac{c}{r_c}.$$

The characteristic scale r_c (related to the mass of the graviton) mimics the action of a cosmological constant. It should be noted, however, that as in the case of $f(R)$ models, even the dynamics predicted by this proposed modification of gravity could be exactly mimicked by odescribing dark energy in terms of an opportunely chosen, minimally coupled, scalar field.

As proposed by (Chevallier & Polarski, 2001; Linder, 2003), lacking definitive and convincing evidence about the causes of the cosmic acceleration, it is convenient to parameterize our ignorance about the dark energy phenomenon by assuming that any cosmological models beyond the standard concordance model can be effectively described assuming that the universe is filled by a uniform perfect fluid X with a generic, time-evolving equation of state given by

$$w = w_o + w_a(1 - a/a_0).$$

Generically, the specific evolutionary pattern from early ($w = w_o + w_a$) to late ($w = w_o$) time should be enough to allow discriminating between alternative dark energy scenarios or at least constraining the free parameters of these proposed alternative theories. As a result, the acceleration equation (1.10) will take the general form

$$\frac{\ddot{a}}{a} = \frac{1}{2}\left\{-(1+3w)\Omega_X - \Omega_m\right\}.$$

Using the CPL parameterization, the expansion rate is expressed as

$$H(z) = H_0\sqrt{\Omega_{k,0}(1+z)^2 + \Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4 + \Omega_{X,0}f_X(z)}, \tag{1.15}$$

where

$$f_X(z) \equiv (1+z)^{3(1+w_o+w_a)}e^{-3w_a\frac{z}{1+z}}.$$

Finally, note that the radiation density term is sourced not only by photons, but by all the relativistic species present in the universe at a given epoch. The expression for the

radiation density parameter, taking into account the existence of N_{eff} species of massless neutrinos, is $\Omega_{r,0} = \Omega_{\gamma,0}(1 + 0.2271 N_{eff})$, where $\Omega_{\gamma,0}$ is the present day photon density parameter ($\Omega_{\gamma,0} = 2.469 \times 10^{-5} h^{-2}$).

1.2 Dynamics of cosmic structures

If we assume that cosmic space-time is filled with uniform fluids, we can treat a complex structure such as the universe as a simple system characterized by only 2 degrees of freedom: the scale factor $a(t)$ and the spatial curvature k. However the universe that we observe is far from displaying the uniformity postulated by the CP. The cosmos is highly structured, showing an hierarchy of masses, of clustering scales and of... interpretational problems. In the next session we will briefly present the theory that allows us to make sense of this complexity. Since the goal of this thesis is to present new cosmological tests based on the analysis of the inhomogeneous sections of the universe, in this section I take the opportunity to introduce the formalism and the basics theoretical instruments that are needed to understand the more technical chapters of this manuscript.

How is it possible to create the landscape of remarkable density contrasts that surrounds us starting from a primordial universe which is highly symmetric? An old and almost unchallenged theoretical idea is that large cosmic inhomogeneities (galaxies, clusters super clusters) are the result of the gravitational amplification of tiny, primordial density perturbations.

Cosmic microwave background measurements have shown that at the decoupling epoch ($z_{dec} \sim 1100$, the epoch at which photons were released by the hot primordial plasma and free-streamed towards us) the amplitude of baryonic density anisotropies $\delta = d\rho/\rho$ was nearly 10^{-5}. The smallness of these departures from uniformity is due to the fact that before decoupling, when radiation and matter where strongly interacting via Thomson scattering, perturbations in the baryonic component of the plasma could not grow. This is because the gravitational attraction within primordial density inhomogeneities was counteracted by the enormous pressure exerted by the hot primordial radiation field.

If we assume that only baryonic matter populate the universe, then standard theory is unable to explain why we observe non linear structures, that is fluctuations with amplitude $\delta > 1$ in the present day universe. Simply, 13 Gyrs is not enough time for gravity to amplify these primordial fluctuations by at least five order of magnitudes. This issue can be bypassed by hypothesizing the existence of a dark matter component which does not interact electromagnetically with photons. The dynamics of this exotic matter components is therefore unaffected by the Thompson scattering, the mechanism that couples photons and baryons. As a consequence dark matter fluctuations are not constrained to be small at the decoupling epoch. After baryons are released, they start falling into the dark matter gravitational potential wells and the baryon fluctuations quickly equalize in amplitude the dark matter fluctuations ($\delta_B \sim \delta_{DM}$).

With the introduction of dark matter, also two orthogonal paradigms were proposed in order to explain the detailed mechanics of cosmic structure formation: the bottom-up and the top-down scenarios. According to the first paradigm, large scale structures are formed in a hierarchical way, i.e. the first structures to form in the universe are low mass objects that are then assembled together by gravity into more massive and complex structures. In the opposite scenario, also called the monolithic paradigm, the largest structures are the former to form. They subsequently fragment into a variety of smaller mass objects. These two scenarios makes explicit predictions about the nature of the dark matter particles. The bottom-up scenario requires dark matter particle to be non relativistic at sufficiently early time (cold dark matter (CDM)), wheres in the top-dawn scenario, the dark matter candidates are hot particles (HDM) with small mass (in the range $10-30$ eV). Since power spectrum of matter perturbations predicted by these two scenarios is different, one can therefore discriminate between them by reconstructing this statistics for example using large redshift surveys. Current observations, although not conclusive, seems to suggest the better viability of CDM models.

I will now present the formalism that describes the gravitational amplification of matter perturbations in an homogeneous expanding universe. Since the matter component is non-relativistic and pressure-less ($P \ll \rho c^2$), and since we are interested in perturbations on scales much smaller then the Hubble horizon ($\lambda \ll \frac{c}{H}$), the Newtonian approximation holds and the system describing their evolution (assuming an Eulerian point of view) is (Jain & Bertschinger, 1994)

$$\frac{\partial \delta}{\partial \tau} + \vec{\nabla}.[(1+\delta)\vec{v}] = 0 \tag{1.16}$$

$$\frac{\partial \vec{v}}{\partial \tau} + \mathcal{H}\vec{v} + (\vec{v}.\vec{\nabla})\vec{v} = -\vec{\nabla}\phi_e \tag{1.17}$$

$$\nabla^2 \phi_e = 4\pi G \rho a^2 \delta. \tag{1.18}$$

where $\delta(\vec{x})$ is the matter perturbation at position $\vec{x}$, $\vec{v}$ is the field of 3-D velocity perturbations (simply peculiar velocities), τ is the conformal time ($d\tau = dt/a(t)$) and $\mathcal{H}$ is the Hubble parameter in conformal time ($\mathcal{H} = aH$). Exploiting the definition of the reduced density parameter, $4\pi G\rho a^2 = \frac{3}{2}\mathcal{H}^2\Omega_m$, the last equation (1.18) can be rewritten as

$$\nabla^2 \phi_e = \frac{3}{2}\Omega_m \mathcal{H}^2 \delta.$$

By applying the gradient operator to equation (1.17), and by combining the resulting expression with equation (1.18) one obtains

$$\frac{\partial \vec{\nabla}\cdot\vec{v}}{\partial \tau} + \mathcal{H}\vec{\nabla}\cdot\vec{v} + \vec{\nabla}\cdot[(\vec{v}\cdot\vec{\nabla})\vec{v}] = -\frac{3}{2}\Omega_m\mathcal{H}^2\delta,$$

while the evolution of matter and velocity perturbations is described by

$$\frac{\partial \delta}{\partial \tau} + \vec{\nabla} \cdot \vec{v} = -\vec{\nabla}.(\delta \vec{v}) \tag{1.19}$$

and

$$\frac{\partial \vec{\nabla} \cdot \vec{v}}{\partial \tau} + \mathcal{H}\vec{\nabla} \cdot \vec{v} + \frac{3}{2}\Omega_m \mathcal{H}^2 \delta = -\vec{\nabla}.[(\vec{v}.\vec{\nabla})\vec{v}], \tag{1.20}$$

which is a non-linear system in the coupled unknowns δ and $\vec{v}$.

1.2.1 Solutions in the linear regime

The linear solution of equations (1.19) and (1.20) is obtained by setting the LHS terms to 0

$$\begin{cases} \dfrac{\partial \delta}{\partial \tau} + \theta &= 0 \\ \dfrac{\partial \theta}{\partial \tau} + \mathcal{H}\theta + \dfrac{3}{2}\Omega_m \mathcal{H}^2 \delta &= 0 \end{cases},$$

where θ is defined as the divergence of the velocity field. Combining the above expressions, one obtains the differential equation which rules the evolution of matter perturbation in a generic cosmology (containing pressure-less matter)

$$\frac{\partial^2 \delta}{\partial \tau^2} + \mathcal{H}\frac{\partial \delta}{\partial \tau} - \frac{3}{2}\Omega_m \mathcal{H}^2 \delta = 0, \tag{1.21}$$

or the equivalent expression (using cosmic time)

$$\frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} - \frac{3}{2}\Omega_m H^2 \delta = 0. \tag{1.22}$$

Since at linear order time and spatial dependences can be separated, i.e. the solution can be written as $\delta(\vec{x}, t) = D(t)\epsilon(\vec{x})$ (Peebles, 1980), the growing mode $D(t)$ evolves according to the equation

$$\frac{\mathrm{d}^2 D}{\mathrm{d}t^2} + 2H\frac{\mathrm{d}D}{\mathrm{d}t} - \frac{3}{2}\Omega_m H^2 D = 0. \tag{1.23}$$

Solution to this equation have been studied since the 70's. For instance Groth & Peebles (1975) derived the analytical solution for an open universe, dominated by pressure-less matter and without cosmological constant. They also investigated the effects induced by including a contribution from relativistic species, finding an analytical solution for a flat universe composed of dark matter and radiation. Anyway, radiation does not affect the growth of structures after the decoupling epoch, and we will neglect its presence in what follows.

Solving analytically the equation (1.23) is challenging but presents an obvious cosmological interest; the knowledge of the growth history of structures provides a very promising way of constraining the value of fundamental cosmological parameters (e.g. Wang &

Steinhardt, 1998). Heath (1977) found an integral representation of equation (1.23) in the general case of a pressure-less universe. He noticed that if radiation is neglected, the acceleration equation can be expressed in the following form

$$\frac{\mathrm{d}^2H}{\mathrm{d}t^2} + 2H\frac{\mathrm{d}H}{\mathrm{d}t} - \frac{3}{2}\Omega_m H^3 = 0. \tag{1.24}$$

Therefore the expansion rate $H(t)$ is a possible solution of equation (1.23). As the Hubble parameter decreases with increasing cosmic time, this solution corresponds to the decaying mode. As a consequence, any primordial perturbation will be suppressed by cosmic expansion. However, one can find a growing solution,

$$D(z) \propto H(z)\int_z^\infty \frac{(1+z')}{H^3(z')}\mathrm{d}z', \tag{1.25}$$

using the 'variation of the constant' method. Despite the simplicity of equation (1.25) it is more convenient to use the growth rate f in order to characterize the linear growing mode. By definition the growth rate is the logarithmic derivative of the growing mode $f \equiv \frac{\mathrm{d}\ln D}{\mathrm{d}\ln a}$. Itwas introduced first by (Peebles, 1980), then further investigated and analytically justified by Fry (1985), Lahav et al. (1991), Lightman & Schechter (1990). As these authors have shown that in many cosmological models the growth rate can be expressed in term of the growth index γ

$$f \simeq \Omega_m^\gamma(z), \tag{1.26}$$

where the growth index can be allowed to slowly vary with respect to cosmic time and encodes a part of the cosmological dependency. Once the growth index is known the growing mode can be calculated using

$$D(z) = A\exp\left\{\int_0^z \Omega_m^\gamma(z')\mathrm{d}\ln(1+z')\right\}.$$

If for a ΛCDM model there is no gain in using the approximation (1.26) to calculate the growing mode, in contradiction it becomes very usefull when modifying the gravity model or when taking into account a dark energy component with a time varying equation of state. As a matter of example if one consider a dark energy model ($w \neq -1$) with a constant equation of state, the accelaration equation (1.24) used to obtain the integral representation (1.25) becomes

$$\frac{\mathrm{d}^2H}{\mathrm{d}t^2} + 2H\frac{\mathrm{d}H}{\mathrm{d}t} - \frac{3}{2}\Omega_m H^3 = \frac{3}{2}(1+3w)(1+w)\Omega_X)H^3.$$

It shows that when $w \neq -1$ and $w < -1/3$ the Hubble rate is nolonger a solution of the equation (1.23). Since in Appendix A of (Linder & Jenkins, 2003) they show that, in general case, using definition (1.26) offers an accurate way of approximating the growth

rate for a wide range of cosmological models, including modifications of gravity it turns that the use of the growth index eases, indeed the calculation of the growing mode.

Moreover Linder & Jenkins (2003) have shown that adding constraint on the growth index to the constraint of expansion history helps to distinguish between a SUGRA model (massive gravity) and a quintessence model. Or, for instance, in Linder (2005) they pointed out that it allows to discriminate between a minimaly coupled scalar field model and a brane world gravity theory (DGP).

1.2.2 Solutions in the weakly non-linear regime

Although linear description allows extracting cosmological information about how inhomogeneities evolve on large cosmological scales, for a wide range of cosmological applications , especially if we are interested in more local dynamical processes, it is necessary to go beyond. Moreover, the study of the density and velocity evolution in the weakly non-linear regimes helps to refine our understanding of the hierarchical character of the structure formation processes in the universe. In this section I describe the second order Eulerian perturbation theory that stands as a founding pillar for the work I will present later in this manuscript (see chapter 3). I will also contrast this formulation with the Lagrangian approach and the spherical collapse approximation, that is alternative dynamical descriptions commonly adopted when exploring the weakly non-linear domain.

As shown above, matter and velocity perturbations are non-linearly coupled via the fluid equation (1.19) and the equation of motion (1.20). To proceed further, it is convenient to express this system in Fourier space. The spatial Fourier transform $\vec{f_k}$ of a vectorial field $\vec{f}$ which depends on position $\vec{x}$ is (in my notation convention)

$$\vec{f_k} = \mathcal{F}\left[\vec{f}(\vec{x})\right] = \frac{1}{(2\pi)^3}\int \vec{f}(\vec{x}) e^{i\vec{k}\cdot\vec{x}} d^3\vec{x}.$$

Accordingly, the inverse Fourier transform is given by

$$\vec{f}(\vec{x}) = \int \vec{f_k} e^{-i\vec{k}\cdot\vec{x}} d^3\vec{k}.$$

The continuity equation (1.19) in Fourier space is (see Appendix A)

$$\frac{\partial \delta_k}{\partial \tau} + \theta_k = -\iint \alpha_{12}\theta_{k_1}\delta_{k_2}\delta^D(\vec{k}-\vec{k}_{12}) d^3\vec{k}_1 d^3\vec{k}_2,$$

where δ^D is the Dirac distribution and α_{12} is a scalar combination of mode vectors $\vec{k}_1$ and $\vec{k}_2$

$$\alpha_{12} = \vec{k}_{12}.\frac{\vec{k}_1}{{k_1}^2},$$

and where $\vec{k}_{12} = \vec{k}_1 + \vec{k}_2$. It is then convenient to define $\Theta_k = -\frac{\theta_k}{\mathcal{H}}$, where $\mathcal{H}$ is the Hubble parameter in conformal time (therefore $\mathcal{H}\partial\tau = \partial \ln(a)$), so that in terms of these new

variable

$$\frac{\partial \delta_k}{\partial \ln a} - \Theta_k = \iint \alpha_{12} \Theta_{k_1} \delta_{k_2} \delta^D(\vec{k} - \vec{k}_{12}) d^3\vec{k}_1 d^3\vec{k}_2. \tag{1.27}$$

On the other hand, by applying the Fourier transform to the equation of motion (1.20) one obtains

$$\frac{\partial \theta_k}{\partial \tau} + \mathcal{H}\theta_k + \frac{3}{2}\mathcal{H}^2\delta_k = -\iint \beta_{12}\theta_{k_1}\theta_{k_2}\delta^D\left(\vec{k} - \vec{k}_{12}\right) d^3\vec{k}_1 d^3\vec{k}_2,$$

where

$$\beta_{12} \equiv \frac{\vec{k}_{12}.\vec{k}_2}{{k_2}^2}\frac{\vec{k}_1.\vec{k}_2}{{k_1}^2}.$$

so that

$$\frac{\partial \Theta_k}{\partial \ln a} + (2 - \epsilon_X - \Omega_m)\frac{\Theta_k}{2} - \frac{3}{2}\Omega_m\delta_k = \iint \beta\left(\vec{k}_1, \vec{k}_2\right)\Theta_{k_1}\Theta_{k_2}\delta^D\left(\vec{k} - \vec{k}_{12}\right) d^3\vec{k}_1 d^3\vec{k}_2. \tag{1.28}$$

where $\epsilon_X \equiv (1 + 3w)\Omega_X$.

Note that by re-introducing cosmic time as a variable, the dependence on cosmological parameters appears explicitly on the LHS. In particular, for an Einstein-de Sitter universe, the coefficients of the LHS are time-independent and it is possible to re-write the equations (1.27) and (1.28) using the matrix formalism introduced by Crocce & Scoccimarro (2006) (see Appendix A). In such an approach the kernels on the RHS of these equations are brought into a symmetric form

$$\begin{cases} \dfrac{\partial \delta_k}{\partial \ln a} - \Theta_k = \iint \dfrac{1}{2}[\alpha_{12}\Theta_{k_1}\delta_{k_2} + \alpha_{21}\Theta_{k_2}\delta_{k_1}]\delta^D(\vec{k} - \vec{k}_{12}) d^3\vec{k}_1 d^3\vec{k}_2 \\ \dfrac{\partial \Theta_k}{\partial \ln a} + (2 - \epsilon_X - \Omega_m)\dfrac{\Theta_k}{2} - \dfrac{3}{2}\Omega_m\delta_k = \iint \beta^s_{12}\Theta_{k_1}\Theta_{k_2}\delta^D\left(\vec{k} - \vec{k}_{12}\right) d^3\vec{k}_1 d^3\vec{k}_2 \end{cases},$$

where $\beta^s_{12} \equiv \frac{\beta_{12} + \beta_{21}}{2}$. A significant fraction of the non-linear contributions arise from the mode coupling functions α_{12} and β^s_{12} that mix in a complicated way the Fourier modes. A possible way of solving this system on non linear equations is to split the solution into different contributions

$$\delta(\vec{x}, t) = \sum_{n=1}^{\infty} \delta^{(n)}(\vec{x}, t) \tag{1.29}$$

$$\theta(\vec{x}, t) = \sum_{n=1}^{\infty} \theta^{(n)}(\vec{x}, t), . \tag{1.30}$$

sort them in order of importance (i.e. $X^{(n)} \gg X^{(n+1)}$) and hope that the system can, in

principle, be solved order by order. This is indeed what happens in the case of an Einstein-de Sitter universe (e.g. Moutarde, 1991; Goroff, 1986; Makino, 1992; Jain & Bertschinger, 1994; Bernardeau et al., 2002). Although this model does not correctly describe observations, it helps figuring out the way of proceeding when looking for the general solution. It is useful to take the Fourier transform of the perturbative expansions (A.7) and (A.8) and to look for separable solutions of the form (e.g. Bernardeau et al., 2002)

$$\delta_k(t) = \sum_{n=1}^{\infty} D_n(t)\delta_k^{(n)} \tag{1.31}$$

$$\Theta_k(t) = -\mathcal{H}f\sum_{n=1}^{\infty} E_n(t)\theta_k^{(n)}, \tag{1.32}$$

where the time dependency at each order (n) is encoded in the functions D_n and E_n that, by definition, are scale independent. The factorized time dependent part of the divergence of the velocity field (i.e. the term $-\mathcal{H}f$) imposes that, in linear regime, $E_1 = D_1$. By substituting the expressions (1.31) and (1.32) into the non linear system one can obtain a recursive system of differential equations (Bernardeau et al., 2002, see full calculation in Appendix A or in)

$$\frac{\mathrm{d}D_n}{\mathrm{d}\ln D_1}\delta_k^{(n)} - E_n\theta_k^{(n)} = \int \mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\Delta_n(\vec{k}_1,\vec{k}_2)\delta^D(\vec{k}-\vec{k}_{12}) \tag{1.33}$$

$$\frac{\mathrm{d}E_n}{\mathrm{d}\ln D_1}\theta_k^{(n)} + [\frac{3}{2}\frac{\Omega_m}{f^2} - 1]E_n\theta_k^{(n)} + \frac{3}{2}\frac{\Omega_m}{f^2}D_n\delta_k^{(n)} = \int \mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\Sigma_n(\vec{k}_1,\vec{k}_2)\delta^D(\vec{k}-\vec{k}_{12}), \tag{1.34}$$

where

$$\Delta_n(\vec{k}_1,\vec{k}_2) \equiv \frac{1}{2}\sum_{m=1}^{n-1}\left\{\alpha_{12}E_{n-m}D_m\theta_{k_1}^{(n-m)}\delta_{k_2}^{(m)} + \alpha_{21}E_mD_{n-m}\theta_{k_2}^{(n-m)}\delta_{k_1}^{(m)}\right\},$$

and

$$\Sigma_n(\vec{k}_1,\vec{k}_2) \equiv \beta_{12}^s\sum_{m=1}^{n-1}E_{n-m}E_m\theta_{k_1}^{(n-m)}\theta_{k_2}^{(m)}.$$

In equations (A.13) and (A.14) the time dependency of each order is expressed through the linear growing mode D_1 ($f\,\mathrm{d}\ln a = \mathrm{d}\ln D_1$). This change of variable $t \to D_1$ is motivated on one hand by an analogy with the Einstein-de Sitter case and on the other hand by the formal structure of the continuity equation. Indeed in the case of an Einstein-de Sitter universe the system can be analytically solved when the time dependency is expressed through the scale factor $a(t)$, which, in this cosmological model, represents the linear growing mode D_1 of the perturbations. It is therefore tempting to consider D_1 as the privileged variable which allows simplifying the system. Moreover through this change of variable the coefficients in the continuity equation becomes time-independent.

As stressed by (Bernardeau et al., 2002), the temporal and spatial functions (eq. 1.31 end 1.32) can be separated only if the coefficients of LHS terms do not depend on time which, considering eq. (A.14), happens only if $f = \Omega_m^{0.5}$. In such a case, it is as straightforward as in the $\Omega_m = 1$ case to solve the complete hierarchy of differential equations with respect to D_1. One obtains that $D_n = D_1^n$ and $E_n = D_n$. Interestingly this ansatz is very close to the approximated expression that describes the time evolution of the growth rate, that is $f = \Omega_m^{0.55}$ (see eq. 17 of Linder, 2005).

The system of equation has been solved up to third order in the case of closed universe without cosmological constant (Bouchet, 1992) and in the case of a flat universe with a cosmological constant (Bouchet, 1995). Even if the ansatz of time and space separability does not holds in these cosmological models, these authors showed that separable solutions can still be found if a Lagrangian approach (as opposed to the Eulerian point of view described here) is adopted (Zel'dovich, 1970; Buchert, 1989).

It can be shown that, even if the time dependency of the coefficients makes the resolution of the system somewhat complicated, order by order solutions (in the Eulerian space) can still be found even when we drop the separability hypothesis. To second order one obtains

$$\frac{\mathrm{d}\delta_k^{(2)}}{\mathrm{d}\ln D_1} - \theta_k^{(2)} = D^2 \iint \mathrm{d}^3\vec{k}_1 \mathrm{d}^3\vec{k}_2 \frac{\delta_{k_1}^{(1)}\delta_{k_2}^{(1)}}{2} \left\{\alpha_{12} + \alpha_{21}\right\} \delta^D(\vec{k} - \vec{k}_{12}) \quad (1.35)$$

$$\frac{\mathrm{d}\theta_k^{(2)}}{\mathrm{d}\ln D_1} + [\frac{3}{2}\frac{\Omega_m}{f^2} - 1]\theta_k^{(2)} + \frac{3}{2}\frac{\Omega_m}{f^2}\delta_k^{(2)} = D^2 \iint \mathrm{d}^3\vec{k}_1 \mathrm{d}^3\vec{k}_2 \delta_{k_1}^{(1)}\delta_{k_2}^{(1)}\beta_{12}^s \delta^D(\vec{k} - \vec{k}_{12}), \quad (1.36)$$

where $\delta_k^{(2)}$ and $\theta_k^{(2)}$ exhibit a non trivial mixing of time and wave vectors. Several ways exists to solve the second order perturbation system (eq. 1.35 and 1.36) (see a discussion in appendix A) and they converge to the same representation of the second order solution

$$\delta_k^{(2)} = D^2 \iint F_2\left(\vec{k}_1, \vec{k}_2\right) \delta_{k_1}^{(1)}\delta_{k_2}^{(1)}\delta^D\left(\vec{k} - \vec{k}_{12}\right) d^3\vec{k}_1 d^3\vec{k}_2, \quad (1.37)$$

where

$$F_2\left(\vec{k}_1, \vec{k}_2\right) = \left(\frac{3\nu_2}{4} - \frac{1}{2}\right) + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\left(\frac{k_2}{k_1} + \frac{k_1}{k_2}\right) + \left(\frac{3}{2} - \frac{3\nu_2}{4}\right)\left(\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\right)^2, \quad (1.38)$$

which is the second order perturbation theory kernel. All the cosmological dependency is encoded in the ν_2 coefficients that can be effectively calculated by using the spherical collapse dynamics (Bernardeau, 1994a; Folsalba & Gaztañaga I, 1998; Gaztañaga & Folsalba II, 1998; Folsalba & Gaztañaga III, 1998; Kamionkowski & Buchalter, 1999). An orthogonal way of solving the system in the Eulerian space was proposed by Catelan et al. (1995). Instead of solving the equations for the density and velocity perturbations (after elimi-

nating the gravitational potential), they first solved the equations for the gravitational potential and the velocity field and recovered the amplitude of density perturbations from these solutions.

I also found a way to solve the second order equations in Eulerian space without using the spherical model approximation, nor the intermediate solution for the gravitational potential. To this purpose, I expanded on an idea proposed by Kamionkowski & Buchalter (1999), but I developed it using a different formalism. Formalism and calculations are presented in Appendix A. Here I show that, interestingly, this independent approach allows demonstrating that the second order Kernel F_2 has the form

$$F_2\left(\vec{k}_1,\vec{k}_2\right) = (1-g) + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right) + g\left(\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\right)^2, \tag{1.39}$$

where g is the particular solution of the second order differential equation

$$\frac{\mathrm{d}^2 g}{\mathrm{d}\ln^2 D_1} + \frac{3}{2}\left(2+\frac{\Omega_m}{f^2}\right)\frac{\mathrm{d}g}{\mathrm{d}\ln D_1} + \left(2+\frac{3}{2}\frac{\Omega_m}{f^2}\right)g = 1.$$

By identifying the terms appearing in equation (1.39) and in equation (1.38), one can see that $3\nu_2 + 4g = 6$ is the relation that allows mapping the results obtained using the spherical collapse with the results obtained using the Eulerian approach from end to end.

I now briefly review the influence of specific background cosmological models on the results obtained from second order perturbation theory. To this purpose I focus on the term ν appearing in the second order PT kernel. Bernardeau & Brax (2011) showed that, for a flat universe with a cosmological constant, $\nu_2(\Omega_m) = 4/3 + (2/7)\Omega_m^{-1/143}$. Since, $\nu_2 \equiv \nu_2(\Omega_m = 1) = 34/21$ in the Einstein-de Sitter case, it follows that the relative difference between $\nu_2(\Omega_m)$ and ν_2 can be written as

$$\frac{\nu_2(\Omega_m)}{\nu_2} - 1 = \frac{6}{34}\left(\Omega_m^{-1/143} - 1\right).$$

Since the relative variation is lower than 0.3% (for Ω_m greater then 0.1) it is possible to neglect any cosmological dependency of the Kernel and rewrite it as

$$F_2\left(\vec{k}_1,\vec{k}_2\right) = \frac{5}{7} + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right) + \frac{2}{7}\left(\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\right)^2, \tag{1.40}$$

This form is extremely useful because it allows to predict the value of fundamental high order statistics of the matter density field as I will show in the next section. It is also useful because, recently, (Bernardeau & Brax, 2011) generalized the second order perturbation theory in order to take into account alternative gravitational theories and gave the expression of the ν_2 coefficients for such modified gravity (MG) models

$$\nu_2^{MG}/\nu_2^{GR} - 1 = -\frac{105}{4641}(\gamma - 0.55)(1 - \Omega_m)\Omega_m^{-0.45}, \tag{1.41}$$

where γ is the growth index in the considered MG model and 0.55 is the value of the growth index in standard GR. Equation (1.41) predicts that the amplitude of the deviations is lower than 0.5% for a DGP model ($\gamma = 0.68$) for a realistic matter density ($\Omega_m = 0.3$). This weak sensitivity implies that the 2-PT kernel (1.40) can be safely adopted even when considering modified gravity models.

1.3 Statistics of cosmic structures

In the previous section, I briefly reviewed the theory that allows extracting information about the dynamics of matter fluctuations in Fourier space and in the weakly non-linear regime. However, it is impossible, from an observational point of view, to follow the evolution of a single perturbation across different cosmic epochs. To assess the soundness of the theoretical predictions, that is to compare theoretical models and observations, we need to resort to statistical methods. In this section, after having introduced some general definitions and after having discussed some relevant properties and relations, I will present some of the most useful statistics used to characterize the large scale distribution of galaxies. In this context, I will pay particular attention to discuss a relatively unexplored statistical tools, the high order multi-point cumulant moments of the galaxy density field. Some central results obtained in this thesis, in fact, are based on applying these statistics in a cosmological context.

1.3.1 One-point statistics

A discrete spatial distribution of points , such as the distribution of galaxies, can be usfully analyzed using count-in-cell techniques. The number N of galaxies contained in a randomly positioned cell in comoving space is a random variable with probability distribution function P_N. For instance, if galaxies were distributed randomly in the universe, i.e. with the same probability of being found at a give position $\mathbf{x}$, the number-count N follows a Poisson law,

$$P_N = \frac{\lambda_o^N}{N!}e^{-\lambda_o}.$$

The shape of the distribution law is fully characterized by a single parameter, λ_o, that is related to the cell size, to the size of the galaxy survey and to the total number of objects contained in the survey. Since, however, galaxies cluster under the action of gravity, a single parameter PDF is not able to capture all the non-linear information imprinted in the spatial distribution. A PDF can also be described via its moments of order n

$$\langle N^n \rangle \equiv \sum_{N=0}^{\infty} N^n P_N,$$

where the moment of order one i $\bar{N}$ is the mean number of objects. It is possible to define other kind of moments with interesting mathematical and physical properties. For example, the central moments of order n are defined as

$$m_n = \langle (N - \bar{N})^n \rangle,$$

so that $m_1 = 0$. In addition, the central moment of order 2 is usually called the variance and provides a quantification of the dispersion of the random variable N around its mean value $\bar{N}$. Another useful definition is that of the factorial moments

$$f_n = \langle (N)_f^n \rangle = \langle N(N-1)(N-2)...(N-n+1) \rangle.$$

Note that, if the random variable N follows a Poisson law, then the factorial moments of order n are given by λ_o^n. This simple property may be used to test whether a given spatial distribution of points is random. An even more usufull way of statistically characterizing a random variable are the cumulant moments k_n. Although their formal representation is not as trivial as the previously introduced kind of moments, they are extensively used. In fact, they can be defined by expanding the moments in different connected parts which means that the knowledge of the cumulant moment at a given order provides an independent (which does not depend on lower order moments) statistical information. For example the cumulant of order two is obtained by subtracting the square of the moment of order one to the moment of order two. This way the information provided by k_2 does not depend on the first order moment (by definition $k_1 \equiv \bar{N}$). It leads to

$$k_2 = \langle N^2 \rangle - \bar{N}^2,$$

which reduces to the variance. Then at third order it follows

$$k_3 = \langle N^3 \rangle - 3k_2\bar{N} - \bar{N}^3.$$

As a matter of exemple the cumulant moments of the Poisson distribution are equal at all order to the parameter λ_o.

In the following, I introduce the formalism that, among other advantages, also allows to compute these various moments in a fast and efficient way.

1.3.2 Generating functions

The *probability generating function* is a powerful mathematical tool that allows to generate, in a simple and systematic way, the different types of moments introduced above. It is

constructed via a Taylor expansion

$$G(\lambda) = \sum_{N=0}^{\infty} \frac{1}{N!} \left. \frac{d^N G(\lambda)}{d\lambda^N} \right|_{\lambda=0} \lambda^N,$$

where, by construction,

$$\frac{1}{N!} \left. \frac{d^N G(\lambda)}{d\lambda^N} \right|_{\lambda=0} \equiv P_N.$$

Note that

$$G(\lambda) = \sum_{N=0}^{\infty} P_N \lambda^N = \langle \lambda^N \rangle. \tag{1.42}$$

where we have exploited the fact that P_N is a well normalized PDF.

In a similar way, the *moment generating function* $M(t)$ is defined via the following Taylor expansion

$$M(t) = \sum_{n=0}^{\infty} \left. \frac{d^n M(t)}{dt^n} \right|_{t=0} \frac{t^n}{n!},$$

where, by definition,

$$\left. \frac{d^n M(t)}{dt^n} \right|_{t=0} \equiv \langle N^n \rangle.$$

Moments, therefore, naturally follows by simply taking successive derivatives of the characteristic function. It is straightforward to show (see Appendix D for the details) that

$$M(t) = \langle e^{Nt} \rangle.$$

A change of variable in the equation (1.42) is enough to establish a relation between the probability generating function and the moment generating function

$$M(t) = G(\lambda = e^t). \tag{1.43}$$

The *generating function of factorial moments* $F(t)$ is defined as

$$F(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n F(t)}{dt^n} \right|_{t=0} t^n,$$

where, by definition

$$\left. \frac{d^n F(t)}{dt^n} \right|_{t=0} \equiv \langle (N)_f^n \rangle.$$

In Apendix (E) I show that it implies that

$$F(t) = \left\langle (1+t)^N \right\rangle,$$

which can be expressed with respect to the probability generating function with a simple

change of variable

$$F(t) = G(\lambda = 1 + t). \tag{1.44}$$

Finally, the *generating function of cumulant moments* $C(t)$ is defined as

$$C(t) \equiv \ln(M(t)). \tag{1.45}$$

so that cumulant moments be simply computed as

$$\langle N^n \rangle_c \equiv \left. \frac{d^n C(t)}{dt^n} \right|_{t=0}.$$

As an example, let's consider the Poisson distribution. Its probability generating function is

$$G(\lambda) = e^{\lambda_o(\lambda-1)}$$

from which we can immediately deduce that the moment generation function is $M(t) = \exp\{\lambda_o(e^t - 1)\}$, the cumulant moment generating function $C(t) = \lambda(e^t - 1)$ and the factorial moment generating function $F(t) = e^{\lambda_o t}$. On one hand, from $M(t)$ we can deduce that the expectation of the variable N is equal to λ_o. On on the other hand, the peculiar analytical expression of the generating function of cumulant moments shows that, at any order, the cumulant moments are always equal to the $\langle \bar{N} \rangle$. Finally, by calculating the derivatives of $F(t)$, one can verifies that $f_n = \bar{N}^n$.

1.3.3 Multi-point statistics

Since our final goal is to contrast observations against theory, we need to relate the formalism that describes discrete statistics (galaxies) with that describing continuous processes (i.e. the cosmic field of matter). In this section, I review some relevant properties of the continuous stochastic fields. Since the value a stochastic field assumes at a particular spatial position is not, in general, independent from the values at other locations, a rigorous description of the cosmic matter field requires a multi-point approach. I therefore show how the formalism presented above can be generalized to describe correlations in the stochastic field.

Be $\lambda(\mathbf{x})$ the intensity function describing the overall mass distribution in the universe, i.e. a stochastic field representing the density of matter at a given position $\mathbf{x}$ (e.g. Martinez & Saar, 2002) . We consider $\lambda(\mathbf{x})$ as a particular (homogeneous and isotropic) realization drawn from an ensemble $\mathcal{E}$ and indicate with $\mathcal{F}[\lambda(\mathbf{x})]$ its probability density functional (PdF).

Since the intensity field is positively defined, $\mathcal{F}[\lambda(\mathbf{x})]$ is by definition non Gaussian. Its complete characterization requires the knowledge of the entire (formally infinite) hierarchy of the K-point expectation values

$$\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})....\lambda(\vec{x_N})\rangle = \left\langle \prod_{i=1}^{N} \lambda(\vec{x_i}) \right\rangle .$$

Note that, in the limiting case ($\vec{x_1}$=$\vec{x_2}$=$\vec{x_3}$=....=$\vec{x_N}$) we recover the one-point moment of order N. In an analogous way, the K-point connected expectation value of the intensity field is

$$\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})....\lambda(\vec{x_N})\rangle_c = \left\langle \prod_{i=1}^{N} \lambda_R(\vec{x_i}) \right\rangle_c .$$

Given the extensive use I will do in this thesis of this last quantity, in the following I describe its properties in a somewhat more detailed way.

The cluster expansion technique

The N-point cumulant moments (i.e. the N-point connected expectation value) of the intensity field Λ, can be illustrated via the use of the cluster expansion technique. This tool is based on the fact that the expectation value taken at N positions contains lower order information which is stacked in the expectation values of $n < N$ points. The cluster expansion technique allows singling out the non independent (i.e. the connected) moments for any set of N points. This can be seen as an ordering of the information contained in the moments, so that the connected expectation values are irreducible, i.e. not anymore dependent on lower order information. As an example, I am going to treat the cases $N = 2$ and $N = 3$. In order to identify the connected parts of the 2-point moment $\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\rangle$, one can remark that if the distance between two points increases to infinity, then the values taken by the random field λ, at position $\vec{x}_1$ and $\vec{x}_2$, are independent. Thus $\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\rangle = \langle\lambda(\vec{x_1})\rangle_c\langle\lambda(\vec{x_2})\rangle_c$ represents the lower order contribution. The connected expectation value, instead, represents the remaining contribution. More formally

$$\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\rangle = \langle\lambda(\vec{x_1})\rangle_c\langle\lambda(\vec{x_2})\rangle_c + \langle\lambda(\vec{x_1})\lambda(\vec{x_2})\rangle_c,$$

where, by definition, $\langle\lambda\rangle_c \equiv \langle\lambda\rangle$. For $N = 3$ we can identify three kinds of configurations

- the tree points are infinitely separated $\langle\lambda(\vec{x_1})\rangle\langle\lambda(\vec{x_2})\rangle\langle\lambda(\vec{x_3})\rangle$
- only one point is infinitely separated from the others $\langle\lambda(\vec{x_1})\rangle\langle\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle_c$ (plus two permutations)
- the connected part of the three points $\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle_c$

This classification leads to

$$\begin{aligned}\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle = \quad & \langle\lambda(\vec{x_1})\rangle\langle\lambda(\vec{x_2})\rangle\langle\lambda(\vec{x_3})\rangle + \langle\lambda(\vec{x_1})\rangle\langle\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle_c \\ & +\langle\lambda(\vec{x_2})\rangle\langle\lambda(\vec{x_1})\lambda(\vec{x_3})\rangle_c + \langle\lambda(\vec{x_3})\rangle\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\rangle_c \ . \\ & +\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle_c\end{aligned}$$

or, in an equivalent way, to

$$\begin{aligned}\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle_c = \quad & \langle\lambda(\vec{x_1})\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle - \langle\lambda(\vec{x_2})\rangle\langle\lambda(\vec{x_1})\lambda(\vec{x_3})\rangle_c \\ & -\langle\lambda(\vec{x_3})\rangle\langle\lambda(\vec{x_1})\lambda(\vec{x_2})\rangle_c - \langle\lambda(\vec{x_1})\rangle\langle\lambda(\vec{x_2})\lambda(\vec{x_3})\rangle_c \ , \\ & -\langle\lambda(\vec{x_1})\rangle\langle\lambda(\vec{x_2})\rangle\langle\lambda(\vec{x_3})\rangle\end{aligned}$$

This last expression clarifies that the cumulant moments are obtained by subtracting lower order independent contributions. Although this technique helps us to grasp the physical meaning of cumulants, it does not help to compute them in a fast and efficient way. To accomplish this task, it is more convenient to make use, as for the 1-point statistics, the generating function formalism.

Joint K-point cumulant moments of the continuous fluctuation field

Since the probability that a continuous stochastic variable assume a given specific value is null, it is no longer possible to define the probability generating function. However, we can still define the cumulant moments and the factorial moments of the stochastic field $\lambda(\vec{x})$. In the discrete case the generating functions of all kind of moments can be deduced from the probability generating function, in the continuous case the generating functions of cumulants moments and factorial moments can be derived form the moment generating function $\mathcal{M}_\lambda[J]$

$$\langle\lambda(\mathbf{x}_1)\,...\,\lambda(\mathbf{x}_K)\rangle = \frac{\delta^K\mathcal{M}_\lambda[J]}{\delta J(\mathbf{x}_1)\,...\,\delta J(\mathbf{x}_K)}\Big|_{J(\mathbf{x}_1)=...=J(\mathbf{x}_K)=0} \tag{1.46}$$

where

$$\mathcal{M}_\lambda[J] = \int \mathcal{D}\lambda(\mathbf{x})\mathcal{F}[\lambda(\mathbf{x})]e^{i\int d\mathbf{x}J(\mathbf{x})\lambda(\mathbf{x})} = \left\langle e^{i\int d\mathbf{x}J(\mathbf{x})\lambda(\mathbf{x})}\right\rangle$$

is the K-point moment generating functional, and where $\mathcal{D}\lambda(\mathbf{x})$ represents a suitable measure introduced in $\mathcal{E}$ such that the total probability turns out to be normalized to 1 (Matarrese, Lucchin & Bonometto, 1986).

The cosmological principle guarantees the existence of a non-zero value for the expected value of the λ field. It is thus convenient to characterize inhomogeneities in the matter distribution in terms of the local dimensionless density contrast

$$\delta(\mathbf{x}) \equiv \frac{\lambda(\mathbf{x})}{\langle\lambda(\mathbf{x})\rangle} - 1. \tag{1.47}$$

In the following, we assume that the ensemble average over $\mathcal{E}$ is equivalent to averaging a particular realization over different spatial positions. If this operation is applied on a fair sample, $\langle\delta(\mathbf{x})\rangle = 0$. As a consequence, if the intensity function $\lambda(\vec{x})$ represents the matter density, $\langle\lambda\rangle = \bar{\rho}$.

An useful tool for probing the large-scale cosmological structure, because of their relatively simple connection to both theory and observations, are the joint K-point cumulant moments of order $\mathbf{n} = (n_1, n_2.....n_K)$ of the cosmic over density field

$$\kappa_{n_1,...,n_K}(\mathbf{x}_1, ..., \mathbf{x}_K) \equiv \langle\delta^{n_1}(\mathbf{x}_1)\,...\,\delta^{n_K}(\mathbf{x}_K)\rangle_c, \tag{1.48}$$

and, in particular, the irreducible K-point auto-correlation functions

$$\kappa_{1,...,1}(\mathbf{x}_1, ..., \mathbf{x}_K) \equiv \langle\delta(\mathbf{x}_1)\,...\,\delta(\mathbf{x}_K)\rangle_c. \tag{1.49}$$

generally denoted as ξ_K and shortly called correlation functions. By definition, the generating functional of the K-point cumulant moment of the over density field is the logarithm of the moment generating functional

$$\mathcal{C}[J] \equiv \ln \mathcal{M}_\delta[J]. \tag{1.50}$$

The advantage of computing connected averages instead of statistical averages ($\mu_{n_1,...,n_K} \equiv \langle\delta^{n_1}(\mathbf{x}_1)\,...\,\delta^{n_K}(\mathbf{x}_K)\rangle$) is that cumulants are zero if the random variables representing the value of the stochastic field at different space positions are statistically independent. Conversely, a cumulant is not zero if and only if the random variables in it are statistically "connected". A specific consequence of this properties is that $\kappa_{n_1,...,n_K} \to 0$ as any subset of positions $\mathbf{x}_i$ are displaced to infinite separation. Finally, cumulants can be explicitly represented in terms of only the lower order moments, that is as a function of $\mu_{m_1...m_K}(\mathbf{x}_1, ..., \mathbf{x}_K)$ with $0 \leq m_i \leq n_i$. This follows from taking successive functional derivatives of the generating functional $\mathcal{C}[J]$

$$\kappa_{n_1,...,n_K} = \frac{\delta^N \mathcal{C}[J]}{\delta J^{n_1}(\mathbf{x}_1)\,...\,\delta J^{n_K}(\mathbf{x}_k)}\Big|_{J(\mathbf{x}_1)=...=J(\mathbf{x}_K)=0} \tag{1.51}$$

where $N = \sum_{i=1}^{K} n_i$. The result can be formally written as (Meeron, 1957)

$$\kappa_{n_1,...,n_K} = -\prod_j \nu_j! \sum_{l=1}^{N} \sum_{\gamma_i, m_{ij}} (\sum \gamma_i - 1)!(-1)^{\sum \gamma_i} \prod_{i=1}^{l} \frac{1}{\gamma_i!} \left\{ \frac{\langle \prod_{j=1}^{K} \delta^{n_{ij}}(\mathbf{x_j})\rangle}{\prod_j n_{ij}!} \right\}^{\gamma_i}. \tag{1.52}$$

In this last equation, γ_i and m_{ij} are non-negative integers satisfying the set of equations

$$\sum_{i=1}^{l} \gamma_i m_{ij} = n_j$$

and bookkeeping all the possible decompositions

$$(\delta^{n_1}(\mathbf{x}_1)...\delta^{n_K}(\mathbf{x}_K)) \rightarrow \prod_{i=1}^{l}(\delta^{m_{i1}}(\mathbf{x}_1)...\delta^{m_{iK}}(\mathbf{x}_K))^{\gamma_i}.$$

In this study we are interested in the joint cumulant moments taken at two different locations $\mathbf{x}_1$ and $\mathbf{x}_2$ up to the order $N = n_1 + n_2 = 7$. For these statistics, from now on simply called correlators and indicated as κ_{nm}, eq. 1.52 gives

$$\begin{aligned}
\kappa_{11} &= \mu_{11} \qquad (1.53)\\
\kappa_{12} &= \mu_{12}\\
\kappa_{13} &= \mu_{13} - 3\,\mu_{11}\mu_2\\
\kappa_{22} &= \mu_{22} - 2\,{\mu_{11}}^2 - {\mu_2}^2\\
\kappa_{14} &= \mu_{14} - 6\,\mu_{12}\mu_2 - 4\,\mu_{11}\mu_3\\
\kappa_{23} &= \mu_{23} - 6\,\mu_{12}\mu_{11} - \mu_3\mu_2 - 3\,\mu_{12}\mu_2\\
\kappa_{15} &= \mu_{15} - 5\,\mu_{11}\mu_4 - 10\,\mu_{12}\mu_3 - 10\,\mu_{13}\mu_2 + 30\,\mu_{11}{\mu_2}^2\\
\kappa_{24} &= \mu_{24} - 4\,\mu_{12}\mu_3 - 8\,\mu_{13}\mu_{11} - 6\,\mu_{22}\mu_2 - 6\,{\mu_{12}}^2 + 6\,{\mu_2}^3 + 24\,{\mu_{11}}^2\mu_2 - \mu_4\mu_2\\
\kappa_{33} &= \mu_{33} - 6\,\mu_{13}\mu_2 - {\mu_3}^2 - 9\,{\mu_{12}}^2 + 12\,{\mu_{11}}^3 - 9\,\mu_{22}\mu_{11} + 18\,\mu_{11}{\mu_2}^2\\
\kappa_{16} &= \mu_{16} - 6\,\mu_{11}\mu_5 - 20\,\mu_{13}\mu_3 - 15\,\mu_{12}\mu_4 - 15\,\mu_{14}\mu_2 + 120\,\mu_{11}\mu_2\mu_3 + 90\,\mu_{12}{\mu_2}^2\\
\kappa_{25} &= \mu_{25} + 40\,{\mu_{11}}^2\mu_3 - 10\,\mu_{14}\mu_{11} - 5\,\mu_{12}\mu_4 - 20\,\mu_{13}\mu_{12} + 30\,\mu_{12}{\mu_2}^2 - 10\,\mu_{22}\mu_3\\
&\quad -10\,\mu_{23}\mu_2 - \mu_5\mu_2 + 120\,\mu_3{\mu_2}^2 + 20\,\mu_{11}\mu_2\mu_{12}\\
\kappa_{34} &= \mu_{34} - 4\,\mu_{13}\mu_3 - 12\,\mu_{13}\mu_{12} - 18\,\mu_{22}\mu_{12} - 3\,\mu_{14}\mu_2 + 24\,\mu_{11}\mu_2\mu_3 + 36\,\mu_{12}{\mu_2}^2\\
&\quad -6\,\mu_{23}\mu_2 + 6\,\mu_3{\mu_2}^2 + 72\,{\mu_{11}}^2\mu_{12} - 12\,\mu_{23}\mu_{11} - \mu_3\mu_4 + 72\,\mu_{11}\mu_2\mu_{12}
\end{aligned}$$

where I have used the fact that $\langle\delta\rangle = 0$ and where I used the reduced notations $\kappa_{nm} = \langle\delta_R^n(\vec{x}_1)\delta_R^m(\vec{x}_2)\rangle_c$ and $\mu_{nm} = \langle\delta_R^n(\vec{x}_1)\delta_R^m(\vec{x}_2)\rangle$. In the 1-point limiting case one recovers the expressions of the cumulant moments κ_N given by Fry (1984b) (cfr. eq. 17). Note, also, that correlators are symmetric with respect to exchanging the indexes. Technically, I computed the above relations between high order 2-point cumulants and moments by using

$$\kappa_{nm} = \frac{\delta^{n+m}\ln\{\mathcal{M}[J]\}}{\delta J^n(\mathbf{x}_1)\delta J^m(\mathbf{x}_2)}\Big|_{J(\mathbf{x}_1)=J(\mathbf{x}_2)=0}. \qquad (1.54)$$

and by taking successive derivatives with the software *Mathematica*. I also inverted equations (1.53). In Appendix B I provide the expressions of the 2-point moments as a function of the 2 point cumulant moments up to order 7.

1.3.4 The 2-point autocorrelation function in Fourier space : the power spectrum

A very useful way of dealing with the correlator of order $(1,1)$, i.e. the 2-point autocorrelation function ξ, is to work out its Fourier space analog, i.e. the power spectrum If I Fourier transform the fluctuation field at a position $\vec{x}_1$

$$\delta_{k_1} = \int \delta(\vec{x}) e^{-i\vec{k_1}\cdot\vec{x}} \mathrm{d}^3\vec{x}$$

it follows that it is possible to express the 2-point correlation function in Fourier space as

$$\langle \delta_{k_1} \delta_{k_2} \rangle = \delta^D(\vec{k}_1 + \vec{k}_2) \frac{1}{(2\pi)^3} \int \xi(\vec{r}) e^{i\vec{k}_2\cdot\vec{r}} \mathrm{d}^3\vec{r}.$$

It is then natural to define the power spectrum $P(\vec{k})$ as the Fourier transform of the 2-point correlation function which leads to

$$\langle \delta_{k_1} \delta_{k_2} \rangle = \delta^D(\vec{k}_1 + \vec{k}_2) P(\vec{k}).$$

As a result, once the spectral decomposition $P(k)$ is known, the second order statistics such as 2-point correlation function or the variance of the matter field ($\vec{r} = 0$) can be both predicted.

Assuming that primordial fluctuations (resulting from quantum fluctuations amplified by inflation) are distributed according to a multivariate Gaussian distribution, the Wick's theorem (e.g. Bernardeau, 2008) guarantees that the full hierarchy of N-point correlation functions are completely specified once the power spectrum is known. The main issue is therefore to know the initial power spectrum of the perturbations.

At the end of inflation the primordial power spectrum which is predicted by most of the inflationary models is of the type $P_o(k) = Ak^{n_s}$ with scalar spectral index $n_s \sim 1$. Note that, in particular case $n_s = 1$ $P_o(k)$ is called scale free primordial power spectrum (Harrison, 1970; Zel'dovich, 1970) because it corresponds to the spectral index for which the associated gravitational potential is scale independent. Observationally the spectral index is well constrained by the CMB data. In particular according to WMAP measurements $n_s = 0.96 \pm 0.014$.

The shape of this primordial power spectrum, is subsequently modified because of the specific way in which matter perturbations grow in an expanding universe. During the inflationary epoch, the size of the horizon is constant, and density perturbations, whose size grows as the scale factor of the universe, exit the horizon. If they re-enter before the epoch of equivalence, the epoch at which the radiation and matter density of the universe are identical, the amplitude of the dark matter perturbations cannot grow until the equivalence epoch is reached. Dark matter perturbations stagnates, in a radiation dominated universe. Since fluctuations characterized by a larger wavelength enter the horizon later than smaller

one, large perturbations re-entering into the horizon after the radiation epoch are not affected by this characteristic freezing: only dark matter perturbations on small scales are suppressed. The growth dynamics characterizing the baryonic component is quite different. The tight coupling between electrons and photons besides preventing fluctuations from growing, force standard matter to oscillate up to the epoch of decoupling.

The physics responsible for the modifications of the primordial power spectrum between the end of inflation and the decoupling epoch can be simply encapsulated in a transfer function of matter and baryons $T(k)$. This function maps the primordial fluctuations into the initial one

$$\delta_{k,i} = \delta_{k,o} T(k),$$

where $k = ||\vec{k}||$. It follows that the linear power spectrum P_L can be deduced from the primordial power spectrum P_o as

$$P_L(k) = P_o(k) T^2(k).$$

In the following of this thesis, I will assume that linear perturbations are described by the linear (dimensionless) power spectrum

$$\Delta_L^2 = 4\pi A k^{n_s+3} T^2(k), \tag{1.55}$$

where A is the normalization factor of the primordial power spectrum, n_s the primordial spectral index and $T^2(k)$ the transfer function (Bardeen et al., 1986; Efstathiou, Bond & White, 1992; Eisenstein & Hu, 1998).

1.3.5 Hierarchy of N-point autocorrelation functions

The Hierarchical ansatz is a theoretical model for the N-point correlation function (e.g. Fry (1984a,b)). According to this hypothesis, the N-point correlation function can be expressed as a sum over products of the (lower order) 2-point correlation functions

$$\xi_N(\mathbf{x}_1, ..., \mathbf{x}_N) = \sum_{k=1}^{t_N} Q_{N,k} \sum_{N-trees} \prod_{labeling\ edges} \xi_{ij}, \tag{1.56}$$

where t_N is the number of possible topologies that connect the N points, $Q_{N,k}$ are the reduced amplitudes, i.e. a set of constants associated to the chosen topology, and where the product is made over all the possible way of connecting the N points once a topology is considered. For convenience, one defines $\xi_{ij} \equiv \xi_2(\mathbf{x}_i, \mathbf{x}_j)$. Note, also, that topologies or graphs connecting the N points have no cycles (not closed graphs). As a consequence the number of product is $N-1$. Although the HA hypothesis has been shown to describe in a poor way the clustering of matter, it still remains a valuable pedagogical tool that allows

me to introduce some fundamental relations whose validity is confirmed by a coherent use of non-linear perturbation theory.

In figure (4), taken from Fry (1984b), the possible topologies up to order 7, for $N = 3$ are listed. Since it exists only one possible graph that connect the 3 points, we define $Q_3 \equiv Q_{3,1}$ which is usually called the reduced amplitude of the 3-point correlation function. In this specific case, the expression (1.56) reduces to

$$\xi_3(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = Q_3 \left\{ \xi_{12}\xi_{23} + \xi_{12}\xi_{13} + \xi_{13}\xi_{23} \right\}. \tag{1.57}$$

Note that, in the specific case in which the labels are identical, we can define the reduced $1-$point cumulant moment S_3 (hereafter simply the reduced skewness) as

$$\langle \delta^3 \rangle_c = S_3 \langle \delta^2 \rangle_c^2, \tag{1.58}$$

where $S_3 \equiv 3Q_3$. In a similar way, by taking a less restrictive limiting case (i.e. by identifying the labels 2 and 3), equation (1.57) naturally defines the reduced $2-$point cumulant moment C_{12} (hereafter simply correlator of order $(1, 2)$)

$$\langle \delta(\mathbf{x}_1)\delta^2(\mathbf{x}_2) \rangle_c = C_{12} \langle \delta^2 \rangle_c \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle_c \left\{ 1 + \frac{\eta}{2} \right\}, \tag{1.59}$$

where $C_{12} \equiv 2Q_3$ and $\eta \equiv \frac{\langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2) \rangle_c}{\langle \delta^2 \rangle_c}$. Note that, when no smoothing is applied to the density contrast of matter, the variance of the fluctuations $\langle \delta^2 \rangle_c$ goes, by definition, to infinity. Therefore, $\eta \to 0$ and it can be neglected in equation (1.59).
By considering the same limiting cases that lead to equations (1.58) and (1.59) and by repeatedly applying them to higher order correlation functions, we can define both the reduced cumulant of order N

$$\kappa_N = S_N \kappa_2^{N-1} \tag{1.60}$$

and the reduced correlators of order $n + m = N$

$$\kappa_{nm} = C_{nm} \kappa_{11} \kappa_2^{n+m-2}. \tag{1.61}$$

In the following I show the great interest of using the correlators in order to characterize the non Gaussian multi-point fields. First, assuming that Q_3 is a constant, and given the relations

$$C_{12}(r) = 2Q_3, \tag{1.62}$$

$$S_3 = 3Q_3. \tag{1.63}$$

we can deduce a simple relation between third order statistics

$$S_3 = \frac{3}{2} C_{12}. \tag{1.64}$$

and see that the reduced amplitude can be measured either with the cumulant or the correlator. Now, at fourth order, equation (1.56) can be reduced to the three following limiting cases ($\mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}_4$), ($\mathbf{x}_1 = \mathbf{x}_3$ and $\mathbf{x}_2 = \mathbf{x}_4$) and finally ($\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_3 = \mathbf{x}_4$). We therefore obtain that

$$\begin{cases} C_{13} &= 6Q_{4,1} + 3Q_{4,2} \\ C_{22} &= 4Q_{4,1} \\ S_4 &= 12Q_{4,1} + 4Q_{4,2} \end{cases} . \tag{1.65}$$

By combining equations (1.65) we can establish the link between 1-point and 2-point statistics at order four

$$S_4 = C_{22} + \frac{4}{3} C_{13}.$$

Equations (1.65) show that the complete hierarchy at fourth order involves two possible topologies characterized by $Q_{4,1}$ and $Q_{4,2}$, while the third equation emphasizes that these two quantities are strongly degenerated, and the simple knowledge of the 4^{th} order cumulant is not enough to capture all the information stacked in the 4^{th} order $2-$point statistics (as it was the case at third order (see equation 1.64)).

1.3.6 The continuum-discrete connection: sampling a stochastic field

A discrete distribution can be thought as resulting from the sampling of an underlying intensity function $\lambda(\vec{x})$. Let's interpret the intensity function as the probability density of finding a given object in an elementary volume $d^3\vec{x}$. The (continuous) probability of finding objects in a cell of finite volume v is

$$\Lambda(\vec{x}) = \int_{v(\vec{x})} \lambda(\vec{x}') d^3\vec{x}', \tag{1.66}$$

where $v(\vec{x})$ represents the volume of a cell positioned at position $\vec{x}$. This quantity (Λ) can be naturally interpreted as the parent continuos field which is discretely sampled by galaxies. More formally, $P[N|\Lambda]$ represents the conditional probability of finding N discrete objects in a cell centered at position $\mathbf{x}$ where the underlying continuous field assumes the value $\Lambda(\mathbf{x})$. It is, generally assumed, in what is called the "fair sample model of the universe" that such sampling is well described by the local Poisson process (LPP, see Layser, 1956). Indeed, this model assumes that counting galaxies in cells is like counting rare events. Therefore, the probability of finding N galaxies in a cell is given by the integral relation

$$P_N = \int_0^\infty P[N|\Lambda] P(\Lambda) d\Lambda, \tag{1.67}$$

where $P[N|\Lambda] = \frac{\Lambda^N}{N!}e^{-\Lambda}$. From the above expressions it follows that $\bar{N} \equiv \bar{\Lambda}$, i.e. the moment of order one of the discrete field is equivalent to the equal order moment of the continuous field.

Equation (2.6) can be used to obtain information on the stochastic field Λ once we know the discrete counting probability P_N. The standard way to characterize the stochastic field Λ is by establishing a mapping between the moments of the continuous field ($\langle\Lambda\rangle$) and the moments of the underlying discrete galaxy field ($\langle N\rangle$). This is equivalent to find the corrections that minimize the sampling noise and restore the original continuous signal. These relations can be obtained using the generating function technique, and this task is particularly easy if we assume a LPP. While the corrections for Poisson sampling noise are well known in the case of 1-point statistics, establishing these relations in the case of multi-point statistics is less immediate. Szapudi et al. (1995) for example worked out this mapping in the case of two-point moments. Here I briefly summarize how they obtained their results.

The probability generating function of a discrete random variable N is $G(\lambda) = \langle\lambda^N\rangle$ where, by definition, $\frac{\mathrm{d}G}{\mathrm{d}\lambda}\Big|_{\lambda=0} = N!P_N$. By changing variable $\lambda = 1+t$ we obtain the factorial moment generating function $F(t)$. For two points statistics, the probability generating function is

$$G(\lambda_1, \lambda_2) = \langle\lambda_1^{N_1}\lambda_2^{N_2}\rangle\ . \tag{1.68}$$

On the other hand, one can generalize the expression (2.6) to obtain

$$P_{N_1,N_2} = \int_0^{+\infty}\int_0^{+\infty} P\left[N_1|\Lambda_1, N_2|\Lambda_2\right] P\left[\Lambda_1, \Lambda_2\right] \mathrm{d}\left[\Lambda_1\right] \mathrm{d}\left[\Lambda_2\right]\ , \tag{1.69}$$

By assuming the LPP, and that the sampling does not introduce correlations, we have $P\left[N_1|\Lambda_1, N_2|\Lambda_2\right] = \frac{\Lambda_1^{N_1}}{N_1!}e^{-\Lambda_1}\frac{\Lambda_2^{N_2}}{N_2!}e^{-\Lambda_2}$ which we can substitute in (1.69)

$$P_{N_1,N_2} = \int_0^{+\infty}\int_0^{+\infty} \frac{\Lambda_1^{N_1}}{N_1!}e^{-\Lambda_1}\frac{\Lambda_2^{N_2}}{N_2!}e^{-\Lambda_2} P\left[\Lambda_1, \Lambda_2\right] \mathrm{d}\left[\Lambda_1\right] \mathrm{d}\left[\Lambda_2\right].$$

and by multiplying each side of the previous expressions by $\lambda_1^{N_1}$ and $\lambda_2^{N_2}$ and by summing over N_1 and N_2 we obtain

$$\begin{aligned} G(\lambda_1, \lambda_2) &= \int_0^{+\infty}\int_0^{+\infty} P\left[\Lambda_1, \Lambda_2\right] \sum_{N_1,N_2}^{\infty} \frac{(\lambda_1\Lambda_1)^{N_1}}{N_1!}e^{-\Lambda_1}\frac{(\lambda_2\Lambda_2)^{N_2}}{N_2!}e^{-\Lambda_2}\mathrm{d}\left[\Lambda_1\right] \mathrm{d}\left[\Lambda_2\right] \\ &= \int_0^{+\infty}\int_0^{+\infty} P\left[\Lambda_1, \Lambda_2\right] e^{(\lambda_1-1)\Lambda_1+(\lambda_2-1)\Lambda_2}\mathrm{d}\left[\Lambda_1\right] \mathrm{d}\left[\Lambda_2\right] \\ G(1+t_1, 1+t_2) &= \mathcal{M}(e^{t_1}, e^{t_2}). \end{aligned} \tag{1.70}$$

It is therefore convenient to define the factorial generating function for 2-point statistics as $F(t_1, t_2) \equiv G(1 + t_1, 1 + t_2)$. From this definition, we obtain that

$$\begin{aligned} F(t_1,t_2) &= \sum_{N_1,N_2}^{\infty} P_{N_1,N_2}(1+t_1)^{N_1}(1+t_2)^{N_2} \\ &= \sum_{N_1,N_2}^{\infty} P_{N_1,N_2} \sum_{n,m}^{\infty} N_1(N_1-1)...(N_1-n+1)N_2(N_2-1)...(N_2-n+1)\frac{t_1^n}{n!}\frac{t_2^m}{m!} \\ &= \sum_{n,m}^{\infty} \frac{t_1^n t_2^m}{n!m!} \sum_{N_1,N_2}^{\infty} P_{N_1,N_2}(N_1)_f^n (N_2)_f^m \\ &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{t_1^n t_2^m}{n!m!} \langle (N_1)_f^n (N_2)_f^m \rangle \end{aligned} \quad , \tag{1.71}$$

Noting that the definition of the factorial generating function for 2-point statistics is

$$F(t_1,t_2) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{t_1^n t_2^m}{n!m!} \left.\frac{\mathrm{d}^{n+m}F}{\mathrm{d}t^n \mathrm{d}t^m}\right|_{t_1=0,t_2=0} , \tag{1.72}$$

and since the 2-point factorial moments of order n, m are defined as $f_{n,m} \equiv \left.\frac{\mathrm{d}^{n+m}F}{\mathrm{d}t^n \mathrm{d}t^m}\right|_{t_1=0,t_2=0}$, we obtain that

$$f_{n,m} = \langle (N_1)_f^n (N_2)_f^m \rangle \ . \tag{1.73}$$

From the relation (1.70) linking the factorial generating function to the moment generating function, we can deduce that the 2-point moments of the stochastic field $\Lambda(\vec{x})$ can be estimated by computing the 2-point factorial moments (eq. 1.73) of the discrete random variable N, that is

$$\langle \Lambda_1^n \Lambda_2^m \rangle = \langle N_1(N_1-1)...(N_1-n+1)N_2(N_2-1)...(N_2-m+1) \rangle \equiv \langle (N_1)_f^n (N_2)_f^m \rangle \ . \tag{1.74}$$

This result generalizes the relation

$$\langle \Lambda^n \rangle = \langle N(N-1)...(N-n+1) \rangle \equiv \langle (N)_f^n \rangle . \tag{1.75}$$

that holds in the case of 1-point statistics. If the sampling process does not satisfy the LPP approximation, the relation between the continuous and discrete moments will be in general different.

1.3.7 The continuum-discrete connection: statistics of smoothed fields

In order to ease the comparison with an intrinsically discrete process such as the distribution of galaxies, as well as to facilitate the incorporation of the biasing scheme (e.g. eq. 1.93) into the study of matter fluctuations, it is useful to smooth the mass distribution on a spatial scale R. This is done by convolving the over density field δ with a (normalized) window function of size R.

$$\delta_R(\mathbf{x}) = \int \delta(\mathbf{x}')W\left[\frac{\mid \mathbf{x}-\mathbf{x}' \mid}{R}\right]d^3\mathbf{x}'. \tag{1.76}$$

Smoothing and averaging are non commutative operations. As a consequence, while the relations given in eq. 1.53 retains their validity when applied to filtered fields, the amplitudes of the smoothed dynamical variables become, instead, scale dependent.

Consider, for example, the lowest-order non-zero, 1− and 2−point cumulant moments of a smoothed density field, that is the variance of the mass fluctuations on a scale R

$$\sigma_R^2 = \kappa_{2,R} = \langle \delta_R^2(\mathbf{x}) \rangle_c, \tag{1.77}$$

and the covariance of the smoothed mass over density field

$$\xi_R(r) = \kappa_{11,R} = \langle \delta_R(\mathbf{x})\delta_R(\mathbf{x}+\mathbf{r}) \rangle_c. \tag{1.78}$$

Suppose, further, that mass fluctuations are small ($|\delta| \ll 1$). The problem is to work out theoretical predictions for the evolution of these relevant cosmological quantities. It is easy to show, using the Newtonian approximation, and the fact that smoothing in real space is equivalent to multiplication in Fourier space

$$\delta_{R,k} = \hat{W}_{TH}(kR)\delta_k,$$

that

$$\sigma_R^2(z) = \sigma_8^2(0)D^2(z)\mathcal{F}_R, \tag{1.79}$$

and

$$\xi_R(r,z) = \sigma_8(0)^2 D^2(z)\mathcal{G}_R(r). \tag{1.80}$$

The normalization of these equations is conventionally fixed at a scale $r_8 = 8h^{-1}$Mpc, $D(t)$ represents the linear growing mode, while the effects of filtering are incorporated in the functions

$$\mathcal{F}_R = \frac{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kR)d\ln k}{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kr_8)d\ln k} \tag{1.81}$$

and

$$\mathcal{G}_R(r) = \frac{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kR)\frac{\sin(kr)}{kr}d\ln k}{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kr_8)d\ln k} \tag{1.82}$$

where $\hat{W}$ is the Fourier transform of the window function. For a spherical Top-Hat filtering function W_{TH} defined as

$$\begin{cases} W_{TH}\left[\frac{|\vec{x}-\vec{x}'|}{R}\right] = \frac{4R^3}{3\pi} & \text{if } |\vec{x}-\vec{x}'| \leq R \\ W_{TH}\left[\frac{|\vec{x}-\vec{x}'|}{R}\right] = 0 & \text{if } |\vec{x}-\vec{x}'| > R \end{cases},$$

where R si the radius of the corresponding sphere and called smoothing radius. Note that the window function is normalised such that $\int W_{TH}(|\vec{x}-\vec{x}'|/R)\mathrm{d}^3\vec{x} \equiv 1$ for any position $\vec{x}'$. Its corresponding Fourier transform is given by

$$\hat{W}(kR) = \frac{3}{(kR)^3}\{\sin(kR) - kR\cos(kR).\} \tag{1.83}$$

In Figure 1.4 I show the scaling of the smoothed 2-point correlation function (as a function of both R and r) at two different cosmic epochs. There is an overall qualitative resemblance between the r dependence of the 2-point correlation function $\xi(r)$ and the R dependence of its smoothed version $\xi_R(r)$. More interestingly, the characteristic non-monotonic scaling induced by the baryon acoustic oscillations (BAO) that are frozen in the large scale matter distribution survives to the smoothing procedure and stands out also in the second order correlator as soon as r approaches $\sim 100h^{-1}$Mpc.

Using linear theory, therefore, it is possible to relate the second order statistics of matter fluctuations with some characteristic parameters of the primordial universe such as the spectral index n_s, the normalization factor A, or the cosmological parameters of the transfer function (essentially, the physical matter densities $\Omega_m h^2$ and $\Omega_b h^2$ (Eisenstein & Hu, 1998). However, while the knowledge of the power spectrum can characterize the fluctuations at decoupling (when fluctuations are Gaussian), the matter perturbation at late epochs (that is at low redshift) cannot be fully characterized in terms of the correlation function. If we want to understand how the non-linear action of gravity shaped the present day large-scale structure, we must then analyze the whole hierarchy of moments of the mass field.

1.3.8 The correlation hierarchy of smoothed fields predicted by the WNLPT

In this section, I review how the WNLPT allows to obtain, in a coherent and rigorous way, the correlation hierarchy (that I derived above using the HA). To this purpose, I will follow the original analysis proposed by Bernardeau (1992) and then generalized to the smoothed matter field in Bernardeau (1994a). Note that the later approach provides a natural way to include the effects of filtering, therefore it allows me to generalize the

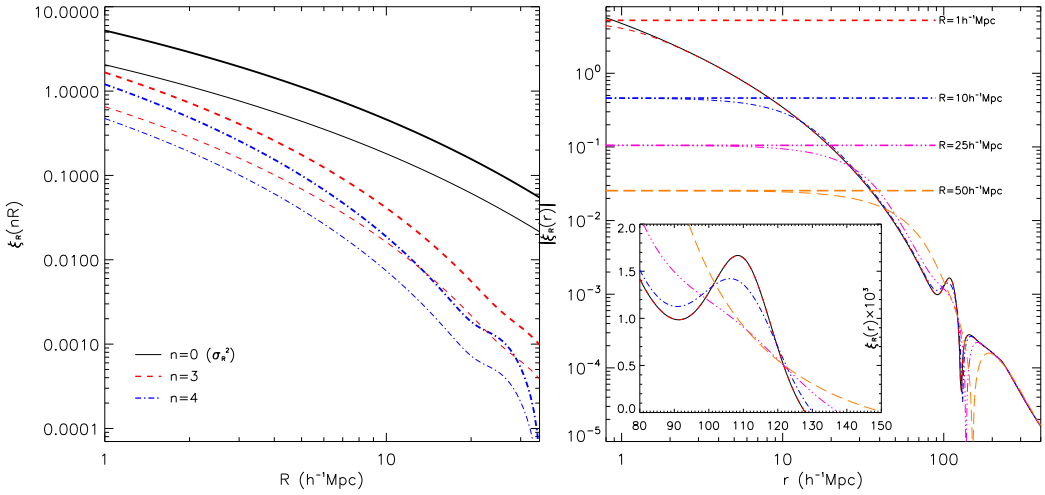

Figure 1.4: *Left:* the 2-point correlation function of the smoothed mass density field ξ_R is shown as a function of the smoothing scale R. We plot ξ_R for different values of the correlation length $r = nR$ i.e. $n = 0$ (black solid lines), $n = 3$ (red dashed lines), and $n = 4$ (blue dot-dashed lines) and at two different redshifts: $z = 0$ (thick lines) and $z = 1$ (thin lines). We adopt the linear power spectrum model of Eisenstein & Hu (1998) and we assume the following set of cosmological parameters: $\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, $H_0 = 72$ km s^{-1} Mpc^{-1}, $\sigma_8(0) = 0.79$, $\Omega_b = 0.044$ and the spectral index $n_s = 0.96$. Note that, for $n = 0$, $\xi_R(0) = \sigma_R^2$, while for a correlation length that goes to infinity the correlation function of the smoothed density field tends to zero. The characteristic bump induced by the baryon acoustic oscillations becomes clearly visible as soon as n is large enough. *Right:* the 2-point correlation function $\xi_R(r)$ is shown as a function of the correlation length r for fields smoothed on different scales, that is $R = 1, 10, 25$ and $50h^{-1}$Mpc. As the smoothing scale R tends to zero, one recovers the 2-point correlation function of matter particles, while as the filtering scale becomes larger, the BAO peak is progressively suppressed. Note also that for $r \to 0$, the value of $\xi_R(r)$ saturates to the variance of the field on the given scale R. On the opposite sense ($r \to \infty$) all the curves converge to the value of the 2-point correlation function of matter particles.

notion of reduced cumulants and correlators to smoothed fields.

Here I show how one can deal with the third order hierarchy of one-point statistics, whereas in Appendix (C) I present the third order two-point hierarchy. In section 1.2.2 I described how to calculate the second order (Fourier) contribution to the density contrast (eq. 1.37) when the matter field is not smoothed. In Appendix (C) I show that if the density contrast field is filtered on a given scale R, then the second order contribution predicted by the WNLPT can be written in real space as

$$\delta_R^{(2)}(\vec{x}) = \iint F_2\left(\vec{q}_1, \vec{q}_2\right) \delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{q}_2)\hat{W}[|\vec{q}_1 + \vec{q}_2|R]e^{i(\vec{q}_1+\vec{q}_2)\cdot\vec{x}}\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2, \tag{1.84}$$

where $\hat{W}$ is the Fourier transform of the filtering function defined in (1.83). This expression can be generalised at order n (Bernardeau, 1994b)

$$\delta_R^{(2)}(\vec{x}) = \iint F_n\left(\vec{q}_1, ..., \vec{q}_n\right) \delta^{(1)}(\vec{q}_1)...\delta^{(1)}(\vec{q}_n)\hat{W}[|\vec{q}_1 + ... + \vec{q}_n|R]e^{i(\vec{q}_1+...+\vec{q}_n)\cdot\vec{x}}\mathrm{d}^3\vec{q}_1...\mathrm{d}^3\vec{q}_n, \tag{1.85}$$

thus the smoothed density contrast is obtained using the series

$$\delta_R(\vec{x}) = \sum_{n=1}^{\infty} \underbrace{D^n(t)\delta_R^{(2)}(\vec{x})}_{\equiv\delta_n(\vec{x})}. \tag{1.86}$$

From eq. 1.86, one can express the one-point moment of order N

$$\langle\delta_R^N\rangle = \langle(\delta_1 + \delta_2 + ... + \delta_n + ...)^N\rangle.$$

By taking $N = 3$, we can express the cumulant moment of order 3 ($\langle\delta^3\rangle_c = \langle\delta^3\rangle$) and obtain

$$\langle\delta_R^3\rangle = \langle\delta_1^3\rangle_c + 3\langle\delta_1^2\delta_2\rangle + ...,$$

where terms have been sorted by order of magnitude. Note that, since the initial condition are assumed to be Gaussian, the first term is equal to zero

$$\langle\delta_R^3\rangle \simeq 3\langle\delta_1^2\delta_2\rangle.$$

Calculation presented in appendix (C) shows that

$$\langle\delta_R^3\rangle = \left(\frac{34}{7} + \gamma_R\right)\langle\delta_R^2\rangle^2, \tag{1.87}$$

where $\gamma_R \equiv \frac{\mathrm{d}\ln\sigma_R^2}{\mathrm{d}\ln R}$. Note that if the field is not smoothed ($\gamma_R = 0$) this result verify the hierarchical ansatz ($\langle\delta^3\rangle \propto \langle\delta^2\rangle^2$). However, if the density field is smoothed, the hierarchical ansatz is no longer a good approximation because of the presence of the

logarithmic derivative of the variance, i.e. of a scale-dependent term. Anyway, it is still convenient to describe the N-point correlation function using the hierarchical ansatz formalism, that is, to assume that the reduced amplitudes are not constants and vary with respect to the considered topology. This way we can still define the reduced cumulant $S_{N,R}$ and the correlators $C_{nm,R}$. Anyway As a result, the correlators must exhibit a dependency on both the smoothing radius R and the correlation length r.

1.3.9 The correlation amplitudes of smoothed fields predicted by the WNLPT

For a top-hat filter, $\delta_R(\mathbf{x})$ is just the volume average of the density contrast over a sphere of radius R. Note that the smoothed correlators of order $N = (n, m)$ retain some of the information contained in the N^{th} order correlation function. As a matter of fact,

$$\kappa_{nm,R} = \frac{1}{V_R^{n+m}} \int_{V_R(\mathbf{x}_1)} d\mathbf{y}_1 ... d\mathbf{y}_n \int_{V_R(\mathbf{x}_2)} d\mathbf{y}_{n+1} ... d\mathbf{y}_{n+m} \xi_{n+m}(\mathbf{y}_1, ..., \mathbf{y}_{n+m}) \tag{1.88}$$

where we assumed a top-hat filter of volume V_R. From a physical point of view $\kappa_{nm,R}$ is the average of the correlation function of order $n+m$ over two distinct volumes separated by $|\mathbf{x}_1 - \mathbf{x}_2|$.

Computing the amplitude of the smoothed reduced cumulants and correlators at the next order (i.e. $S_{3,R}$ and $C_{12,R}$) requires results from the weakly non-linear perturbation theory. If the primordial mass field is Gaussian and fluctuations with wavelength $<< R$ are suppressed using a top-hat filter, then the third order reduced moment, which is often referred to as the skewness of the density field, is (Juszkiewicz, Bouchet & Colombi , 1993; Bernardeau, 1994b)

$$S_{3,R} = \frac{34}{7} + \gamma_R \tag{1.89}$$

while, in the LS limit, that is for separations $r >> R$, the reduced correlator of the same order is (Bernardeau, 1996)

$$C_{12,R}(r) = \frac{68}{21} + \frac{1}{3}\gamma_R + \frac{1}{3}\beta_R(r). \tag{1.90}$$

The complete calculation of the correlator $C_{12,R}$ is detailed in appendix (C). As stressed before, the effect of filtering is to introduce additional, scale-dependent, coefficients

$$\gamma_R \equiv \frac{d \log \sigma_R^2}{d \log R} \tag{1.91}$$

$$\beta_R(r) \equiv \frac{d \log \xi_R(r)}{d \log R}, \tag{1.92}$$

such that the reduced moment $S_{3,R}$ effectively depends on the local slope of the linear power spectrum of density fluctuations (decreasing with the slope of the power spectrum),

while the reduced correlator $C_{12,R}$ acquires a specific and characteristic non-local dependence. In the following we parameterize the distance between the centers of independent smoothing spheres as $r = nR$, where n is a generic real parameter (usually taken, without loss of generality, to be an integer). According to the analysis of Bernardeau (1996), the LS regime is fairly well recovered as soon as $n \geq 3$. Interestingly, since WNLPT results hold for large R, such a small n defines a separation scale r that is already accessible using current redshift surveys such as the Sloan Digital Sky Survey.

The β_R contribution in eq. (1.92) is usually neglected (Bernardeau, 1996) since, in the LS limit, $\xi_R(r)$ is simply the 2-point correlation function of the un-filtered field, a function that effectively vanishes for large separations. This is a critical simplification and the domain of its validity deserves more in-depth analysis. By assuming a power-law spectrum of effective index $n_e = -1.2$ (i.e. $\gamma_R = -(n_e + 3) = -1.8$), we obtain that, on all scales R, the amplitude of $\beta_R(nR)$ becomes negligible (≤ 0.08) as soon as $n \geq 3$. Anyway this rapid convergence to zero is a peculiar characteristic of a scale-free power spectrum. If we consider a more realistic power spectrum (cfr. eq. 1.55) on scales that are accessible to both semi-linear theory and current large scale data (i.e. $10 < R < 30h^{-1}$ Mpc, and $n \sim 3$), the amplitude of the β_R contribution is still significant and varies non monotonically as a function of the length scales r on which the cell correlation is estimated. This is illustrated in Figure 1.5 where we contrast the scaling of $\beta_R(nR)$ and γ_R for different vales of R and n. The systematic error in the estimation of C_{12} that is induced by neglecting the β_R-term on relevant cosmological scales, is larger than the error with which this statistics can already be estimated from current data (see Figure 3.3 in chapter CM). For example, the amplitude of β_R, for characteristic values $n = 3$ and $R = 10(/25)h^{-1}$Mpc, is $\sim 15(/30)\%$ that of γ_R. Interestingly, one can see that, as for γ_R, also the value of $\beta_R(r)$ does not depend on cosmic time, at least at linear order. This redshift-independence follows immediately from eqs. (1.80) and (1.92).

Notice, finally, that WNLPT theory results are expected to hold in the correlation length range in which C_{12} can be unambiguously defined, that is up to the scale where the correlator k_{11} crosses zero. This requirement sets an upper limit to the effective correlation scale n that can be investigated using predictions of WNLPT. In this large scale context, also notice that non-linear effects contributing to the baryon acoustic peak in the the 2-point correlation function, would modify predictions obtained on the basis of the simple linear model of eq. (1.79). For these reasons we limit the present analysis to correlation scales $nR \leq 100h^{-1}$ Mpc.

1.4 Galaxy Bias

The relative simplicity with which fundamental predictions about amplitude and scaling of relevant clustering statistics can be obtained from first principles, must not make us overlook the fundamental difficulty that hampers large scale structure studies. This is

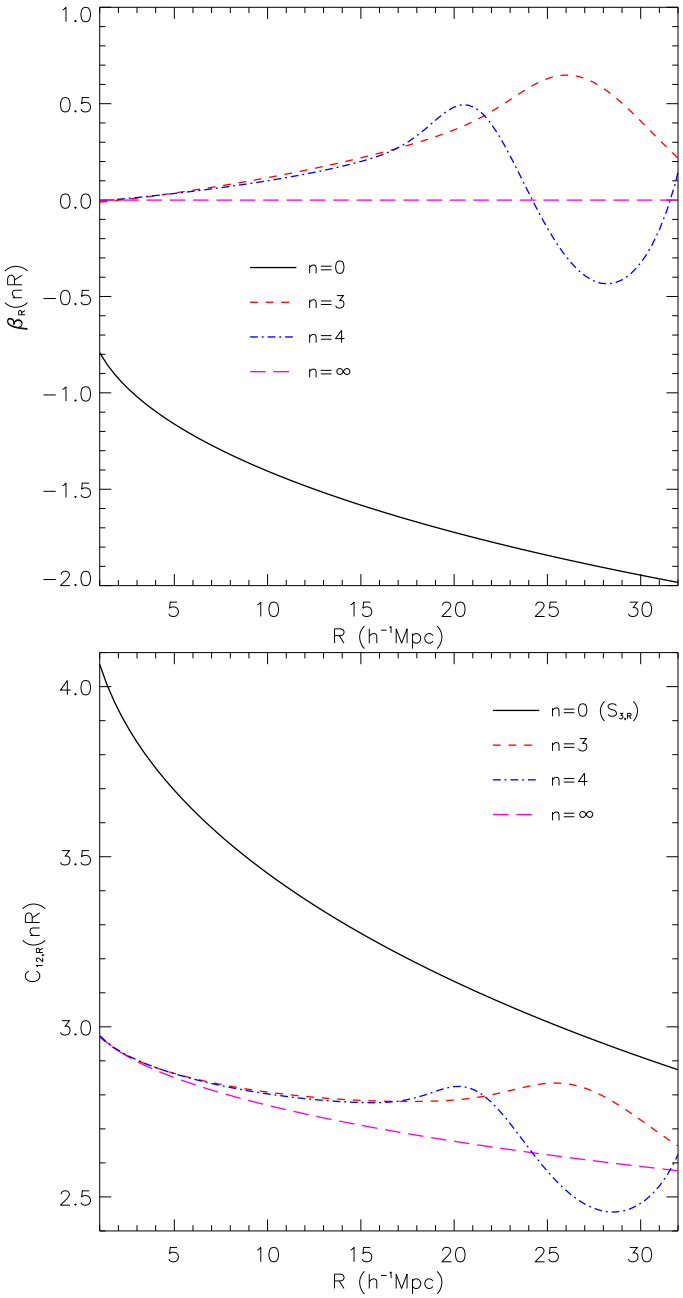

Figure 1.5: *Left :* the logarithmic derivative of $\xi_R(r)$ with respect to R (cfr. eq. 1.92) is shown as a function of the smoothing scale R for different values of the correlation length $r = nR$ (i.e. for $n = 0$ solid line, $n = 3$ short-dashed line, $n = 4$ dot-dashed line and $n = \infty$ long-dashed line). We have assumed the linear power spectrum model of Eisenstein & Hu (1998) and the same parameters value listed in the caption of Figure 1. The non-monotonic scaling of $\beta_R(nR)$ at a scale R$\sim 25h^{-1}$Mpc is induced by the baryon acoustic oscillations. Note that, in the linear regime, $\beta_R(r)$ does not depend on the cosmic epoch, i.e. it is a redshift independent quantity. *Right :* the scaling of the reduced correlator of order 3 is shown as function of the smoothing scale R for different values of the correlation length $r = nR$. Note that for $n = 0$ the correlator reduces to the skewness $S_{3,R}$ and that for $n = \infty$ the expression of $C_{12,R}(nR)$ reduces to the one adopted by Bernardeau (1996).

the fact that the perfect, continuous fluid in terms of which we model the large scale distribution of matter cannot be directly 'observed'. As an example, let's imagine that we are able to locate in the universe all existing galaxies (which is far from being true!) and that we know with an infinite precision their masses. Without any knowledge of how luminous galaxies trace the underlying continuous distribution of matter, even this ultimate galaxy sample would be of little use. Galaxies are intrinsically biased tracer of the matter content of the universe, and the problem of unveiling how mass and galaxy fields map into each others is the so called cosmological biasing issue. If not properly addressed, biasing might bias the theoretical inferences that we draw from data.

There are essentially two different approaches to deal with the biasing issue. One approach focuses on the local physics responsible for the formation of galaxies within dark matter haloes. Once this physics is known, the biasing scheme can be determined and measurements can be compared to theoretical predictions. An alternative approach consists in parameterizing our ignorance about biasing via some parameters (the bias coefficients), and treat them as nuisance parameters in the analysis.

We now discuss in more detail this second way of tackling the biasing issue. An operational definition of bias, that is useful for investigating the hierarchical clustering of matter, is conventionally given in terms of continuous density fields, provided that the galaxy distribution is smoothed on scales R large enough compared to those where non-gravitational physics operates. One can thus expand the dimensionless density fluctuations of galaxies $\delta_{g,R}(\mathbf{x})$ at position $\mathbf{x}$ in Taylor series of the underlying mass over density δ_R at the same point

$$\delta_{g,R}(\mathbf{x}) = \sum_{i=0}^{N} \frac{b_i}{i!} \delta_R^i(\mathbf{x}) \tag{1.93}$$

where b_i are the bias coefficients. It has been shown that, in this large scale limit, such a local transformation preserves the hierarchical properties of 1-point matter statistics (Fry & Gaztañaga, 1993).

The problem of interfacing theory (mass) with observations (light) stems from the fact that the equations that give access to the value of fundamental gravitational quantities (such as the *rms* of linear matter fluctuations on a given scale R ($\sigma_R = \sqrt{\langle \delta_R^2 \rangle}$) or the growth rate of linear perturbations $f = d\ln\delta/\ln a$, where a is the cosmic scale factor) are also the very same equations that allow us to extract the value of the bias parameters b_i. Since the relevant physical and cosmological quantities are generally degenerate with the bias parameters, it is not immediately obvious how to fix their values. Because of this, the traditional approach consists in assuming that gravitational and cosmological parameters are independently known and to fix the amplitude of the bias coefficients b_i (e.g. Lahav et al., 2002).

The viability of the opposite route, that is investigating the coherence of the physical

model given an *a-priori* knowledge of the bias function, has been systematically explored only recently. Zhang et al. (2007) and Song & Percival (2009), for example, have proposed to assess the soundness of GR by constructing statistical indicators that are in principle insensitive to the linear biasing parameter, i.e. the lower order term in equation (1.93). Guzzo et al. (2008), on the contrary, tested GR by comparing the observed and predicted growth rate of matter fluctuations using data from deep redshift surveys. To fulfill their goal they adopted the value of the linear biasing parameter provided by independent observations, such as the level of anisotropy in the Cosmic Microwave Background (Komatsu et al., 2011) or the mean number density of galaxy clusters (Borgani et al., 2001; Schuecker et al., 2003). Both these strategies suffer from the fact that there are now convincing evidences about the non-linear character of the bias function (Marinoni et al., 2005; Gaztañaga et al., 2005; Marinoni et al., 2008; Kovac et al., 2009). This testing scheme is also far from being economic, requiring data from multiple and independent probes of the large scale structure, redshift surveys, imaging surveys, CMB observations.

An orthogonal, more general approach, aims at extracting from redshift surveys both the value of σ_R and the bias parameters b_i. Several authors have shown that, if the initial perturbations are Gaussian and if the shape of third order statistics such as the reduced skewness S_3 (Gaztañaga, 1994; Gaztañaga & Frieman, 1994), the bispectrum (Fry, 1994; Scoccimarro, 1998; Feldman et al., 2001; Verde et al., 2002) or the 3-point correlation function (Gaztañaga et al., 2005; Gaztañaga & Scoccimarro, 2005; Pan & Szapudi, 2005) is correctly described by results of the weakly non-linear perturbation theory, then one can fix the amplitude of b_i up to order 2 in a way that is independent from the overall amplitude of clustering (e.g. σ_8) and depends only on the shape of the linear power spectrum.

Chapter 2

The scale of cosmic isotropy

2.1 Introduction

The cosmological principle (CP), the assertion that the cosmic mass distribution appears homogeneous and isotropic, that is uniform, to a family of typical observers that move with the same average velocity of the surrounding matter (fundamental or comoving observers) has far reaching consequences in cosmology (Wienberg, 1972). It entails that the geometry of space-time is highly symmetric and completely described by the simple Robertson & Walker metric (Robertson, 1929; Walker, 1936). Furthermore, it implies that space expands at a rate that is set by the equations of Friedman & Lemaitre (Friedmann, 1922; Lemaitre, 1931).

The very first surveys of the three-dimensional distribution of optical galaxies showed that the topology of the large-scale structure is very complex and irregular (Geller & Huchra, 1989; Gionannelli & Haynes, 1991). Because of this departure from exact uniformity, the CP is regarded as a coarse-grained model of the universe, a statistical description of the mass distribution that applies only on sufficiently large scales where the finest details of the galaxy clustering pattern become irrelevant.

More recently, two-dimensional observations of the Cosmic Microwave Background (CMB) (Bennett et al., 1994) have shown that the universe is extremely isotropic about us (to roughly 1 part in 10^5), confirming earlier claims based on the analysis of the spatial distribution of local ($z \sim 1$) sources (e.g. (Gregory & Condon, 1991)). What is challenging is to show that the universe is isotropic also about distant observers. As difficult as it may seem, it is important to attack the problem. Indeed, while isotropy at a specific position does not imply cosmic homogeneity (and *viceversa*), isotropy about every fundamental observer does imply overall homogeneity (Ehlers, 1993). Lacking direct evidence for everywhere isotropy, the case for the CP rests more on philosophical rather than on empirical evidences; it is enough to postulate that we are not privileged observers (the so called Copernican principle) to deduce that if the universe appear isotropic about our position, it must also appear isotropic to observers in other galaxies (Ellis, 1976).

The tremendous explanatory power of the standard model of cosmology cannot be advocated as an indirect demonstration of the CP, since they are not the only solutions of the Einstein equations which are able to fit cosmological observations. In particular, many authors have speculated that some effects of the accelerated expansion of the universe (Riess et al., 1998; Perlmutter et al., 1999; Bernardis et al., 2000; Eisenstein et al., 2005; Marinoni & Buzzi, 2010), which remains fundamentally unexplained in terms of microscopic physics, could be mimicked by allowing CP violations (see a review in Célérier, 2007). This intriguing possibility has motivated recent attempts of rooting the CP on a more solid basis. Interestingly, there are some encouraging proposals in this direction which are based on the analysis of the large-scale maps of CMB anisotropies (Goodman, 1995; Caldwell & Stebbins, 2008; Jia & Zhang, 2008; Zhang & Stebbins, 2011), of galaxies (Clarkson, Bassett & Lu, 2008; Uzan, Clarkson & Ellis, 2008; Romano, 2007; Bolejko & Wyithe, 2009) and of supernovae (Clifton, Ferreira & Land, 2008).

Even if we postulate the CP, the picture is not complete unless we identify the averaging scale that is implicit in this assumption, i.e. the scale on which the FLRW model provides an effective, coarse-grained description of the universe (Stoeger, Maartens & Ellis, 1995). It is generically asserted that the CP holds on domains that are large enough to encompass the biggest gravitational structures of the universe. Yet, few studies have attempted to narrow in on the length value above which clumpiness gives way to uniformity (Wu, Lahav & Rees, 1999).

Past efforts were mostly based on the analysis of the two-point correlation properties of galaxy samples (Davis, 1997; Guzzo, 1997). This approach, however, suffers from severe theoretical drawbacks. Since the average number density of the sample is needed as input, the method presupposes the premise to be tested, i.e. a constant density distribution of matter (Pietronero, Montuori & Sylos-Labini, 2007). Moreover, it does not provide an unambiguous definition of the cross-over scale (Gaite et al., 1999). As a consequence, the inferred homogeneity length-scales depend on the size of the analyzed sample and range from values as low as $30h^{-1}$Mpc up to $200h^{-1}$Mpc (Borgani, 1995; Cappi et al., 1998; Amendola & Palladino, 1999; Bharadwaj et al., 1999; Martinez et al., 2001; Pan & Coles, 2001). More recently, orthogonal techniques have been explored which are based on the count-in-cells analysis of observations confined to a spatial hyper-surface of constant time (e.g. Bagla, Yadav & Seshadri, 2005). These methods are insensitive to light cone effects, i.e. possible biases arising from comparing galaxy fluctuations at different cosmic epochs, and seem to indicate a transition to homogeneity at a scale of 70^{-1}Mpc (Hogg, 2005; Yadav et al., 2005) but see Sylos-Labini (2009a,b) for an opposite conclusion. In particular, the counting method advocated by Scrimgeour et al. (2012) allows to estimate the homogeneity scale independently from the sample size.

It is widely believed that, since we cannot point telescopes from any other place but the solar system, it is not possible to establish if also distant observers see an isotropic universe. While this argument is certainly true for apparent 2D quantities such as, for

example, the CMB temperature (but see Goodman, 1995), we show here that it does not apply to 3D maps of the spatial distribution of galaxies. Specifically, we quantify the typical dimension above which independent observers see an isotropic 'bath' of galaxies. Besides establishing an operational definition of the isotropy scale, our approach also provides an overall consistency test of a fundamental facet of the CP, i.e. that we are not privileged observers of the universe.

2.2 The Method

I identify paths of external length irradiating from a given arbitrary target galaxy to every other n^{th} closest neighbours (see Fig. 2.1). The amplitude of the angle t between these directions and the observer line-of-sight (*los*) to the target is computed by assuming that the local properties of a homogeneous and isotropic universe are described in terms of the infinitesimal Robertson & Walker (Robertson, 1929; Walker, 1936) line element

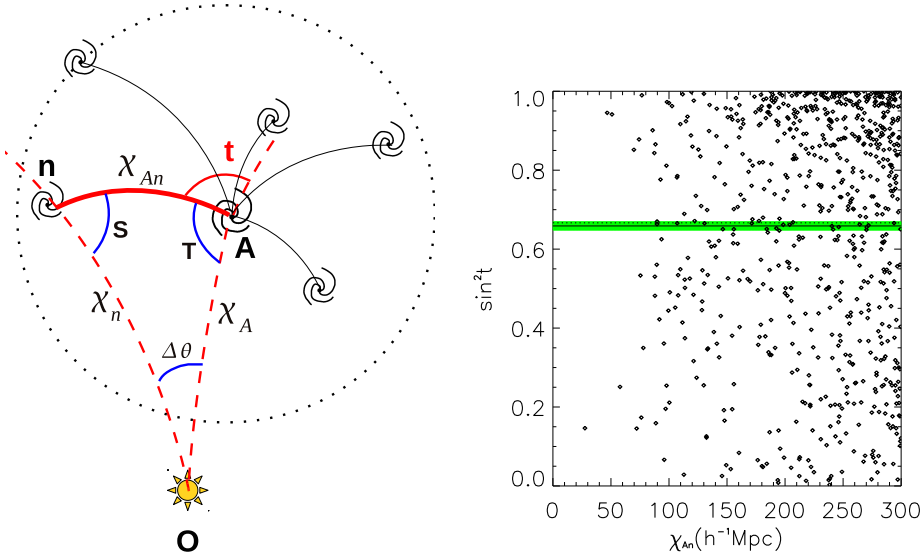

Figure 2.1: *Upper:* We determine the geodesic connections between a given target galaxy A and all the surrounding galaxies that lie inside a sphere of radius R centred on A. The target galaxy A and its n^{th} closest neighbour subtend an angle $\Delta\theta$ at the observer position O. The tilting angle t measures the inclination of the geodesic separation χ_{An} between A and n with respect to the observer line-of-sight (dashed line). If the CP holds on the scale R we expect this '*spaghetti* to be isotropically oriented about any given target. In other terms we expect the *los* angle t to be isotropically distributed, i.e. its PDF is $\varphi(t) = (\sin t)/2$. *Lower:* the distribution of sin^2t as a function of the geodesic separation χ_{An} in a sphere of radius $R = 300h^{-1}$Mpc randomly positioned in the LRG SDSS sample. The dotted line represents the theoretically expected average ($\mu = 2/3$), while the solid line represents the IGI value, that is the average of the plotted points. The shaded area represents the 1σ uncertainty of the IGI value.

$$ds^2 = (cdt)^2 - a^2(t)[\mathrm{d}\chi^2 + \Sigma_k^2(\chi)(\mathrm{d}\theta^2 + \sin^2\theta d\phi^2)]$$

where, c is the light speed, k is the scalar spatial curvature, χ is the radial geodesic comoving distance, $a(t)$ is the cosmic expansion factor, and where, using the Kronecker symbol, $S_k(\chi) = \delta_{k,1}\sin\chi + \delta_{k,0}\chi + \delta_{k,-1}\sinh\chi$. The spatial part of the metric is invariant under a *quasi-translation* transformation of its coordinates (Wienberg, 1972). We can translate the reference frame from the terrestrial observer O to the target A and express the coordinates $\vec{x}_{n/A}$ of its n^{th} neighbor as

$$\begin{aligned}\vec{x}_{n/A} &= \vec{x}_{n/O} - \vec{x}_{A/O}\Bigg\{\left[1 - kx^2_{n/O}\right]^{1/2} + \\ &+ \left[1 - (1 - kx^2_{A/O})^{1/2}\right]\frac{\vec{x}_{n/O}\cdot\vec{x}_{A/O}}{x^2_{A/O}}\Bigg\}. \end{aligned} \tag{2.1}$$

By orienting the axes in such a way to minimize the number of non-zero components (I choose $\vec{x}_{n/A} = (S_k(\chi_{An})\ ,\ 0,\ 0)$, $\vec{x}_{n/O} = (S_k(\chi_n)\sin\Delta\theta\ ,\ 0,\ S_k(\chi_n)\cos\Delta\theta)$ and $\vec{x}_{A/O} = (0\ ,\ 0,\ -S_k(\chi_A))$), and by exploiting the identity

$$C^2_k(\chi) + kS^2_k(\chi) = 1 \tag{2.2}$$

we obtain

$$\begin{aligned}S^2_k(\chi_{An}) &= S^2_k(\chi_n)\sin^2\Delta\theta + \\ &+ \left[S_k(\chi_n)C_k(\chi_A)\cos\Delta\theta - S_k(\chi_A)C_k(\chi_n)\right]^2. \end{aligned} \tag{2.3}$$

Further, by repeatedly applying this relation to the 3 edges of the triangle $(\hat{S}, \hat{T}, \hat{\Delta\theta})$ shown in Fig. 2.1, and by isolating, after some algebra, identical terms in the resulting expressions one obtains the generalized law of sinus

$$\frac{\sin\Delta\theta}{S_k(\chi_{An})} = \frac{\sin T}{S_k(\chi_n)} = \frac{\sin S}{S_k(\chi_A)}. \tag{2.4}$$

Since $t = \pi - T$, it finally follows from eqs (2.3) and (2.4) that

$$\sin^2 t = \frac{1}{1 + \left[C_k(\chi_A)\ \cot\Delta\theta - \frac{S_k(\chi_A)}{S_k(\chi_n)}\frac{C_k(\chi_n)}{\sin\Delta\theta}\right]^2}. \tag{2.5}$$

If the CP holds, the *los* angle t has a comoving space Probability Distribution Function (PDF) of a characteristic type ($\varphi(t) = (\sin t)/2$), namely, it is a random variable isotropically distributed with respect to any fundamental observer. Therefore, the expectation value $\mu = \langle\sin^2 t\rangle$ is cosmology independent and equal to 2/3. I define the indicator of

galaxy isotropy (IGI) as the estimator m_R constructed by averaging equation (2.5) over n galaxies inside a sphere of comoving radius R that is centred around any given observer in the universe. On scales R where the CP applies, we expect the measure of m_R to converge to the predicted value $\mu = 2/3$.

The testing protocol is as follows: I assume that the CP holds and I implement standard statistical inference methods to try to falsify it and reject its validity. In detail, I assume the existence of a length-scale R above which the empirical IGI estimates (m_R) are statistically identical to the theoretical prediction (μ). I thus formulate a null hypothesis h_0 according to which the two quantities are not different. I quantify the goodness of the agreement by means of χ^2 statistics, and, following standard convention, I fix the rejection threshold of h_0, i.e. the risk of reaching the wrong conclusion, at the 5% level. This means that the hypothesis that the universe is isotropic above a scale R cannot be rejected by data if the probability P of obtaining a worse (larger) χ^2 value is greater than 5%. On the contrary, an eventual failure in identifying the scale of isotropy would unambiguously point at the incoherence of the FLRW model.

Homogeneity and isotropy are properties that characterize the large-scale distribution of *matter* on a 3D spatial hyper-surface at a given instant of time. Since light propagates at a finite speed, the most distant regions of the 3D volume directly accessible to observations are also the furthest in time. As a consequence, the number density of galaxies, an observable that is modulated by local physical processes with their own specific time-scales, is expected to vary as a function of distance. This is a known issue that hampers most of the tests of the CP (Heavens, Jimenez & Maartens, 2011). In the following we show that, by focusing our attention on the angular distribution of galaxies, instead of their number density fluctuations, we can tackle the past light cone issue.

If the CP holds true, as I assume here, the galaxy spatial number density $\rho_s(r)$ within spherical shells of width Δr centered on the terrestrial observer must be independent from the distance r. Note that *shell-homogeneity*, that is $\rho_s = const$, does not imply homogeneity, i.e. invariance under general spatial translations, while the opposite is true. More importantly, the radial constancy of ρ_s does not imply everywhere isotropy (isotropy about arbitrary comoving observers) that is the fundamental facet of the CP that we want to test. As a matter of fact, the distribution of galaxies that surrounds us can be characterized by a constant ρ_s and yet be anisotropic. I therefore remove past-light cone artifacts, by imposing that the comoving number density of galaxies be strictly constant within concentric shells centered on us. In practice, I analyze a volume limited catalog of galaxies, that is a sample of objects brigther than a given minimum absolute luminosity, and I additionally remove, with a random rejection process, any residual radial gradient in the distribution of galaxies. As I show in the following, this technique preserves the clustering properties of the galaxy distribution, and does not falsely impose homogeneity where there is none.

The fair sample model of the universe (Layzer 1956), assumes that the galaxy distribu-

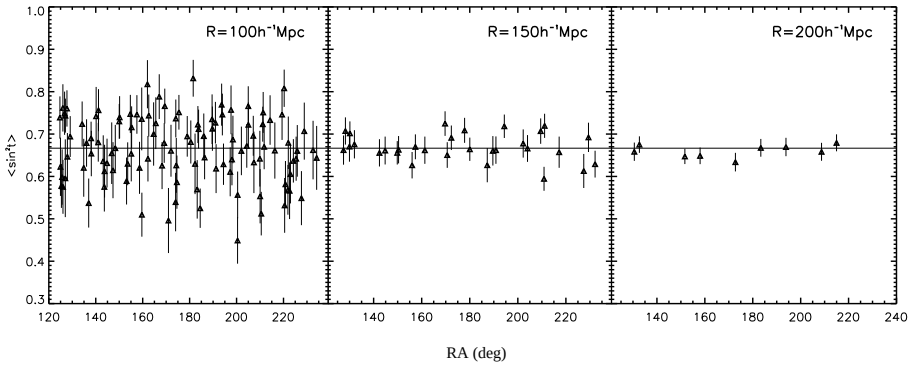

Figure 2.2: The IGI value measured by different observers, labeled by their right ascension coordinate, is plotted. Each estimate is performed within non-overlapping spheres of comoving radius R (shown in the inset) randomly thrown in the volume covered by the SDSS galaxy survey. Errors are computed as the standard deviation of the mean, and well trace the theoretically expected figure $\sigma = \sqrt{4/(45n)}$. The average IGI value ($m_R = \langle \sin^2 t \rangle$) is $0.678 \pm 0.005, 0.672 \pm 0.006$, and 0.660 ± 0.007 from left to right. The solid line shows the expectation value predicted under the assumption that the CP holds (i.e. $\mu = 2/3$). On small scales data scatters widely, while on scales where the CP is expected to hold, data fit the theoretical prediction. A goodness of fit statistical analysis yields $\chi^2/dof = (2.1, 1.12, 0.65)$ from the the smaller to the larger scale.

tion is a discrete stochastic process resulting from the Poissonian sampling of an underlying continuous matter density field $\Lambda(\mathbf{x})$. Accordingly, a galaxy sample that traces an underlying continuous field of PDF $Q(\Lambda)$ can be modeled as a discrete stochastic process in which the probability of counting N galaxies within a given arbitrary cell is

$$P_N = \int P(N|\Lambda) Q(\Lambda) d\Lambda, \tag{2.6}$$

where the conditional sampling probability is

$$P(N|\Lambda) = \frac{\Lambda^{-N}}{N!} e^{-\Lambda}. \tag{2.7}$$

In the ideal universe described by the FLRW model, the continuous matter fluid $\Lambda(\mathbf{x})$ satisfies exactly the CP and its PDF is the Dirac delta $Q(\Lambda) = \delta^D(\Lambda - \Lambda_0)$. Equations (2.6) and (2.7) therefore imply that P_N follows a Poissonian statistic, that is the spatial distribution of galaxies is random.

I now demonstrate that a spatially random sample of size K cannot be generated by randomly sampling, with probability p, a parent population of size N whose spatial distribution is clustered, that is inhomogeneus. Let the sampling process be described by

$$K = \Phi(\mathbf{x}_1) + \Phi(\mathbf{x}_2) + ... + \Phi(\mathbf{x}_N), \tag{2.8}$$

where Φ is a random variable (taking on the values 0 or 1) distributed according to the Bernoulli probability law (that is $p^{\Phi}(1-p)^{1-\Phi}$, where p is the probability that $\Phi = 1$ in a single trial), and where N is a non-negative integer-valued random variable distributed according to the Poisson distribution of average counts $\bar{N}$.

The probability generating function (PGF) of P_N is

$$\mathcal{G}(z) \equiv \sum_{i=0}^{\infty} P_i z^i, \tag{2.9}$$

and its expression in the case of the Bernoulli and Poisson processes is

$$\mathcal{G}_{\Phi}(z) = (1-p) + pz \tag{2.10}$$

and

$$\mathcal{G}_N(z) = e^{(z-1)\bar{N}}, \tag{2.11}$$

respectively.

It can be shown (Kendall & Stuart, 1983) that the PGF of the sum of independent, identically distributed random variables Φ, i.e. the PGF of the random variable K in equation (2.8) is

$$\mathcal{G}_K(z) = \mathcal{G}_N(\mathcal{G}_{\Phi}(z)). \tag{2.12}$$

Let's assume that $\mathcal{G}_K$ is the PGF of a Poissonian distribution with average parameter $\bar{n}$. By taking the derivative of eq. (2.12) we obtain

$$\bar{n} e^{\left[\frac{(\mathcal{G}_{\Phi}+p-1)}{p}-1\right]\bar{n}} = \frac{d\mathcal{G}_N}{d\mathcal{G}_{\Phi}} p \tag{2.13}$$

which, upon integration gives

$$\mathcal{G}_N(u) = e^{(u-1)\frac{\bar{n}}{p}} + C. \tag{2.14}$$

The value of the arbitrary constant C can be set to 0 using the additional condition $\mathcal{G}_N(1) = 1$ (see eq. (2.9)). Therefore, a spatially random distribution cannot be the result of the random sampling of a non-Poissonian distribution. Reciprocally, one can show that a random sampling of a Poisson parent distribution, results in a subsample of elements with Poisson distribution.

I now show that also the N-point moments of the continuous mass distribution $\Lambda(\mathbf{x})$ are not modified by a random sampling process. Given the counts $N_i \equiv N(\mathbf{x}_i)$ in a cell at position $\mathbf{x}_i$, the N point PGF is immediately obtained by generalizing the expression given in eq. (2.9)

$$G(z_1, z_2, ..., z_n) = \sum_{N_1} \sum_{N_2} ... \sum_{N_n} P_{N_1,N_2,...N_n} z_1^{N_1} z_2^{N_2} ....z_n^{N_n}, \tag{2.15}$$

where

$$P_{N_1,N_2,...,N_n} = \int P(N_1|\Lambda_1)....P(N_n|\Lambda_n))Q(\Lambda_1)...Q(\Lambda_n)d\Lambda_1.....d\Lambda_n, \tag{2.16}$$

and where $\Lambda_i \equiv \Lambda(\mathbf{x}_i)$. If the sampling is random, the N-point PGF is given by

$$G(z_1, z_2, ..., z_n) = \langle e^{\Lambda_1(z_1-1)} e^{\Lambda_2(z_2-1)} ......e^{\Lambda_n(z_n-1)} \rangle. \tag{2.17}$$

The moment generating function associated to the discrete, N-point, counts $N_1, N_2, ..., N_n$ follows immediately by substituting $z_i \to e^{z_i}$ in the argument of the PGF (Szapudi, Szalay & Boschán, 1992)

$$M(z_1, z_2, ..., z_n) = \langle e^{\Lambda_1(e^{z_1}-1)} e^{\Lambda_2(e^{z_2}-1)}, ..., e^{\Lambda_n(e^{z_n}-1)} \rangle. \tag{2.18}$$

The N-point moment of the galaxy distribution, is calculated as the functional derivative of the moment generating function (see section 1.3.3)

$$\langle N(\mathbf{x}_1)N(\mathbf{x}_2).....N(\mathbf{x}_n)\rangle \equiv \frac{\delta^n M}{\delta z_1 \delta z_2 ....\delta z_n}\Big|_{z_1=z_2=.....=z_n=0}. \tag{2.19}$$

By substituting eq. (2.18) into the previous one, we see that the N-point galaxy moment is equivalent to the corresponding statistics computed for the underlying continuous matter field, i.e.

$$\langle N(\mathbf{x}_1)N(\mathbf{x}_2).....N(\mathbf{x}_n)\rangle = \langle \Lambda(\mathbf{x}_1)\Lambda(\mathbf{x}_2).....\Lambda(\mathbf{x}_n)\rangle. \tag{2.20}$$

As a result, the whole hierarchy of N-point galaxy correlation functions computed from a random (discrete) sample trace with fidelity the N-point correlation function of matter. Again, an inhomogeneous spatial distribution cannot be turned into a homogeneous one by a random sampling process.

2.3 Data: the SDSS DR7 sample

I apply the method to the seventh release of the Sloan Digital Sky Survey (Abazajian et al., 2009) which is comprised of $\sim 930,000$ galaxies over a field of view of 9380 deg^2. Our analysis is limited to luminous red galaxies (LRG, Eisenstein et al., 2001) distributed in the North Galactic contiguous area defined by $120 < RA < 240$, $7 < DEC < 56$ (see figure 2.3). A sample with a nearly constant density of galaxies is obtained by volume limiting the SDSS dr7 catalog in the redshift range $0.22 < z < 0.5$. This sample extends on a comoving radial size $\Delta r \sim 700 h^{-1}$Mpc (in what follows, I consider a cosmological model characterized by the reduced density of matter $\Omega_m = 0.27$ and dark energy $\Omega_\Lambda = 0.73$,

and I assume that the value of the Hubble constant is $H_0 = 72$ km s^{-1}Mpc^{-1}.) The upper redshift limit is fixed by the requirement of measuring the IGI with approximately the same average precision of nearly 1% over all the interval $100 < R < 200 Mpc$.

The strict equality $\rho_s(r) = const$ is then imposed by interpolating the observed number density ρ_s of objects in spherical shells centered on us (and with thickness a hundred times smaller than the effective depth of the sample), and by randomly rejecting galaxies using a Monte Carlo process with selection function $\phi(r) = \min(\rho_s)/\rho_s(r)$. The final LRG sample contains a total of ~ 6500 objects, has a mean number density $\rho = 6.14 \cdot 10^{-6} h^3 \mathrm{Mpc}^{-3}$ and covers an effective field of view of 4860 deg^2.

As stressed before an angular window and a radial coverage define the survey geometry. As shown in figure (2.1), the extragalactic observer A have to probe if he is seeing an isotropic distribution of galaxies around him. Thus it is crucial to make sure that this virtual observer won't be able to see any boundaries of the survey which could lead to artificial anisotropies inducing a break down in the testing process. The easiest way of proceeding is to generate a random distribution of points (which materializes the potential observers) within the boundaries of the survey. Then I must verify if each observer is positioned (with respect to any boundary) at a larger distance than the tested scale. As one may guess, this way of proceeding is computationally expensive and since I had to deal with placing spheres into a survey in most of my thesis work, I indeed optimized this process. Basically, the idea is to predict in advance where to place observers in order to avoid any boundary effect. It is straightforward and intuitive to do so in a flat Euclidean space. On the contrary, in curved space this operation is less trivial.

The first step is to define the distance between two points A and B in curved spaces. The most general (i.e. valid for both flat and curved spaces) expression (eq. 2.3) of the normalized comoving distance χ_{AB} between two objects A and B subtending an angle γ at the observer position O (assumed to be the origin of the coordinate system) is

$$\begin{aligned} S_k^2(\chi_{AB}) = \; & C_k^2(\chi_A) S_k^2(\chi_B) + S_k^2(\chi_A) C_k^2(\chi_B) + k S_k^2(\chi_A) S_k^2(\chi_B) \sin^2\gamma \\ & -2 C_k^2(\chi_A) C_k^2(\chi_B) S_k^2(\chi_A) S_k^2(\chi_B) \cos\gamma, \end{aligned} \tag{2.21}$$

where $C_k^2(\chi) + k S_k^2(\chi) = 1$ for any value of k. The expression (2.21) can be written in a more synthetic way as

$$C_k^2(\chi_{AB}) = \left\{ C_k(\chi_A) C_k(\chi_B) + k S_k(\chi_A) S_k(\chi_B) \cos\theta \right\}^2 . \tag{2.22}$$

Now the position of a point (galaxy) M can be defined in terms of the radial comoving distance χ, the right ascension α and the declination δ. It is possible to map this equatorial coordinate system into the standard Cartesian one by applying the following transformations

$$\begin{cases} S_k(x) = S_k(\chi)\cos\delta\cos\alpha \\ S_k(y) = S_k(\chi)\cos\delta\sin\alpha \\ S_k(z) = S_k(\chi)\sin\delta \end{cases}, \tag{2.23}$$

where $(x,\ y,\ z)$ are the Cartesian coordinates of the point M. The relation between the angle γ and the equatorial angular coordinates of the galaxies A and B is

$$\cos\gamma = \cos\delta_A\cos\delta_B\cos(\alpha_A - \alpha_B) + \sin\delta_A\sin\delta_B, \tag{2.24}$$

a relation showing that the separation angle does not depend on the absolute position in right ascension of A or B.

As stressed before, using the distance formula (2.22 or 2.21) it is possible to predict the region of the space unaffected by boundary issues. A boundary can be defined as a surface of constant right ascension α_o, constant declination δ_o or constant radial comoving distance χ_o. Thus, to avoid any boundary effects, I define the distance between any previously cited surface and a generic point M as the distance between this point and a virtual point lying on the boundary surface. This particular virtual point must be chosen as the closest to M. From expression (2.22) I deduce that

$$kS_k^2(\chi_{AB}) = 1 - \left\{C_k(\chi_A)C_k(\chi_B) + kS_k(\chi_A)S_k(\chi_B)\cos\theta\right\}^2, \tag{2.25}$$

and therefore I conclude that minimizing the distance χ is equivalent to finding the maximum/minimum of the right hand side of equation (2.25) which is equivalent as finding the maximum/minimum of

$$f_\delta(\chi,\alpha) \equiv C_k(\chi_M)C_k(\chi) + kS_k(\chi_M)S_k(\chi)\left\{\cos\delta_M\cos\delta_o\cos(\alpha_M - \alpha) + \sin\delta_M\sin\delta_o\right\}, \tag{2.26}$$

in the case of a boundary surface defined by $\delta = \delta_o$. This procedure can be generalized for the two other boundaries as

$$f_\alpha(\chi,\delta) \equiv C_k(\chi_M)C_k(\chi) + kS_k(\chi_M)S_k(\chi)\left\{\cos\delta_M\cos\delta\cos(\alpha_M - \alpha_o) + \sin\delta_M\sin\delta\right\}, \tag{2.27}$$

$$f_\chi(\alpha,\delta) \equiv \cos\delta_M\cos\delta\cos(\alpha - \alpha_M) + \sin\delta_M\sin\delta. \tag{2.28}$$

in the cases of boundary surfaces defined by $\alpha = \alpha_o$ and by $\chi = \chi_o$ respectively. After some algebra, the distance in curved spaces between any boundary surface point and a point with coordinates (χ,α,δ) can be expressed as

$$\begin{aligned} S_k(\chi_{\delta_o}) &= S_k(\chi)\sin|\delta-\delta_o| \\ S_k(\chi_{\alpha_o}) &= S_k(\chi)|\cos\delta\sin(\alpha-\alpha_o)| \ , \\ \chi_{\chi_o} &= |\chi-\chi_o| \end{aligned} \tag{2.29}$$

where χ_{δ_o}, χ_{α_o} and χ_{χ_o} are the comoving distances involving respectively a declination surface, a right ascension surface and a shell surface. The equation (2.29) can therefore be used to calculate the effective boundaries by imposing that for any observer

$$\begin{cases} \chi_{\delta_o} \geq R' \\ \chi_{\alpha_o} \geq R' \\ \chi_{\chi_o} \geq R' \end{cases} ,$$

where R' is the reduced comoving radius of the considered sphere ($R' = R/a_o$). By taking the limiting case $k = 0$ I recover the more intuitive formulas that hold in a flat universe.

2.4 Analysis of SDSS data and Comparison to Theoretical Models

For meaningful error interpretation, it is imperative to acquire independent estimations of the observable m_R. Consequently, I do not apply the scheme to every galaxy in the sample, i.e. I do not carve spheres of comoving radius R around each 'extraterrestrial' observer to determine whether they see the same degree of isotropy. Instead, I only select as observers, those target galaxies that are at the center of non-overlapping spheres. As an example, given the geometry of the largest contiguous volume in the SDSS survey, I can place a maximum number $N = 107, 30, 9, 4, 3$ of independent observers, exploring the isotropic distribution of galaxies on length-scales $R = 100, 150, 200, 250$ and $300h^{-1}$Mpc respectively. Each of these observers are geodesically connected, on average, to $n = 26, 87, 206, 401, 695$ galaxies respectively.

2.4.1 Random catalogs

Before analyzing real data, I have first applied the method to synthetic samples simulating spatially random (statistically uniform) galaxy distributions. The point here is to detect the minimum radius R below which our technique is noise-limited and the scale of everywhere isotropy cannot be resolved. Using Montecarlo techniques, I have generated various uniform mock catalogues with galaxy number densities in the range $10^{-4} - 10^{-6}h^3\text{Mpc}^{-3}$. We have found that, as expected, when the scale R is larger than the mean inter-particle separation $\lambda = \rho^{-1/3}$, the distribution of the t angle statistically converges towards an isotropic PDF. Quantitatively, as soon as $R > 1.5\lambda$, that is when on average $\sim 4 \cdot (1.5)^3$ galaxies are geodesically connected to the observer, the risk of reaching the wrong con-

clusion in rejecting the isotropy hypothesis h_0 becomes larger than 5%. In particular, the everywhere isotropy of a random (uniform Poissonian) distribution of galaxies with the same density of the LRG sample investigated in this study can be unambiguously detected on scales larger than $\sim 85h^{-1}$ Mpc.

I now explain how I generated a random distribution of points in curved spaces. One can immediately note that if a uniform distribution of points in curved spaces is generated using spherical coordinates, then only the radial distribution is affected by curvature. Indeed the curvature has no effect on the projection along the line of sight (i.e. the angular distribution). As discussed above, a light cone is defined by its boundaries in right ascension (α_{min}, α_{max}), in declination (δ_{min}, δ_{max}) and in radial comoving distance ($a_o\chi_{min}$, $a_o\chi_{max}$). Thus the effective solid angle Ω is given by

$$\Omega = (\alpha_{max} - \alpha_{min})(\sin\delta_{max} - \sin\delta_{min}). \tag{2.30}$$

And the total comoving volume can be expressed as

$$V_T = \frac{\Omega}{4} a_o^3 \left\{ V_k(2\chi_{max}) - V_k(2\chi_{min}) \right\}, \tag{2.31}$$

where k is the curvature (-1, 0 or 1), and V_k is defined as

$$V_k(x) \equiv k\left[x - S_k(x)\right] + \frac{x^3}{3}\delta^K(k),$$

where δ^K stands for the Kronecker function, 1 if $k = 0$ otherwise 0. It is therefore possible to predict the total number of points N_T in order to obtain a random sample of a given number density ρ_o

$$N_T = \rho_o V_T. \tag{2.32}$$

Since the total number of points is known, I am now interested in finding the probability density function $P\left[(a_o\chi)\right]$ of radial comoving distances in a curved space uniformly populated. It corresponds to express the probability of finding a given number N_s of points in a shell of thickness $d(a_o\chi)$ and limited by a solid angle Ω. Thus uniformity of the random positions imposes that

$$P\left[(a_o\chi)\right] d(a_o\chi) = \frac{N_s}{N_T}. \tag{2.33}$$

Since the volume of a shell is $dV_s = \Omega a_o^3 S_k^2(\chi) d(\chi)$, the expected number N_s of points in the corresponding shell can be expressed as $N_s = \rho_o dV_s$. Using equations (2.31), (2.32) and (2.34) it is therefore possible to express the density probability of the radial distances as

$$P\left[(a_o\chi)\right] d(a_o\chi) = \frac{4}{a_0} \frac{S_k^2(\chi)}{V_k(2\chi_{max}) - V_k(2\chi_{min})} d(a_o\chi). \tag{2.34}$$

The comoving distances of a random distribution must be distributed according to the probability density $P\left[(a_o\chi)\right]$. However, in practice, it is covenient to obtain the distribution of radial comoving distances from a uniform random generator. Be ξ a uniformly distributed variable between 0 and 1. In order to express the radial distance with respect to ξ one can use the conservation of probabilities $P\left[(a_o\chi)\right] d(a_o\chi) = P\left[\xi\right] d\xi$ and by integrating, it follows

$$4 \int_{\chi_{min}}^{\chi} \frac{S_k^2(x)}{V_k(2\chi_{max}) - V_k(2\chi_{min})} \mathrm{d}x = \int_0^{\xi} \mathrm{d}\xi'.$$

I can thus express $(a_o\chi)$ as

$$(a_o\chi) = \frac{a_o}{2} V_k^{-1} \left[\xi \left\{V_k(2\chi_{max}) - V_k(2\chi_{min})\right\} + V_k(2\chi_{min})\right], \tag{2.35}$$

where V_k^{-1} is the reciprocal function (i.e. $V_k^{-1}[V_k(x)] \equiv x$) of V_k which is defined for any k. Although it can be analytically expressed in the case of $k = 0$, in the cases $k = \pm 1$ it is necessary to solve it numerically. This can be computationally expensive if the number of point is very large.

The right ascensions α and declinations δ of the random points are generated as follows. Paying attention to the conservation of the solid angle under rotation of the right ascensions it is obvious that α must be uniformly distributed (i.e. drawn from a uniform distribution) between α_{min} and α_{max}.

It follows that the right ascencion of randomly distributed points can be drawn from a uniform distribution ξ as

$$\alpha = (\alpha_{max} - \alpha_{min})\xi + \alpha_{min}. \tag{2.36}$$

From the expression of the solid angle Ω (eq. 2.30) it is obvious that a shift in declination does not conserve the solid angle. However, we notice that a translation of the sine of the declination leaves the solid angle unchanged. Thus in order to randomly distribute point in declination, it is the sine of the declination which must be drawn from a uniform distribution

$$\delta = \arcsin[\xi(\sin\delta_{max} - \sin\delta_{min}) + \sin\delta_{min}]. \tag{2.37}$$

From equations (2.35), (2.36) and (2.37) I generated random distributions of points in different cosmological models.

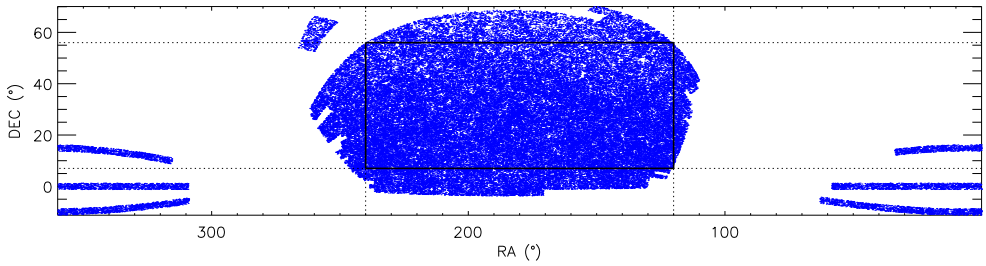

Figure 2.3: Effective angular field extracted from the SDSS northern region, the solid rectangle represents the angular cut we applied in order to simplify the boundaries of the survey.

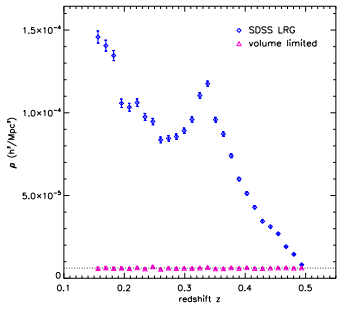

Figure 2.4: Blue diamonds show the density profile of the SDSS LRG sample whereas the magenta triangles show a possible Montecarlo realization of the LRG sample with a constant density profile. The dotted line represents the value $\rho = 6.14 \times 10^{-6} h^3 \mathrm{Mpc}^{-3}$.

2.4.2 SDSS LRG sample

I have then analyzed the luminous red galaxy (LRG) sample extracted from the SDSS dr-7. I first retrieved the whole SDSS DR7 catalog (containing blue and LRGs galaxies) (Strauss et al., 2002) from the SDSS website. I then selected the LRG sample by applying the selection criteria in magnitude and color detailed in section 2 of Eisenstein et al. (2001). I also applied an angular selection (i.e. a cut in Right Ascension and DEClination) in order to select a contiguous volume

$$\begin{aligned} 120^\circ \leq \quad & \mathrm{R}A \quad \leq 240^\circ \\ 7^\circ \leq \quad & \mathrm{D}EC \quad \leq 56^\circ. \end{aligned}$$

Figure (2.3) shows the effective angular cut applied in the northern part of the SDSS survey. The complete sample, referred to as SDSS LRG in figure (2.4), contains 70750 galaxies its radial density profile is provided in figure (2.4).

The IGI value estimated by distant observers on a scale $R = 100, 150$, and $200 h^{-1}$Mpc is graphically shown in Figure 2.2. It is interesting to note that, for any displayed scale R, the distribution of the average IGI values ($\langle m_R \rangle$) peaks at $\mu = 2/3$, while the variance of the distribution decreases as a function of R. The stability of the central value of the distribution shows whether isotropy is present on average, whereas the scatter shows whether

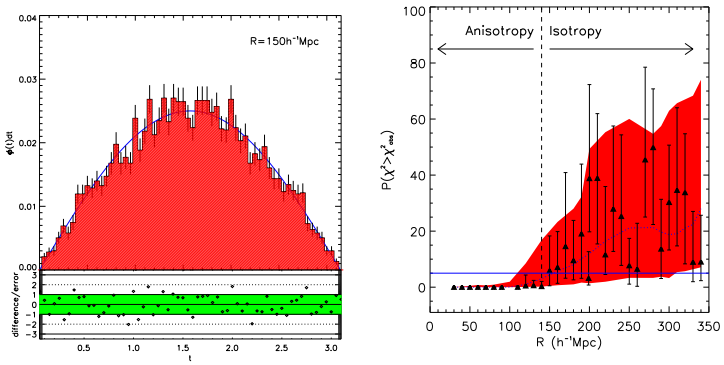

Figure 2.5: *Upper:* the observed PDF of the *los* angle t (histogram) is compared to the isotropic prediction ($\varphi(t) = (\sin t)/2$) on a scale $R = 150h^{-1}$Mpc. The ratio between model deviations and data errors is also plotted (together with the lines indicating 1σ and 2σ deviations). *Lower:* The confidence with which the hypothesis h_0 cannot be rejected on a given scale, i.e. probability that the assumption of everywhere isotropy is compatible with observations. This is computed as the median of the probability P inferred from 1000 re-samplings of the SDSS LRG sample that are shell homogeneous. Errorbars represents the first and third quartile of the distribution of P. The blue line shows the average expectation extracted from the analysis of 50 mocks catalogs simulating the sample. The red envelope shows the confidence threshold bracketing 1σ fluctuations around the average expectation.

isotropy is present for all observers.In accordance with standard theoretical expectations, as the R-scale increases, all the observers are equally likely to observe isotropy, i.e. they loose their specificity and progressively become the 'typical' observer of the universe. The left panel of Figure (2.5) confirms that the galaxy pattern observed from different positions in the universes approaches an isotropic distribution.

The precise scale of transition to isotropy R_{iso} is quantitatively determined as follows. First, by randomly rejecting galaxies from the main LRG sample, we have constructed 1000 subsamples that satisfies to the requirement $\rho_s = const$ (see figure 2.4). This bootstrapping process, allows us to estimate the central moments and the dispersion of the P statistics. I have then positioned the centers of the maximum number of non-overlapping spheres of radius R that fit inside the survey volume. In particular, I require that the position of the extraterrestrial observers change randomly from sample to sample. For each length-scale R probed, I have finally computed the risk of erroneously rejecting the null hypothesis as the median of P over the 1000 realizations. The right panel of Fig 2.5. shows that the median risk is larger than 5% for scales larger than $150h^{-1}$ Mpc. Despite the observed spread in the P-values for large R, essentially due to the low density of the LRG sample, a statistically significant sharp transition towards isotropy at a scale $R_{iso} \sim 150h^{-1}$ is unambiguously detected. Notwithstanding, a larger sample might also definitively exclude the hypothesis that the transition happens at a scale as low as $120h^{-1}$Mpc. In fact, the evidence with which such a low R_{iso} is currently excluded is not yet conclusive.

My conclusions are, within error bars, independent from the density cut I artificially impose to guarantee shell homogeneity. I have verified that samples that are everywhere

isotropic on a scale R continue to be everywhere isotropic on that scale if the density threshold imposed by the requirements of shell homogeneity is enhanced. This conclusion follows from the analysis of subsamples obtained by volume-limiting the SDSS catalogue at a lower redshift z_{max}. Interestingly, and in the opposite sense, if the sample is isotropic on a scale R then it continues to be isotropic on that scale, also even when its density is artificially lowered (by randomly rejecting galaxy members). The isotropic length proves robust up until the investigated scale R becomes smaller than ~ 1.5 times the mean inter-particle separation. Below that threshold, as the analysis of random samples already suggested, the predictive power of our indicator breaks down.

I have compared the measurements to predictions of N-body simulations of the large-scale structure of the universe. To this end, I have analyzed with the same technique 50 independent mock catalogs simulating the distribution of LRG galaxies in an SDSS-like survey. They were constructed by the LasDamas project (McBride et al., 2009) using Λ cold dark matter simulations (with characteristic parameters ΛCDM) simulations ($\Omega_M = 0.25, \Omega_\Lambda = 0.75, h = 0.7, \sigma_8 = 0.8, n_s = 1$).

Fig. 2.5 quantifies the confidence level with which the hypothesis h_0 cannot be rejected on a scale R, and compares it to what is expected in the mock catalogs. Not only is a sharp transition towards isotropy at a scale $R \sim 150h^{-1}$ detected in real SDSS data, it is observed in synthetic galaxy catalogs too. This excellent agreement implies that the scale of isotropy $R \sim 150h^{-1}$ is a length that characterizes not only luminous galaxies, i.e. the visible component of the universe, but also of the most massive dark matter halos. The significance of this conclusion is best understood by considering that the everywhere isotropy inferred from real data alone, does not give insight into the corresponding arrangement of the underlying mass component.

2.5 Conclusions

An acritical acceptance of the Copernican principle might result in what Haynes (1996) called the "Verrazzano bias". As in the case of this explorer who, off the coast of the outer banks of North Carolina, mistakenly believed that he had discovered the Pacific Ocean, it is dangerous to draw definite cosmological conclusions on the basis of limited data collected from a special spatial position.

In this Chapter I have presented a new geometrical tool that allows us to assess whether or not, from the view point of a distant galaxy, the large-scale structure of the universe appear almost identical to its aspect from earth. Virtually all of the previous attempts to identify the coarse graining scale above which the visible distribution of matter comply with the requirements of the CP have focused on the analysis of the so called *homogeneity scale*. I have addressed this same issue from a different angle. I propose to identify this fundamental length with the scale of everywhere isotropy R_{iso}, the scale above which the distribution of galaxies appears isotropic to every comoving observer, that we define as

the smallest scale at which the probability of wrongly rejecting the CP is smaller than 5%.

By analyzing state-of-the-art data, I have found that the galaxy distribution, as traced by luminous red galaxies, appears isotropic to every comoving observer in the universe once the averaging scale is larger than $R_{iso} \sim 150h^{-1}$ Mpc. This figure is in excellent agreement with predictions of the spatial clustering of galaxies in ΛCDM simulations.

The advantage of the method is that it is insensitive to the shape of the radial selection function of the redshift sample analyzed, i.e. to the effective number of objects that sample the underlying clustering of galaxies as a function of redshift. As a matter of fact, it is straightforward to subtract look-back time issues once the focus is shifted from counting objects (the standard methodology of the homogeneity tests) to measuring angles (as implemented by our strategy).

Since the matter distribution converges continuously towards homogeneity/isotropy, it is in fact quite arbitrary to decide which criterion must be adopted to single out an exact scale of transition. In this work I adopt the point of view that the most natural way to test the CP is to assign a probability to the hypothesis that this model is wrong. The goal is to frame the analysis of its coherence within the domain of probability theory, as the intrinsically statistical nature of this cosmological statement explicitly demands. This helps elucidating the meaning of such generic sentences as "....the CP holds on a scales larger than XXX Mpc" and will ease the quantitative comparison of the results obtained with different and independent methods.

Deeper redshift surveys of the universe (such as, for example, BOSS or EUCLID) are currently ongoing and expected to be soon completed. It would be interesting to understand if the scale of everywhere isotropy does scale as a function of cosmic time as predicted by numerical simulations of the gravitational clustering in the universe. This will confirm that the CP is not some temporary assertion about the present day appearance of the universe but a fundamental property of matter distribution at all cosmic epochs. Even more importantly, it will help us to shed light on the physics behind the large-scale uniformity of the universe by answering the question: where does this scale come from?

Chapter 3

The rms of matter fluctuations

According to a widely accepted paradigm, cosmic structures grew from tiny dark matter density fluctuations present in the otherwise homogeneous and rapidly expanding early universe. The standard version of the model incorporates the assumption that these primordial and Gaussian distributed fluctuations are amplified by gravity eventually turning into the rich structure that surrounds us today. This picture in which gravity, as described by general relativity, is the engine driving cosmic growth is generally referred to as the gravitational instability paradigm (GIP). However plausible it may seem, it is important to test its validity.

In the local universe the GIP paradigm has been shown to make sense of a vast amount of independent observations on different spatial scales, from galaxies to superclusters of galaxies (e.g. Peacock et al., 2001; Tegmark et al., 2006; Reyes et al., 2010). Deep galaxy surveys now allow us to test whether the predictions of this assumption are also valid at earlier epochs (Guzzo et al., 2008; Blake et al., 2011). In particular, they allow us to asses weather GR is the correct theory describing the action of gravity on large cosmological scales (e.g. Jain & Zhang, 2008; Uzan, 2009; Acquaviva & Gawiser, 2010).

Indeed, modifications to GR have been proposed as alternatives to explain observations showing that the universe is undergoing accelerated expansion (Dvali, Gabadadze & Porrati, 2000; Capozziello, Cardone & Troisi, 2005; Amendola, Polarski & Tsujikawa, 2007; Buzzi, Marinoni & Colafrancesco, 2008). Non standard gravitational models have also been invoked as an alternative to dark matter (Milgrom, 1983; Bekenstein, 2004) or to its standard physical characterization (Piazza & Marinoni, 2003; Bertacca et al., 2008). Since these modified gravity theories are specifically tuned to explain the uniform cosmic expansion history, a possible way to test their reliability is to analyze the inhomogeneous section of the universe, i.e. cosmological perturbations of matter (Linder, 2005; Zhang et al., 2007).

The central achievement of this section is therefore to detail the computational scheme that allows extracting both the amplitude and redshift scaling of the rms of matter fluctuation (σ_R) in real space, using as data, the three dimensional positions of galaxies. This

is a non-trivial task, compounded by our poor knowledge of the biasing mechanism, that is of how the global matter distribution is traced by its luminous subcomponent. To address this issue I exploit the information encoded in the reduced 2-point correlator C_{12} (see section 1.3.9). Following an original suggestion of Szapudi (1998), I work out the explicit expressions for the bias coefficients up to order 4 as a function of C_{12} and of the reduced 1-point cumulant S_i (up to order 5). Using these results I construct the central formula of this section (cfr. eq. (3.22)), i.e. an estimator of σ_R that is independent from any assumption about the linear power spectrum of matter. As a result, by contrasting the redshift evolution of σ_R with predictions of perturbation theory at linear order I can place constraints on the reliability of the GIP paradigm and in particular on the viability of general relativity on large cosmological scales. This formalism relies upon theoretical assumptions and approximations that I have tested using N-body simulations of luminous red galaxies (Horizon simulations and LasDamas simulations). In this way I can assess the overall coherence of the method and the impact of potential systematics.

3.1 Mapping matter into galaxy: statistical relations

I will begin by discussing the transformation laws that allows us to establish a connection between matter statistics (predicted by theory) and the galaxy statistics (extracted from a redshift survey).

3.1.1 The local, non-linear, galaxy biasing function

Smoothing is not the only process that leaves an imprint on the value of cumulant moments. Their amplitude and shape is also altered if galaxies, instead of massive particles, are used to trace the overall matter distributions. Fry & Gaztañaga (1993) computed the effects of biasing on 1-point (smoothed) cumulant moments up to order 5. In this section, I generalize their results by deriving the expressions of the smoothed 2-point cumulant moments of the galaxy overdensity field up to the same order. These are new observables that can be used to test predictions of the GIP in the weakly non-linear regime, to give insight into the gravity induced large-scale bias, and also to distinguish models with Gaussian initial conditions from their non-Gaussian alternatives.

As stressed in the introduction of this thesis, I assume that bias between galaxy fluctuations and matter fluctuations is described by a local non-linear mapping (Fry & Gaztañaga, 1993). Since the biasing function can be expanded in Taylor series, it can be written in a formal way as

$$\delta_{g,R} = \sum_{i=0}^{\infty} \frac{b_i}{i!} \delta_R^i. \tag{3.1}$$

Here δ_R is the matter fluctuation smoothed on scale R and $\delta_{g,R}$ is the galaxy fluctuation smoothed on the same scale. Note that the zero order basing coefficient b_0 is not

an arbitrary parameter. Its value, instead, depends on the order at which the Taylor series is effectively truncated, and it is computed by requiring that the average of galaxy fluctuations be, by definition, null. Thus, by averaging bias function (eq. 3.1) I obtain

$$b_0 = -\sum_{i=2}^{N} \frac{b_i}{i!} \langle \delta_R^i \rangle.$$

By substituting b_0 in 3.1 I obtain

$$\delta_{g,R} = b_1 \delta_R + \sum_{i=2}^{N} \frac{b_i}{i!} \Delta_i, \tag{3.2}$$

where $\Delta_i = \delta_R^i - \langle \delta_R^i \rangle$.

In the following I will first express the galaxy N-point autocorrelation functions of the galaxy density field as a function of the corresponding statistics for the matter density field. These results follows from applying the cluster expansion technique (see section 1.3.3). I will then discuss how I used these results to establish the relation between the correlators of mass and galaxies.

3.1.2 N-point galaxy autocorrelation functions

The cluster expansion is the technique that I used to obtain the N-point autocorrelation functions of galaxies as a function of the N-point correlation functions of the matter field. Since from now on we only consider smoothed statistics, I simplify my notation by omitting any explicit reference to the smoothing scale R. The $R-$dependence of relevant statistical quantities will be re-emphasized when necessary. Consider the N-point, smoothed, correlation functions of matter up to order 5

$$\begin{aligned} \xi_{ij} &= \kappa_{11}(\mathbf{x}_i, \mathbf{x}_j) \\ \zeta_{ijk} &= \kappa_{111}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) \\ \eta_{ijkl} &= \kappa_{1111}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l) \\ \omega_{ijklm} &= \kappa_{11111}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l, \mathbf{x}_m). \end{aligned} \tag{3.3}$$

At leading order, the corresponding statistics describing the distribution of galaxies, labeled with the suffix g are,

$$\xi_{12,g} = {b_1}^2 \, \xi_{12} \tag{3.4}$$

$$\zeta_{123,g} = {b_1}^3 \, \zeta_{123} + {b_1}^3 c_2 (\xi_{13}\xi_{23} + 2 \text{ perm}) \tag{3.5}$$

$$\eta_{1234,g} = {b_1}^4 \, \eta_{1234} + {b_1}^4 [c_2(\xi_{23}\zeta_{124} + 11 \text{ perm}) + c_3(\xi_{14}\xi_{24}\xi_{34} + 3 \text{ perm})$$

$$
\begin{aligned}
& +{c_2}^2(\xi_{13}\xi_{24}\xi_{34} + 11\ perm)] \\
\omega_{12345,g} &= {b_1}^5\,\omega_{12345} + {b_1}^5[c_2\,(\xi_{15}\eta_{2345} + 19\ perm + \zeta_{125}\zeta_{345} + 14\ perm) \\
& +c_3\,(\zeta_{135}\xi_{25}\xi_{45} + 29\ perm) + c_4\,(\xi_{14}\xi_{24}\xi_{34}\xi_{45} + 4\ perm) \\
& +{c_2}^2\,(\zeta_{135}\xi_{24}\xi_{25} + 119\ perm) + c_2c_3\,(\xi_{14}\xi_{25}\xi_{34}\xi_{24} + 59\ perm) \\
& +{c_2}^3\,(\xi_{13}\xi_{24}\xi_{45}\xi_{15} + 59\ perm)\,].
\end{aligned}
$$

Equation (3.4) reduces to the commonly used definition of the linear bias. Additionally, equation (3.5) can be combined to equation (3.4) in order to express the third order reduced amplitude of galaxies with respect to the linear bias, to the second order bias parameter and to the third order reduced amplitude of matter

$$Q_{3,g} = b_1^{-1}\{Q_3 + c_2\}.$$

Moreover, an original contribution of this thesis is the expression of the four and five points correlation functions of galaxy fluctuations as a function of matter statistics. The utility of these expressions will be made explicit in the following, where we use them to investigate the hierarchical properties of matter fluctuations.

3.1.3 The correlators of the galaxy density field

The mapping between correlators of mass (i.e. the two point cumulant moments (κ_{nm}) of order (nm)) and the galaxy correlators ($\kappa_{nm,g}$) at the leading order, follows immediately by taking the 2-point limiting case of the expressions obtained in the previous section. Listed below are the results up to order $n + m = 4$ (including also non-leading contributions)

$$
\begin{aligned}
\kappa_{11,g} &= b_1^2\kappa_{11} + b_1^2(c_3 + C_{12}c_2)\kappa_{11}\,\kappa_2 + 1/2\,b_1^2c_2^2\kappa_{11}^2 \qquad (3.6)\\
\kappa_{12,g} &= b_1^3(C_{12} + 2\,c_2)\kappa_{11}\,\kappa_2 + b_1^3c_2\kappa_{11}^2 + b_1^3(5/2\,c_3C_{12} + 2\,c_2^2C_{12} + 3\,c_3c_2 + c_4 \\
&\quad +1/2\,c_2C_{22} + c_2^2S_3 + c_3S_3 + c_2C_{13})\kappa_{11}\,\kappa_2^2 + b_1^3(c_2^3 + 1/2\,c_4 + 3\,c_2^2C_{12} \\
&\quad +2\,c_3c_2 + c_3C_{12})\kappa_{11}^2\kappa_2 + b_1^3c_3c_2\kappa_{11}^3 \\
\kappa_{13,g} &= {b_1}^4(6\,{c_2}^2 + C_{13} + 3\,c_2S_3 + 3\,c_3 + 6\,c_2C_{12})\kappa_{11}\,{\kappa_2}^2 + {b_1}^4(3\,C_{12}c_2 \\
&\quad +6\,{c_2}^2){\kappa_{11}}^2\kappa_2 + {b_1}^4c_3{\kappa_{11}}^3 + {b_1}^4(3/2\,c_2C_{14} + 9/2\,C_{12}c_3S_3 \\
&\quad +12\,{c_2}^2c_3 + 6\,{c_2}^3S_3 + 45/2\,c_2c_3C_{12} + 9\,c_2c_4 + 3/2\,c_3S_4 + 9/2\,c_4S_3 \\
&\quad +3/2\,c_5 + 15/2\,{c_3}^2 + 3\,{c_2}^2S_4 + 3\,{c_2}^2C_{22} + 15/2\,{c_2}^2C_{12}S_3 \\
&\quad +9/2\,c_4C_{12} + 6\,{c_2}^2C_{13} + 45/2\,c_3S_3c_2 + 5\,c_3C_{13} + 6\,{c_2}^3C_{12} \\
&\quad +1/2\,c_2C_{23})\kappa_{11}\,{\kappa_2}^3 + {b_1}^4(9/2\,{c_2}^2{C_{12}}^2 + 6\,{c_2}^2C_{13} + 3/2\,c_3C_{22} \\
&\quad +39/2\,c_2c_3C_{12} + 3/2\,c_4C_{12} + 9/2\,c_3S_3c_2 + 3\,{c_2}^4 + 18\,{c_2}^3C_{12} + 3/2\,c_3{C_{12}}^2 \\
&\quad +15/2\,c_2c_4 + 18\,{c_2}^2c_3 + 6\,{c_2}^3S_3){\kappa_{11}}^2{\kappa_2}^2 + {b_1}^4(9/2\,{c_3}^2 + 1/2\,c_5
\end{aligned}
$$

$$+3/2\,c_4C_{12}+9\,c_2c_3C_{12}+6\,{c_2}^2c_3){\kappa_{11}}^3\kappa_2+3/2\,{b_1}^4c_2c_4{\kappa_{11}}^4$$

where $c_i \equiv b_i/b_1$ for $i \geq 2$. In Appendix (F) I list the expressions for the galaxy correlators up to fifth order. For $\mathbf{x}_1 = \mathbf{x}_2$, one recovers the expressions of Fry & Gaztañaga (1993) (cfr. their eq. 9).[1]

3.1.4 Hierarchy of reduced galaxy correlators

The leading term in eq. (3.6) reduces, in the 1-point limiting case, to

$$\kappa_{2,g} = b_1^2\kappa_2. \tag{3.7}$$

I find that, to leading order in the product $\kappa_{11}\kappa_2$ the remaining results, $\kappa_{nm,g}$ for $n+m \geq 3$, preserve the hierarchical properties of matter correlators, i.e. $k_{nm,g} = C_{nm,g}\kappa_{11,g}\kappa_{2,g}^{n+m-2}$, with amplitudes $C_{nm,g}$ given by

$$\begin{aligned}
C_{12,g} &= b_1^{-1}(C_{12}+2\,c_2)\\
C_{13,g} &= b_1^{-2}(C_{13}+3\,c_2(S_3+2C_{12})+3\,c_3+6\,c_2^2)\\
C_{22,g} &= b_1^{-2}(C_{22}+4\,c_2C_{12}+4\,{c_2}^2)\\
C_{14,g} &= b_1^{-3}(C_{14}+4\,c_2(S_4+3\,C_{13})+12\,(c_3+3\,c_2^2)(S_3+C_{12})+12\,c_2C_{12}S_3\\
&\quad +4\,c_4+36\,c_3c_2+24\,c_2^3)\\
C_{23,g} &= b_1^{-3}(C_{23}+2\,c_2(C_{13}+3\,C_{22})+3\,(c_3+c_2S_3)C_{12}+6\,c_2^2(S_3+3\,C_{12})\\
&\quad +6\,c_3c_2+12\,c_2^3).
\end{aligned} \tag{3.8}$$

These relations show that in order to draw any conclusions from the galaxy distribution about matter correlations of order N, properties of biasing must be specified completely to order $N-1$. Note, also, that the equations (3.8) have been obtained in the large separation approximation and fail as soon as $|\mathbf{x}_1 - \mathbf{x}_2| < R$. As a consequence, in the 1-point limit they do not converge to the results of Fry & Gaztañaga (1993) on the amplitude of the reduced cumulants, that is

$$\begin{aligned}
S_{3,g} &= b_1^{-1}\,(S_3+3\,c_2)\\
S_{4,g} &= b_1^{-2}\left(S_4+12\,c_2S_3+4\,c_3+12\,{c_2}^2\right)
\end{aligned} \tag{3.9}$$

[1] Actually I found two misprints in their eq. (9). The term $210c_2^3S_3$ is actually $210c_3^2S_3$ and $180c_2^2S_4$ should read $180c_2^3S_4$. Anyway such terms were subsequently neglected in that analysis, leaving the author's conclusions unaltered.

$$S_{5,g} = b_1^{-3}\left(S_5 + 20\,c_2 S_4 + 15\,c_2 {S_3}^2 + (30\,c_3 + 120\,{c_2}^2)S_3 + 5\,c_4 + 60\,c_2 c_3 + 60\,{c_2}^3\right).$$

Interestingly, it is possible to design a sanity test to verify whether these calculations are correct. In section 1.3.5 I showed that if an non-smoothed field is considered, and if the validity of the HA is assumed, then at third order $S_3 = 3/2 C_{12}$, and, at fourth order, $S_4 = C_{22} + \frac{4}{3} C_{13}$. These properties must hold also for the galaxy field because equations (3.8) and (3.9) show that the non-linear local bias conserves the HA. As a consequence, we should obtain $S_{g,3} = 3/2 C_{g,12}$ and $S_{g,4} = C_{g,22} + \frac{4}{3} C_{g,13}$ if the expressions (3.8) have been correctly derived. This is effectively what we deduced from the above set of equations.

Interrestingly, the hierarchical scaling is not the only matter property which survives to the local non-linear biasing transformation of eq. 1.93. I also find that

$$\begin{aligned} C_{22,g} &= C_{12,g} C_{12,g} \\ C_{23,g} &= C_{12,g} C_{13,g} \end{aligned} \tag{3.10}$$

that is the reduced galaxy correlators $C_{nm,g}$, conserve the factorization property of the matter density field (cfr. eq. 1.73).

3.2 Non-linear bias in real space

The most common methods for estimating the amplitude of the non-linear bias coefficients b_i relies on fitting a theoretical model to higher order statistical observables, such as the 3-point correlation function (e.g. Gaztañaga & Scoccimarro, 2005), the bispectrum (e.g. Verde et al., 2002), or the galaxy probability distribution function (e.g. Marinoni et al., 2005).

Here the approach I take is orthogonal. I explicitly derive analytical relations directly expressing the non-linear bias coefficients b_i as a function of high order observables. To this end I cannot simply invert the set of equations 3.8 (or 3.9), since the N^{th} galaxy reduced correlator (or moment) are a function of $N+1$ bias coefficients (Szapudi, 1998). I add, instead, the expression of the 3rd order reduced correlator $C_{12,g}$ to the system of equations (3.9) and solve the resulting set for the biasing coefficients. I obtain

$$b_1 = \frac{3\,C_{12} - 2\,S_3}{3\,C_{12,g} - 2\,S_{3,g}} \tag{3.11}$$

$$b_2 = \frac{(C_{12} S_{3,g} - S_3 C_{12,g})(3\,C_{12} - 2\,S_3)}{(3\,C_{12,g} - 2\,S_{3,g})^2} \tag{3.12}$$

$$
\begin{aligned}
b_3 \;=\; & (9\,S_{4,g}C_{12}^2 - 12\,S_{4,g}C_{12}S_3 + 4\,S_{4,g}S_3^2 - 12\,S_3S_{3,g}C_{12}C_{12,g} + 24\,S_3S_{3,g}^2C_{12} \\
& -24\,S_3^2S_{3,g}C_{12,g} + 24\,S_3^2C_{12,g}^2 - 9\,S_4C_{12,g}^2 + 12\,S_4C_{12,g}S_{3,g} - 4\,S_4S_{3,g}^2 - 12\,S_{3,g}^2C_{12}^2) \\
& \times(3\,C_{12} - 2\,S_3)(3\,C_{12,g} - 2\,S_{3,g})^{-3}/4 \\
b_4 \;=\; & (-60\,{S_3}^3S_{4,g}C_{12,g} - 108\,S_{5,g}{C_{12}}^2S_3 - 120\,S_4{S_{3,g}}^3S_3 - 600\,{S_{3,g}}^3S_3C_{12}^2 \\
& +495\,S_4S_3C_{12,g}^3 - 600\,{S_{3,g}}^2S_3^3C_{12,g} - 270\,S_{3,g}S_{4,g}C_{12}^3 + 72\,S_{5,g}C_{12}S_3^2 \\
& +108\,S_5C_{12,g}^2S_{3,g} + 1200\,S_{3,g}S_3^3C_{12,g}^2 + 600\,S_{3,g}^3S_3^2C_{12} - 40\,S_4S_{3,g}^3C_{12} \\
& +120\,S_{3,g}S_{4,g}S_3^3 - 72\,S_5C_{12,g}S_{3,g}^2 - 90\,S_4S_{3,g}C_{12}C_{12,g}^2 + 120\,S_4S_{3,g}^2C_{12}C_{12,g} \\
& -930\,S_4S_{3,g}S_3C_{12,g}^2 + 580\,S_4S_{3,g}^2S_3C_{12,g} + 630\,S_{3,g}S_{4,g}C_{12}^2S_3 - 480\,S_{3,g}S_{4,g}C_{12}{S_3}^2 \\
& +180\,S_{3,g}^2S_3C_{12}^2C_{12,g} - 600\,S_{3,g}^2S_3^2C_{12}C_{12,g} + 270\,S_{3,g}S_3^2C_{12}C_{12,g}^2 \\
& -135\,S_3S_{4,g}C_{12}^2C_{12,g} + 180\,S_3^2S_{4,g}C_{12}C_{12,g} + 54\,S_{5,g}C_{12}^3 - 16\,S_{5,g}S_3^3 - 54\,S_5C_{12,g}^3 \\
& +16\,S_5S_{3,g}^3 + 240\,S_{3,g}^3C_{12}^3 - 690\,S_3^3C_{12,g}^3)(3\,C_{12} - 2\,S_3)(3\,C_{12,g} - 2\,S_{3,g})^{-4}/10.
\end{aligned}
$$

Note that linear combination of S_3 and C_{12} which appears in the expression of b_1 would be exactly equal to zero in the no smoothing case. We see, again, that the smoothing process does not conserve the perfect hierarchy of a stochastic field, in the sense that the hierarchical amplitudes (i.e. the reduced correlators and moments) cannot be constant if the field is smoothed.

This set of equations allows to investigate the eventual non-linear character of the biasing function up to of order 4 by exploiting information encoded in the reduced correlators up to order 3 and the reduced moments up to order ≤ 5. Note that, if I set $\beta_R(r) = 0$ in eq. (1.90), then my expressions for b_1 and b_2 (cfr. should eqs. (3.11) and (3.12)) reduce to equations (4) and (5) originally derived by Szapudi (1998). As stressed by this author, in this formalism biasing coefficients are not anymore simple parameters to be estimated (by maximizing, for example, the likelihood of observables that are sensitive to them such as the reduced skewness S_3 (Gaztañaga, 1994; Gaztañaga & Frieman, 1994), the bispectrum (Fry, 1994; Scoccimarro, 1998; Feldman et al., 2001; Verde et al., 2002), the 3-point correlation function (Gaztañaga & Scoccimarro, 2005; Pan & Szapudi, 2005) or the full probability distribution function of the density fluctuations Marinoni et al. (2005, 2008), but they become themselves estimators.

In what follows, I will focus on the linear biasing parameter b_1 only. In my formalism, this quantity is explicitly expressed in terms of the amplitude of third-order statistics (i.e. the reduced cumulants and correlators) and it is independent from the amplitude of second-order statistics such as the *rms* of matter fluctuations. Notwithstanding, b_1 seems to depend on the shape of the power spectrum of the matter density fluctuations via the terms γ_R and $\beta_R(r)$ appearing in eq. (1.90). I will demonstrate in the next section that the inclusion of the correction term $\beta_R(r)$ in the analysis has the additional advantage of making the linear bias coefficient b_1 effectively independent from any assumption about

second order statistics, i.e. independent not only from the amplitude but also from the shape of the linear matter power spectrum. This *fully third-order* dependency of the linear biasing parameter estimator b_1 allows me to construct an estimator of the variance σ_R of the matter density fluctuations smoothed on the scale R.

3.3 Constructing the estimator of σ_m

Now that all the ingredients are collected, I detail how we construct the estimator of the linear matter density fluctuations σ_R. From eq. (1.89), (1.90), (3.7) and (3.11) we obtain

$$\sigma_R = \frac{\tau_{g,R}(r)}{\beta_R(r) - \gamma_R}\sigma_{g,R} \tag{3.13}$$

where

$$\tau_{g,R}(r) = 3C_{12,g,R}(r) - 2S_{3,g,R} \tag{3.14}$$

and where the suffix g indicates that the relevant quantities are evaluated using data. For clarity, the scale dependence of third order statistics is explicitly highlighted. Apparently, the right hand side of the previous equations depends on the overall shape of the *a-priori* unknown matter power spectrum. In reality, the terms $\beta_R(r) - \gamma_R$ can be consistently estimated from observations without any additional theoretical assumption. To show this, I define

$$\alpha_R(r) \equiv \frac{d\log\eta_R(r)}{d\log R} \tag{3.15}$$

where

$$\eta_R(r) \equiv \frac{\xi_R(r)}{\sigma_R^2}. \tag{3.16}$$

As far as matter particles are considered, the previous definitions imply that $\alpha_R(r) = \beta_R(r) - \gamma_R$. By combining the expression for $\xi_{g,R}$ given by eq. 3.6 with its 1-point limiting case i.e.

$$\sigma_{g,R}^2(z) = b_1^2\sigma_R^2(z)\left\{1 + \left(1/2{c_2}^2 + S_{3,R}c_2 + c_3\right)\sigma_R^2(z)\right\}, \tag{3.17}$$

and using the fact that, on scales where WNLPT applies, $\sigma_R^2 \ll 1$, we obtain

$$\eta_{g,R}(r) \sim \eta_R(r) - \left\{(S_{3,R} - C_{12,R})c_2 + 1/2{c_2}^2\right\}\xi_R(r) + 1/2{c_2}^2\eta_R(r)\xi_R(r), \tag{3.18}$$

where $\eta_{g,R}(r) = \xi_{g,R}(r)/\sigma_{g,R}^2$ and where the terms on the RHS have been sorted by order of magnitude. Finally, in the LS limit ($\xi_R(r)$ negligible with respect to $\eta_R(r)$) the previous

equations reduces to $\eta_{g,R}(r) \sim \eta_R(r)$. The level of accuracy of this approximation is presented in Figure 3.1 where I show that, on a typical scale ($R \sim 16h^{-1}$ Mpc), the imprecision is less than 0.5% at any cosmic epoch investigated ($0 < z < 0.6$). Since $\alpha_{g,R}(r) = \alpha_R(r)$, we obtain

$$\sigma_R = \frac{\tau_{g,R}(r)}{\alpha_{g,R}(r)}\sigma_{g,R} \tag{3.19}$$

where $\alpha_{g,R}(r) = d\log\eta_{g,R}(r)/d\log R$.

Count in cells techniques provide an estimate of the terms on the RHS of eq. 3.19. In this regard, a central point worth stressing concerns the continuum-discrete connection. Biasing is not the only obstacle hampering the retrieval of matter properties from the analysis of galaxy catalogs. The formalism needs also to correct for the fact that the galaxy distribution is an intrinsically discrete process (Szapudi & Szalay, 1993). These issues will be thoroughly addressed in section 3.5.1 where I present the strategy that we have adopted in order to minimize the sampling noise.

3.4 From real- to redshift-space

Matter perturbations induce a peculiar velocity field $\vec{v}$ on top of the Hubble flow. As a result, cosmic objects such as typical galaxies are not anymore 'comoving' and their redshift is not an unbiased estimator of the ratio between the scale factor of the universe today and the scale factor at the time at which the received photons were emitted. Peculiar velocities, and in particular their line-of sight component u, alter systematically the cosmological redshift z_c. The observed redshift z_o is

$$z_o = z_c + \frac{u}{c}(1 + z_c), \tag{3.20}$$

where c is the vacuum speed of light. The effects of peculiar velocity perturbations, called redshift distortions, are thoroughly reviewed by Hamilton (1998). In the following, observables constructed by using the observed redshift z_o are referred to as 'redshift space' observables, while those re-constructed in order to erase peculiar velocity effects, are called 'real space' observables.

All the theoretical results presented in the previous sections have been obtained in real-space. Therefore, the feasibility of extracting the value of mass fluctuations σ_R via eq. (3.19) rests upon the possibility of expressing real space variables ($b_1^{-1} = \tau_{g,R}/\alpha_{g,R}$, and $\sigma_{g,R}$) in terms of redshift space observables. On the large (linear) cosmic scales where our formalism applies, the Kaiser model (e.g. Kaiser, 1987) effectively describes the mapping between real and redshift space expressions of second order statistics. The transformation is given by

$$\sigma^z_{g,R} = \left[1 + \frac{2}{3}\frac{f}{b_1} + \frac{1}{5}\left(\frac{f}{b_1}\right)^2\right]^{1/2} \sigma_{g,R}, \tag{3.21}$$

where the suffix z labels measurements in redshift (as opposed to configuration) space, and where f is the logarithmic derivative of the linear growth factor $D(a)$ with respect to the scale factor a. Much easier is the transformation rule for $\alpha_{g,R}$: since the model correction for redshift distortions at second order is the same for 1- and 2-point statistics such correction vanishes in the expression of $\eta_{g,R}$ so that in linear regime it is unaffected by redshift distortions, that is $\alpha^z_{g,R} = \alpha_{g,R}$.

Assessing the impact of peculiar motions on third order statistics is less straightforward. Up to now all the formulas were derived analytically from theory. To address this last issue we now use numerical simulations. By running some tests using a suite of simulated galaxy catalogs (these catalogs will be described in section 3.5.2) I conclude that the amplitude of $S_{3,g}$ and $C_{12,g}$ are systematically (and non-negligibly) higher in z-space and that the relative overestimation systematically increases as a function of the order of the statistics considered (see Figures 3.3 and 3.4). Notwithstanding, from the theoretical side, the expressions of third order statistics ($S_{3,R} = S_3 + \gamma_R$ and $C_{12} = (2S_3)/3 + \gamma_R/3 + \beta_R(r)/3$) imply that the linear combination $3C_{12,R} - 2S_{3,R}$ should be much more insensitive to redshift distortions. Note, in particular, that both γ_R and $\beta_R(r)$ are unaffected by linear motions. However convincing it might seem, this guess applies only to matter particles. In order to draw definitive conclusions about the impact of peculiar motions on $\tau_{g,R}$, I have used N-body galaxy simulations. Guided by synthetic catalogs I demonstrate that the biased galaxy statistics $\tau_{g,R}$ is effectively unaffected by redshift distortions. This conclusion is graphically presented in Figure 3.1 where I show that the relative error introduced by reconstructing the statistics using observed redshifts, instead of the cosmological ones, is progressively smaller as R increases, and it is globally $<\sim 2\%$.

By incorporating these results into the formalism we finally obtain

$$\hat{\sigma}_R = \sigma^z_{g,R}\left[\left(\frac{\alpha^z_{g,R}}{\tau^z_{g,R}}\right)^2 + \frac{2}{3}\frac{\alpha^z_{g,R}}{\tau^z_{g,R}}f + \frac{1}{5}f^2\right]^{-1/2} \tag{3.22}$$

an estimator that is manifestly independent from any assumption about the amplitude and shape of the linear matter power spectrum. Also, this formula is independent from any assumption about the value of the Hubble constant H_0. Only an *a-priori* gravitational model must be assumed to correct for redshift space distortions in the local universe, that is to evaluate the growth rate function $f(z)$. This introduces an additional strong dependence on the cosmological parameters Ω_M and Ω_Λ on top of the marginal one that is forced upon when we compute metric distances in order to estimate $\sigma^z_{g,R}$, $\alpha^z_{g,R}(r)$ and $\tau^z_{g,R}$. More quantitatively, when the input cosmological parameters are chosen in the parameter plane delimited by $0 \leq \Omega_M \leq 1$ and $0 \leq \Omega_\Lambda \leq 1$, the maximum relative

variation of the estimates with respect to their fiducial value in the ΛCDM cosmological model are $\max|df/f| \sim 0.6$, $\max|d\tau/\tau| \sim 0.2$, $\max|d\alpha/\alpha| = 0.3$, and $\max|d\sigma/\sigma| = 0.15$. The weak cosmological dependence of $\tau^z_{g,R}$ follows from the fact that neither the reduced cumulants nor the reduced correlators of mass are effectively sensitive to the background cosmology. We expect them to be essentially unaffected also by sensible modifications of the gravitational theory (Gaztañaga & Lobo (2001), Multamaki et al. (2003), but see for example Freese & Lewis (2002) or Lue, Scoccimarro & Starkman (2001) for more radical scenarios where this expectation is not met). I have also noted that the denser in matter is the cosmological model the more overestimated are the values of both of $\alpha^z_{g,R}$ and $\tau^z_{g,R}$. Since also the amplitude of the cosmological dependence is nearly the same, we expect the ratio $\alpha^z_{g,R}/\tau^z_{g,R}$, i.e. the linear bias parameter, to be nearly insensitive to the underlying cosmological model. I have quantitatively verified this statement in Figure (3.11) of section 3.5.5. In conclusion, the estimator in eq. (3.22) is sensitive to cosmology mainly through the growth rate function $f(z)$ and, to some degree, through $\sigma^z_{g,R}$.

Once the cosmological background is known via independent techniques (e.g Astier et al. (2006); Komatsu et al. (2011); Marinoni & Buzzi (2010))) the strategy I have outlined offers the possibility to estimate in a direct way the amplitude and time evolution of matter fluctuations. The formalism could also be implemented, in a reverse direction, to probe the coherence of the gravitational instability paradigm. Any eventual discrepancy resulting from the comparison of the measurements (cfr. eq. (3.22)) with theoretical predictions (cfr. eq. (1.79)) provides evidence that either the assumed set of cosmological parameters are wrong, either the assumed power spectrum of matter fluctuations is poorly described by linear theory, either the time dependence of the linear growing mode $D(t)$ is deduced in the context of an improper gravitational model.

3.5 Applying the Method

The practical implementation of the method, including successful tests of its robustness, is discussed in this section. I first present a strategy to estimate the correlators of order (n, m) of discrete 3D density fields such as those sampled by galaxy redshift surveys. I then show that, by applying the test formalism to N-body simulations of the large scale structure of the universe, I am able to recover the amplitude and scaling of the linear matter fluctuations $\sigma_R(z)$. A strategy to test the coherence of the results and to validate our conclusions is also designed, applied and discussed.

3.5.1 Statistical estimators of the galaxy correlators

The galaxy distribution is a discrete, 3-dimensional stochastic process. The random variable N models the number of galaxies within typical cells (of constant comoving size) that ideally tesselate the universe or, less emphatically, a given redshift survey. Notwith-

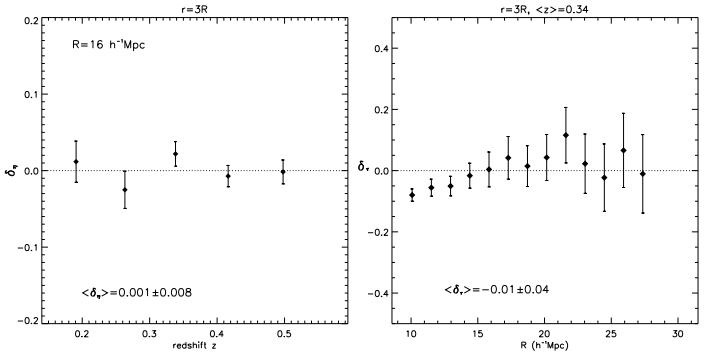

Figure 3.1: *Left:* the relative error arising from estimating the η-function using galaxies instead of matter. The inaccuracy $\delta_\eta = \eta^z_{g,R}/\eta_R - 1$ is evaluated at different epochs and is determined by assuming $R = 16h^{-1}$Mpc and $r = 3R$. Simulated data are extracted from the mock catalogs described in section §6.2 . Error bars are 1σ standard deviation and represent the best accuracy attainable with the available suite of simulations. *Right:* relative difference between the real- and redshift-space estimation of $\tau_{g,R} \equiv 3C_{12,g}(r) - 2S_{3,g}$. The inaccuracy $\delta_\tau = \tau^z_{g,R}(r)/\tau_{g,R}(r) - 1$ is plotted as a function of the smoothing scale R and is computed at a separation $r = 3R$. The dotted line represents the case in which computing the τ statistics using observed redshifts is equivalent to using cosmological redshifts. Real- and redshift-space simulated data are extracted from the mock catalogs described in section §6.2. Error bars represent 1σ standard deviation.

standing, a variable that is more directly linked to theoretical predictions of cosmological perturbation theory is the adimensional galaxy excess

$$\delta_N \equiv \frac{N}{\bar{N}} - 1, \tag{3.23}$$

where $\bar{N}$ is the mean number of galaxies contained in the cells.

To estimate 1-point moments of the galaxy overdensity field ($\mu_{n,g} = \langle \delta^n_N \rangle$), I fill the survey volume with the maximum number (N_t) of non-overlapping spheres of radius R (whose center is called *seed*) and we compute

$$\hat{\mu}_{n,g} = \frac{1}{N_t} \sum_{i=1}^{N_t} \delta^n_{N,i}, \tag{3.24}$$

where $\delta_{N,i}$ is the adimensional counts excess in the i-th sphere.

As far as 2-point statistics are concerned, I add a motif around each previously positioned spheres. The center of each new sphere is separated from the seed by the length $r = nR$ and the pattern is designed in such a way to maximize the number of quasi non-overlapping spheres at the given distance r (the maximum allowed overlapping between contiguous spheres is 2% in volume.) The moments $\mu_{nm,g} = \langle \delta^n_{N,i} \delta^m_{N,j} \rangle$, that is the average of the excess counts over all the i and j spherical cells at separation r, are estimated as

$$\hat{\mu}_{nm,g} = \frac{1}{2N_t N_{mot}} \left\{ \sum_{i=1}^{N_t} \delta^n_{N,i} \sum_{j=1}^{N_{mot}} \delta^m_{N,j} + \sum_{i=1}^{N_t} \delta^m_{N,i} \sum_{j=1}^{N_{mot}} \delta^n_{N,j} \right\}, \tag{3.25}$$

where N_{mot} is the number of spheres at distance r from a given seed, and where I have assumed that the stochastic process is stationary. The expression of the 1- and 2-point cumulant moments $k_{nm,g}$ follow immediately from eq. (1.53).

Since galaxies counts are a discrete sampling of the underlying continuous stochastic field $\lambda_g(\mathbf{x})$ (see Chapter 1) it is necessary to correct our estimators for discreteness effects. In other terms, the quantity of effective physical interest that we want to estimate is $\delta_{g,R}(\mathbf{x}) \equiv \Lambda_g(\mathbf{x})/\bar{\Lambda}_g - 1$ where $\Lambda_g(\mathbf{x}) = \int_{V(\mathbf{x})} \lambda(\mathbf{x}')d^3x'$ is the continuous limit of the discrete counts N in the volume V. To this purpose, following standard practice in the field, I model the sampling as a local Poisson process (LPP, Layser (1956)) and I map moments of the discrete variable N into moments of its continuous limit by using

$$\langle \Lambda_g^n \rangle = \langle N(N-1)...(N-n+1) \rangle = \langle (N)_f^n \rangle, \tag{3.26}$$

in the case of 1-point statistics, and its generalization

$$\langle \Lambda_g^n(\mathbf{x}_1)\Lambda_g^m(\mathbf{x}_2) \rangle = \langle (N_1)_f^n (N_2)_f^m \rangle, \tag{3.27}$$

for the 2-point case. As a result, the estimators of $k_{nm,g} = \langle \delta_{g,R}^n(\vec{x})\delta_{g,R}^m(\vec{x}+\vec{r}) \rangle_c$ corrected for shot noise effects are (see also Angulo et al. (2008))

$$\begin{aligned}
\hat{k}_{2,g} &= \hat{\mu}_{2,g} - \bar{N}^{-1} \\
\hat{k}_{11,g} &= \hat{\mu}_{11,g} \\
\hat{k}_{3,g} &= \hat{\mu}_{3,g} - 3\bar{N}^{-1}\hat{\mu}_{2,g} + 2\bar{N}^{-2} \\
\hat{k}_{12,g} &= \hat{\mu}_{12,g} - \bar{N}^{-1}\hat{\mu}_{11,g} \\
\hat{k}_{4,g} &= \hat{\mu}_{4,g} - 3\hat{\mu}_{2,g}^2 - 6\bar{N}^{-1}\hat{\mu}_{3,g} + 11\bar{N}^{-2}\hat{\mu}_{2,g} - 6\bar{N}^{-3} \\
\hat{k}_{13,g} &= \hat{\mu}_{13,g} - 3\hat{\mu}_{2,g}\hat{\mu}_{11,g} + 2\bar{N}^{-2}\hat{\mu}_{11,g} - 3\bar{N}^{-1}\hat{\mu}_{12,g} \\
\hat{k}_{22,g} &= \hat{\mu}_{22,g} - 2\hat{\mu}_{11,g}^2 - \hat{\mu}_{2,g}^2 + \bar{N}^{-2}\hat{\mu}_{11,g} - 2\bar{N}^{-1}\hat{\mu}_{12,g} \\
\hat{k}_{5,g} &= \hat{\mu}_{5,g} - 10\hat{\mu}_{2,g}\hat{\mu}_{3,g} - 10\bar{N}^{-1}\{\hat{\mu}_{4,g} - 3\hat{\mu}_{2,g}^2\} + 35\bar{N}^{-2}\hat{\mu}_{3,g} - 50\bar{N}^{-3}\hat{\mu}_{2,g} + 24\bar{N}^{-4} \\
\hat{k}_{14,g} &= \hat{\mu}_{14,g} - 6\hat{\mu}_{12,g}\hat{\mu}_{2,g} - 4\hat{\mu}_{11,g}\hat{\mu}_{3,g} - 6\bar{N}^{-1}\{\hat{\mu}_{13,g} - 3\hat{\mu}_{11,g}\hat{\mu}_{2,g}\} + 11\bar{N}^{-2}\hat{\mu}_{12,g} \\
&\quad -6\bar{N}^{-3}\hat{\mu}_{11,g} \\
\hat{k}_{23,g} &= \hat{\mu}_{23,g} - 6\hat{\mu}_{11,g}\hat{\mu}_{12,g} - 3\hat{\mu}_{2,g}\hat{\mu}_{12,g} - \hat{\mu}_{2,g}\hat{\mu}_{3,g} - 3\bar{N}^{-1}\{\hat{\mu}_{22,g} - 2\hat{\mu}_{11,g}^2 - \hat{\mu}_{2,g}^2\} \\
&\quad -\bar{N}^{-1}\{\hat{\mu}_{13,g} - 3\hat{\mu}_{11,g}\hat{\mu}_{2,g}\} + 5\bar{N}^{-2}\hat{\mu}_{12,g} - 2\bar{N}^{-3}\hat{\mu}_{11,g}
\end{aligned} \tag{3.28}$$

where we have set $\hat{\mu}_{0i,g} = \hat{\mu}_{i0,g} \equiv \hat{\mu}_{i,g}$ and $\hat{k}_{0i,g} = \hat{k}_{i0,g} \equiv \hat{k}_{i,g}$.

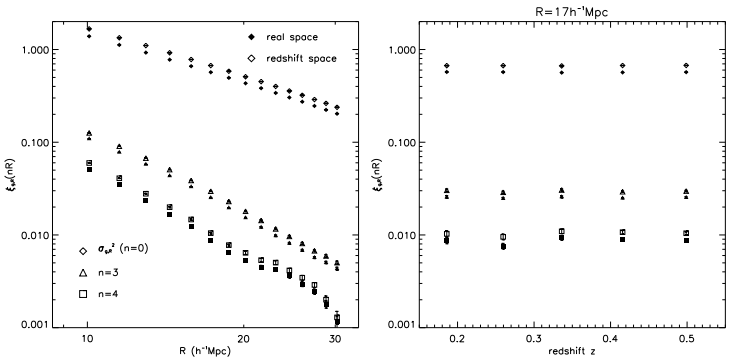

Figure 3.2: *Left:* the correlator of order $(1,1)$ of the smoothed galaxy overdensity field at the average redshift $z = 0.34$ is shown as a function of the smoothing radius R, and for different values of the correlation length ($n = \{0, 3, 4\}$). In the degenerate case $n = 0$ we recover the variance of the galaxy fluctuation field. *Right:* the redshift evolution of $\xi_{g,R}(nR)$ at a given smoothing scale ($R = 17h^{-1}$Mpc) is shown for different values of the correlation length ($n = \{0, 3, 4\}$). The statistics are computed in both real (solid symbols) and redshift space (unfilled symbols). Each point is the average of the results obtained from 8 independent full-sky LRG catalogs. Errorbars are estimated as the standard error of the mean.

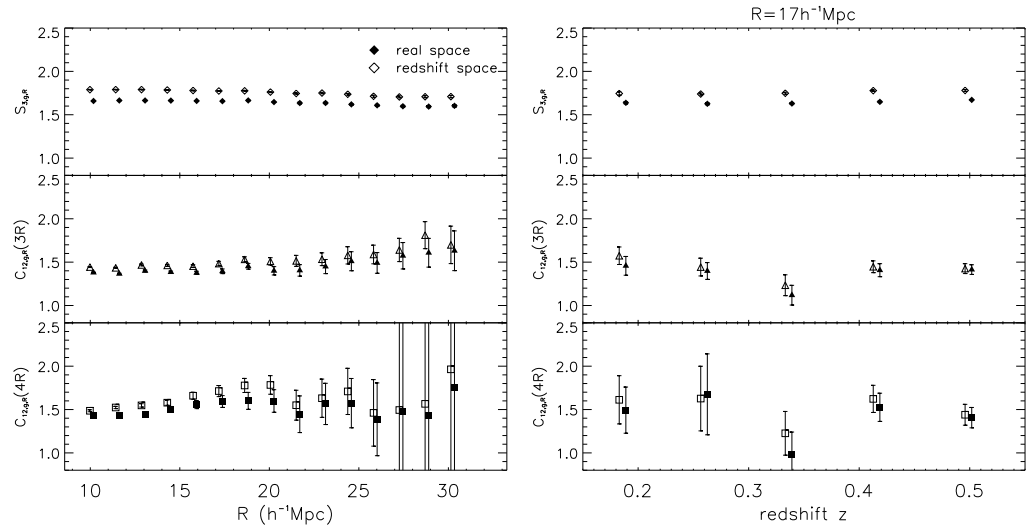

Figure 3.3: *Left:* the reduced correlator of order $(1,2)$ of the smoothed galaxy overdensity field at the average redshift $z = 0.34$ is shown as a function of the smoothing radius R, and for different values of the correlation length ($n = 0$ (upper panel) $n = 3$ (central panel) and $n = 4$ (lower panel)). In the degenerate case $n = 0$ we recover the reduced skewness of the galaxy fluctuation field. *Right:* the redshift evolution of $C_{12,g,R}(nR)$ at a given smoothing scale ($R = 17h^{-1}$Mpc) is shown for different values of the correlation length ($n = \{0, 3, 4\}$). The statistics are computed in both real (solid symbols) and redshift space (unfilled symbols). Each point is the average of the results obtained from 8 independent full-sky LRG catalogs. Errorbars are estimated as the standard error of the mean.

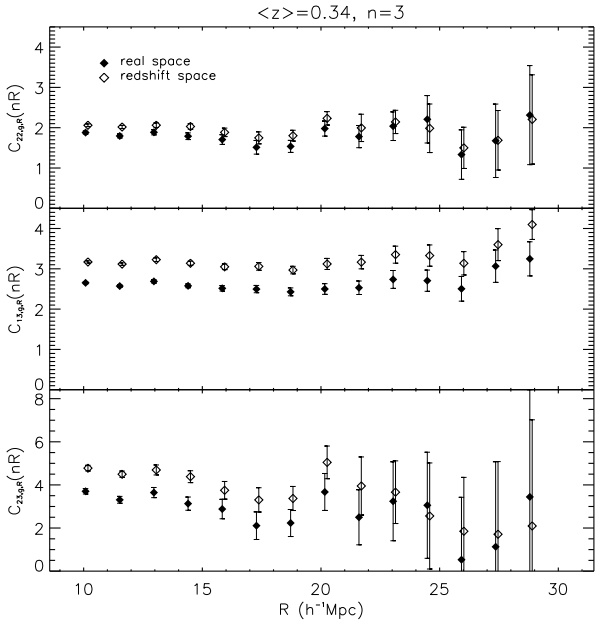

Figure 3.4: Tthe reduced galaxy correlators of order 4 ($C_{13,g,R}$ and $C_{22,g,R}$) and of order 5 ($C_{23,g,R}$) at the average redshift $z = 0.34$ and at the correlation separation $r = 3R$ are shown as a function of the smoothing radius R. The statistics are computed in both real (solid symbols) and redshift space (unfilled symbols). Each point is the average of the results obtained from 8 independent LRG catalogs. Errorbars are estimated as the standard error of the mean.

3.5.2 High order reduced correlators extracted from the Horizon/LasDamas simulations

I use N-body simulations of Luminous Red Galaxies (LRG) of the Sloan Digital Sky Survey (SDSS) to validate the method, to test its end-to-end coherence and to spot the presence of eventual systematics. I have run the whole pipeline on two different numerical experiments, namely Horizon (Kim et al., 2009; Dubinski et al., 2004; Kim & Park, 2006) and LasDamas (McBride et al., 2009). Ideally, this way, the analysis is the least dependent upon the specific simulation strategy and technique.

Horizon simulation

This is a large simulation (4120^3 particles in cubical box) in which LRG galaxies are selected by finding the most massive gravitationally bound, cold dark matter halos. It is characterized by the following set of cosmological assumptions ($\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$, $w_0 = -1$, $w_a = 0$, $H_0 = 72$km/s/Mpc, $\Omega_b = 0.044$, $n_s = 0.96$, $\sigma_8 = 0.79$). I have analyzed 8 nearly independent, full-sky light cones extending over the interval $0.15 < z < 0.55$, each covering a volume of 13 h^{-3}Gpc^3 and containing nearly $3.8 \cdot 10^6$ LRG galaxies. A mass threshold decreasing with redshift was chosen so to force Horizon's comoving number density profile to be constant with redshift and reproduce the density profile observed in the SDSS LRG sample. As a consequence, the simulated sample can be

considered with good approximation as being volume limited. The mean comoving number density is $\bar{n} = 3.0 \times 10^{-4}\ h^3\mathrm{Mpc}^{-3}$ and the mean inter-galaxy separation is $l \sim 19\ h^{-1}\mathrm{Mpc}$. As an example, the average number of LRG galaxies ($\bar{N}$) inside a spherical cell of radius $R = (10, 15, 20, 25)h^{-1}\mathrm{Mpc}$ is approximately $1, 5, 10$ and 20.

In the left panel of Figure 3.2 I show the correlation function of the smoothed galaxy overdensity field $\xi_{g,R} = \langle \delta_{g,R}(\mathbf{x}) \delta_{g,R}(\mathbf{x}+\mathbf{r}) \rangle_c$ for different values of the correlation length ($r = nR$) in both real and redshift spaces. Note that in the degenerate case $n = 0$ we recover the variance of the galaxy density fluctuations on a scale R. If the cell separation increases, the amplitude of the galaxy correlators of order 2 decreases. Their R-scaling is nearly similar to the slope of the analogous statistics computed for the matter density field (see Figure 1.5) with a slope of ~ -2.25 at $R = 10h^{-1}\mathrm{Mpc}$ and ~ -3.4 at $R = 20h^{-1}\mathrm{Mpc}$ for the correlation configuration $n = 3$). Note also the neat appearance of the characteristic baryon acoustic peak at the scale $R \sim 25h^{-1}\mathrm{Mpc}$ when the correlation length is computed for $n = 4$. I interpret these results as a qualitative indication of the fact that, at least on the scales explored by my analysis, the linear biasing parameter is well approximated in terms of a scale-free parameter.

On the right panel of Figure 3.2 I show the redshift dependence of the $\xi_{g,R}$ for a given arbitrary smoothing scale (in this case $R = 17h^{-1}\mathrm{Mpc}$). The constant amplitude of this statistic, together with the fact that the corresponding matter statistics decreases by no more than ~ 0.05 in amplitude over the same redshift interval ($0.15 < z < 0.55$, see Figure 3.8), provide evidences that linear biasing was nearly $\sim 15\%$ stronger at the early epoch $z = 0.55$.

In Figure 3.3 I show the scaling of the 1- and 2-point reduced galaxy cumulant moments of order 3, namely $S_{3,g,R}$ and $C_{12,g,R}(r)$. The slight and systematic decrease of $S_{3,g,R}$ as a function of scales, much less pronounced than that of $S_{3,R}$ (see Figure 1.5), is not compatible with biasing being described by a single constant parameter b_1. Since, we have already argued that b_1 is scale independent, this implies that the biasing function is non linear and the next order biasing coefficient b_2 must show some scale dependency.

LasDamas simulations

This is a suite of large simulations (1280^3 particles in a box of $2400h^{-1}\mathrm{Mpc}$). Among them I choose those simulating the LRG sample of the SDSS (refered to as Oriana). LRG galaxies are simulated as follows: first the most massive gravitationally bound, cold dark matter halos are identified in the matter simulation, then they are populated with artificial galaxies according to an HOD (Halo occupation distribution) statistics modeled to reproduce SDSS observations. The simulation is characterized by the following set of cosmological parameters ($\Omega_m = 0.25,\ \Omega_\Lambda = 0.75,\ w_0 = -1,\ w_a = 0,\ H_0 = 70\mathrm{km/s/Mpc}$, $\Omega_b = 0.04,\ n_s = 1,\ \sigma_8 = 0.8$) which are close to the cosmological parameters of Horizon simulations. The major difference with respect to the Horizon simulation resides in

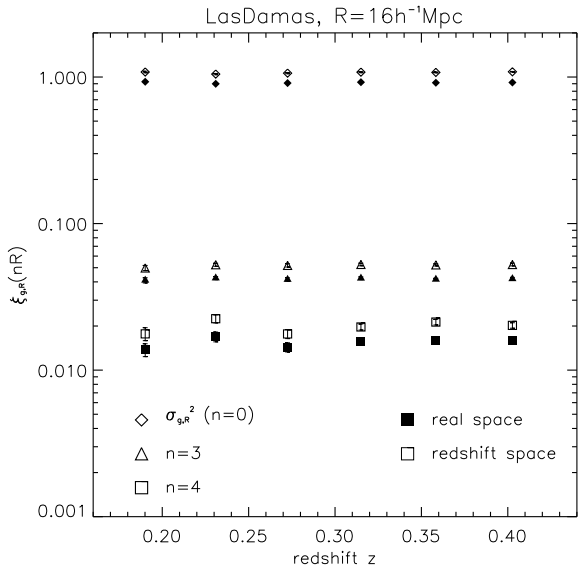

Figure 3.5: The redshift evolution of second order statistics in LasDamas simulations. The filled symbols show the variance ($n = 0$) and the smoothed 2-point correlation ($n \neq 0$) in real-space whereas the empty symbols show the same quantities estimated in redshift-space.

the setting of initial conditions. LasDamas initial conditions are imposed using a 2LPT code, tested by Crocce, Pueblas & Scoccimarro (2006), which is based on second-order Lagrangian perturbation theory. Horizon's initial conditions are set using the Zel'dovich approximation (which apparently might induce transient effects (Crocce, Pueblas & Scoccimarro, 2006) in the clustering of matter). The initial power spectrum of the LasDamas simulations is obtained from CMBfast (Seljak & Zaldarriaga, 1996) and the Oriana simulation, in particular, is started at an initial redshift of z = 49.

In figure (3.5) I show the second order statistics which are constant with respect to the redshift. Figure (3.6) confirms what we already found using the Horizon simulations, that is the statistical quantity $\tau_{g,R}$ is not affected by redshift distortions. This confirms that the approximation $\tau^z_{g,R} = \tau_{g,R}$ can be safely adopted, at least in ΛCDM cosmological models.

Extracting the reduced correlators and moments from the simulations

The analysis of both the Horizon and LasDamas simulations shows that the reduced $C_{12,g,R}(r = nR)$ captures interesting information about the clustering of galaxies (at least on scales R where this indicator is not too noisy). For both the correlation configurations $n = 3$ and $n = 4$ this requirement limits the region of interest to the scales $R \leq 25h^{-1}$Mpc. Despite the fact that measurements on different scales are correlated, it appears that both these statistics are fairly independent from R. We remark that in the LS limit the value of $C_{12,g,R}$ is also independent from the correlation scale r. As a consequence, eq. (3.19)

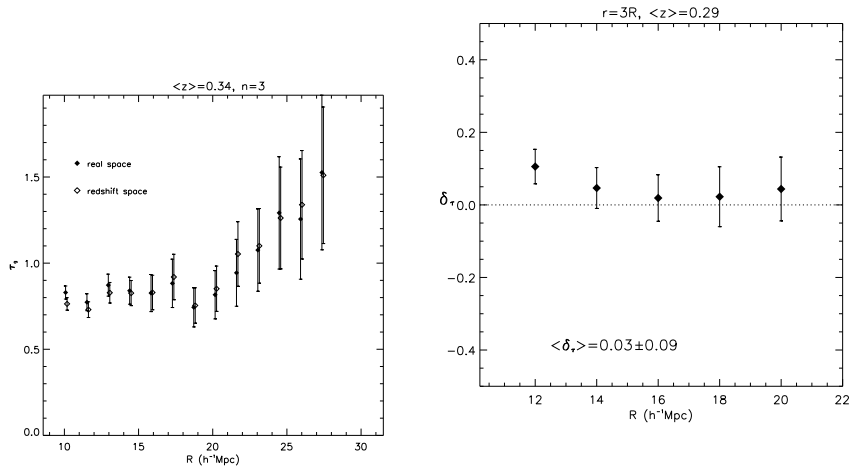

Figure 3.6: *Left:* Comparison between $\tau_{g,R}$ estimated in real space (filled symbols) and in redshift space (empty symbols) as a function of the scale R (at the mean redshift $z = 0.34$ and for a correlation length $r = 3R$) using the LasDAmas simulations. *Right:* Relative difference between the estimates of $\tau_{g,R}$ estimated in real and redshift space.

is best evaluated by adopting the smallest possible value of n, i.e. the one that minimizes the amplitude of the errorbars.

The analysis shows that both the amplitude of $S_{3,g,R}$ and $C_{12,g,R}(r)$ are nearly redshift independent, i.e. mostly independent from the cosmic epoch at which these statistics are computed. This property holds in both real and redshift spaces and mirrors the analogous behavior, predicted by theory, for the reduced cumulant moment of the matter field (see eqs. (1.89) and (1.90).) Because of this we can conclude that, at least up to redshift $z \sim 0.55$, even the next order biasing coefficient, i.e. b_2, is weakly sensitive to time in the Horizon ans LasDamas simulations.

The WNLPT predicts that reduced moments and correlators of the matter field should display hierarchical properties in real space. Figure 3.3 shows that the scaling predicted by eq. (1.87) in section 1.3.8 still holds in redshift space, a results originally found by Lahav et al. 1993 and Hivon et al. 1995 who showed that it holds even on smaller scales than those analyzed here, i.e. on domains where non-linear effects become important. In an analogous way, Figure (3.3) shows that the mapping between real and redshift space also preserves the hierarchical properties of the reduced correlator $C_{12,g}$. This property is not a characteristic of low orders statistics only. In Figure 3.4 I present the estimates of the galaxy reduced correlators up to order 5 in both real and redshift space. Observations in the local universe have shown that the reduced cumulants $S_{N,g,R}$ of the smoothed galaxy field in redshift space display hierarchical clustering properties up to order $N = 6$ (e.g. Baugh et al. 2004). This plot shows that also the reduced correlators $C_{nm,g,R}$ extracted form the Horizon and LasDamas simulation preserve the hierarchical scaling up to order 5 in redshift space.

Finally, Figure 3.4 graphically displays the validity of the factorization property that

I have found in eq. (3.11), i.e. that not all galaxy reduced correlators contain original information. We stress that even if the factorization property $C_{22,g} = C_{12,g}^2$ was shown to hold analytically only in real space, simulations now show that it holds also in redshift distorted space.

I conclude by commenting on the precision of the estimates. The relative error in the estimation of the reduced correlators increases, as expected, with the order of the statistics. Moreover, 2-point statistics are recovered with larger uncertainty than the 1-point statistics of the same order since they are estimated using a smaller number of independent cells. Specifically we find that, the larger the correlation scale, the stronger the sensitivity of the estimates to finite volume effects. On the typical scale $R = 20h^{-1}$Mpc the relative error with which $C_{12,g}$, $C_{13,g}$, $C_{14,g}$, $C_{22,g}$ and $C_{23,g}$ are recovered is 4, 5, 10, 15 and 33% respectively. This can be compared to the precision with which the equal-order reduced moments S_N have been estimated on the same scale, that is 0.7, 2 and 10%, for the order 3, 4 and 5 respectively.

3.5.3 Estimation of $\sigma_R(z)$

In this section I test the efficiency of the estimator given in eq. (3.22). Three major potential issues, if not properly addressed, may affect its reliability. First, it is imperative to test whether we can safely apply WNLPT results in the LS approximation to compute the reduced correlators $C_{nm,R}(r)$ in the limit in which the cell separation is as low as $r \sim 3R$, as the analysis of Bernardeau (1996) suggests. Second, I want to verify that peculiar velocity corrections as well as redshift-space observations do not introduce unexpected biases into the real-space observables. Finally, I want to test if the local Poisson model fairly corrects for the sampling noise in the low counts regime.

To fulfill these goals, I apply the estimator given in eq. (3.22) to the simulated LRG catalogs and gauge the precision with which we can retrieve the real-space amplitude and scaling of σ_R, that is both the local normalization and evolution of the linear matter perturbations embedded in the ΛCDM simulations.

The following argument help to select the range of scales R that are best suited for applying the formalism to the simulated catalogs. I expect that the estimator given in eq. (3.22) will work neatly on sufficiently large scales R (where the WNLPT and the linear modeling of redshift distortions both apply) and on sufficiently large correlation lengths $r = nR$ (where the LS approximation applies). On the smallest scale where the method can be theoretically applied, i.e. $R = 10h^{-1}$Mpc, the amplitude of σ_R, which is of order ~ 0.4 at the average redshift of the sample, is completely dominated by shot noise which is of the order ~ 1. The signal becomes dominant with respect to discrete sampling corrections as soon as R is greater than $\sim 15h^{-1}$Mpc. Moreover, below this last scale, a small, but statistically significant imprecision arises in assuming $\tau_{g,R} = \tau_{g,R}^z$, as shown in Figure 3.1. On the opposite end, the largest scale R accessible is set by the

geometry of the survey and the requirement of sampling the correlation length $r = nR$ with sufficient statistical power. I find that the relative error in my estimate of $C_{12,g,R}(r)$ becomes larger than $\sim 10\%$ (see Figure 3.3) for scales $R > 22h^{-1}$Mpc when $n = 3$ and for scales $R > 18h^{-1}$Mpc when $n = 4$.

In Figure 3.7 I plot the estimates of σ_R based on LRG mock catalogs. I recover the rms of the linear matter fluctuation field on two smoothing scales ($R = 17$ and $20h^{-1}$Mpc) and for three different correlation lengths ($n = 2$, $n = 3$ and $n = 4$). By contrasting my measurements against the theoretical predictions obtained by inserting into eq. (1.79) the parameters used in the Horizon simulations, I find that my reconstruction scheme fails when the correlation length is as low as $n = 2$. This was expected since Bernardeau (1996) already showed that the LS approximation does not hold on such small correlation scales. Effectively, when I probe larger scales ($n = 3$), the reconstruction becomes significantly more accurate (central panel of Figure 3.7), with the estimates of σ_R at the average redshift of the catalog ($z = 0.34$) being affected by a relative error of 13% and 15% on the scales $R = 17$ and $20h^{-1}$Mpc respectively. For $n = 4$ errorbars become too large for the estimates to be also precise. Additionally, we remark that my estimates seem to slightly overestimate the value of σ_R on both the scale analyzed. This is also confirmed by the analysis of an independent mock catalog, i.e. the Las Damas simulation (McBride et al., 2009). As stressed in section 1.3.9, this is due to the increasing inaccuracy in the theoretical prediction of the amplitude of the reduced correlators $C_{12,R}$ on correlation lengths that approaches the scale where ξ_R crosses zero.

I re-emphasize that these estimates are totally independent from any assumption about shape and normalization of the power spectrum of linear matter fluctuations. They are also independent from the amplitude of the present day normalization of the Hubble parameter H_0. My estimates, however, do depend on the set of cosmological parameters $(\Omega_M, \Omega_\Lambda)$ that I have used to assign galaxies to cells (i.e. to smooth the galaxy distribution), and to subtract the effect of redshift space distortions (i.e. to evaluate the growth rate function f).

3.5.4 Estimation of the local value of σ_8

As I have already discussed, the scale $R = 8h^{-1}$Mpc falls outside the range of applicability of the test. Nonetheless, I can extract information about the value of σ_8 at redshift $z \sim 0$ ($\sigma_8(0)$) from measurements of σ_R on larger scales. I do this by fitting eq. (1.79) to data. The price to pay is that the recovered value will depend on the adopted power spectrum model and on the set of parameters on which the power spectrum itself depends, i.e. the reduced Hubble constant h, the primordial spectral index n_s, and the reduced density of baryons (Ω_b). This approach, however offer some advantage: if the relevant cosmological parameters Ω_m, Ω_Λ and h are considered known from independent probes, then I can extract information about the purely gravitational sector of the theory.

The way my proceed is as follows: I assume standard gravity as described by GR, I frame my analysis in the linear regime, i.e. I adopt the phenomenological description of the matter power spectrum given by Eisenstein & Hu (1998), as well as the time evolutionary model for σ_R given in eq. (1.79), and I look for the the best fitting parameter n_s, Ω_b and $\sigma_8(0)$ that minimize the statistical distance between my measurements (cfr. eq. (3.22)) and theoretical predictions (cfr. eq. (1.79)). The outcome of this approach is displayed in Figure 3.9.

If I assume that n_s and Ω_b are fixed to the simulation's values ($n_s = 0.96, \Omega_b = 0.044$), I obtain $\sigma_8(0) = 0.79 \pm 0.08$. This best fitting value is in perfect agreement with the simulated one ($\sigma_8 = 0.79$). Results obtained by performing a joint two-parameter analysis (after fixing the third parameter to the simulated value) are shown in Figure 3.9. The left and central panel of this figure reveal that $\sigma_8(0)$ is only marginally degenerate with respect to both n_s and Ω_b, a fact that highlights the fundamental inefficiency of our probe in constraining the values of both these parameters. This conclusion is reinforced in the right panel of the same figure, which displays a strong degeneracy between n_s and Ω_b together with a loosely constrained confidence region in the corresponding parameters plane. Luckily, this means that the uncertainties with which both these parameters are estimated using more sensitive probes, do not critically affect the precision with which our method constrain the amplitude of $\sigma_8(0)$.

Despite my analysis was performed by slicing the survey volume in independent redshift shells, the limited redshift interval explored allows me to fix only the local amplitude of the linear matter fluctuation field. In a future work I will show that by implementing the method with deeper mock catalogs simulating the region of space that will be surveyed spectroscopically by surveys like BigBOSS and EUCLID, one can further aim at constraining the time evolution of σ_R. Data on a larger redshift interval will allow to constrain not only $\sigma_8(0)$, but also the growth index γ in terms of which the growth rate is usually parameterized ($f = \Omega_m^\gamma(z)$ Peebles (1980)). This will allow to reject possible alternative description of gravity, or, in turn the standard model of gravitation itself.

3.5.5 Consistency tests

I have shown that, by using simulations, it is pretty straightforward to assess whether the proposed measuring strategy is able to recover the underlying value of σ_R. What if, instead, a real redshift survey is considered? Are there specific physical criteria or statistical indicators that guarantee us that the recovered value of σ_R is the *true* one? In other terms I want to shift our attention from the precision of the estimates to their accuracy. Apart from the unbiasedness of the WNLPT results in the LS limit, my test strategy strongly relies on assuming that the correct set of cosmological parameters (Ω_M, Ω_Λ) has been used in the analysis. As a consequence, any imprecision in the measurements of the reduced cosmic densities translates into a biased estimate of σ_R. In this section, we

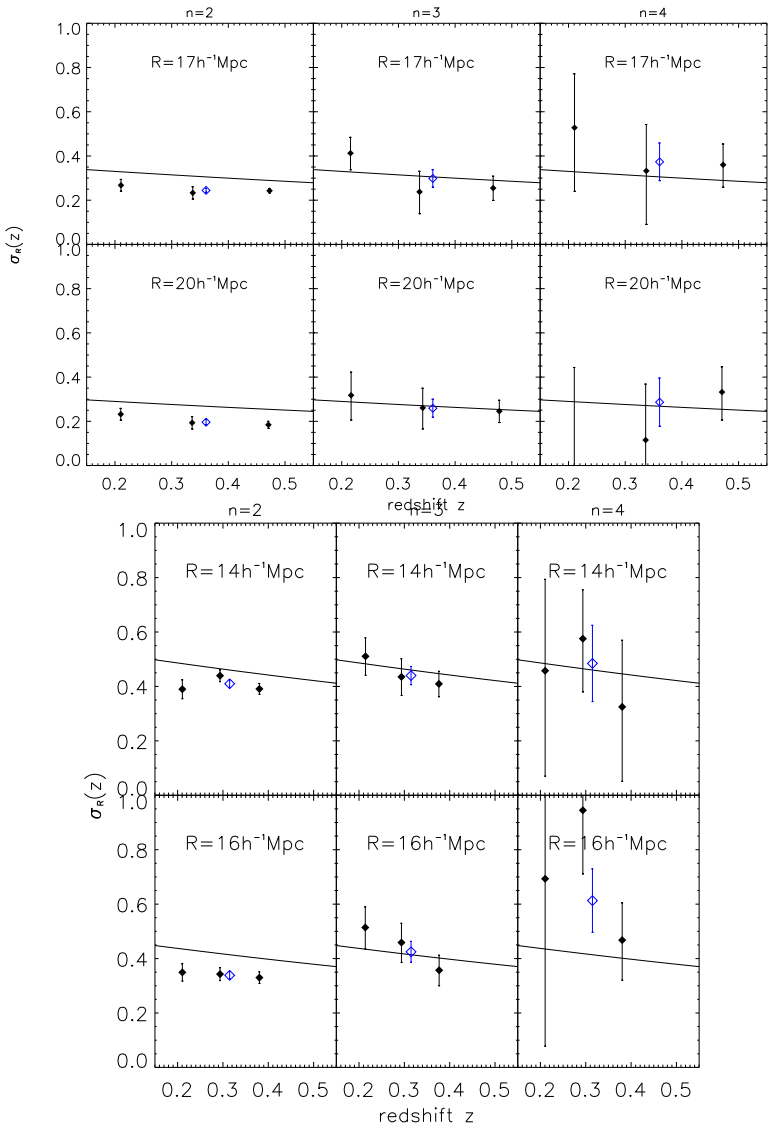

Figure 3.7: Upper plot corresponds to measurements from Horizon and lower corresponds to LasDamas. For both, the real-space *rms* of the matter fluctuation field is shown as a function of redshift. Black filled points represent estimates obtained by implementing eq. (3.22) to the LRG mock catalogs. Solid lines show the theoretical prediction obtained by inserting all the physical and cosmological parameters of the Horizon simulation into eq. (1.79). Count in cell as well as cell correlation analysis have been performed by assuming spherical cells of radius R separated by correlation lenght $r = nR$. We present results obtained for $n = 2$ (left panels), $n = 3$ (central panels) and $n = 4$ (right panels) using two typical cell sizes $R = 17(/14)h^{-1}$ Mpc (upper panels) and $R = 20(/16)h^{-1}$Mpc (lower panels). Each point is the average of the results obtained from eight independent full-sky LRG catalogs containing nearly 3.8 million galaxies each. Errorbars are estimated as the standard error of the mean. At the mean redshift of the catalogs ($z = 0.34$) we also display the average estimate of σ_R obtained in the whole survey volume (unfilled blue points).

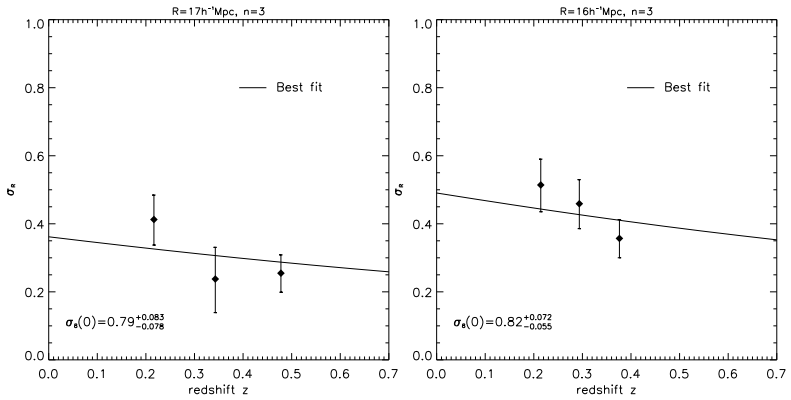

Figure 3.8: The real space *rms* of the linear matter density fluctuations on a scale R is shown as a function of redshift. Left panel shows the value of σ_R extracted from Horizon whereas the right panel shows the one from LasDamas. In both panels the solid curve represents the best fitting theoretical model for the linear scaling of σ_R (cfr. eq. (1.79) obtained after fixing the amplitude of Ω_b, n_s and h to the values of the simulations. The corresponding best fitting value of $\sigma_8(0)$ is shown in the inset together with its standard deviation.

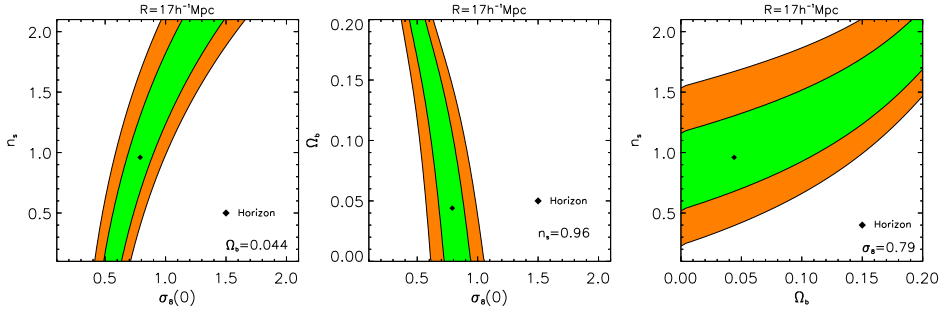

Figure 3.9: The likelihood contours obtained from the joint estimation of $(\sigma_8(0), n_s)$, $(\sigma_8(0), \Omega_b)$ and (Ω_b, n_s). The analysis have been performed calculating reduced correlators on a scale $R = 17h^{-1}$Mpc and at the correlation lenght $n = 3$. Isocontours of differently colored region corresponds to $\mathcal{L}_{min} + 2.30$ and $\mathcal{L}_{min} + 4.61$ where $\mathcal{L}$ is proportional to the log of the likelihood of the data and it is here assumed to be affine to the χ^2 statistic.. On each panel the filled points represent the fiducial value of the Horizon simulation.

design a diagnostic scheme to test the coherence of our results and, at the same time, the soundness of the adopted values for Ω_M and Ω_Λ.

In the approach developed in this thesis, the linear biasing parameter in real space is directly estimated from redshift space observables of intrinsic third-order nature using the estimator

$$b_{1,R} = \frac{\alpha^z_{g,R}}{\tau^z_{g,R}}. \tag{3.29}$$

As discussed in Section 3.4, this estimator has the remarkable property of being approximately independent from cosmology.

Now let's define two new estimators of the real space linear biasing parameter as $\mathring{b}_{1,R} = \sigma_{g,R}/\sigma_R$ and $\tilde{b}_{1,R} = \sqrt{\xi_{g,R}/\xi_R}$, where now I exploit, as it is usual, second order statistics.

Both $\sigma_{g,R}$ and $\xi_{g,R}$ are quantities not directly measurable, nonetheless we can recast the expressions of the real space linear biasing parameters in terms of redshift space observables. By adopting the Kaiser model for linear motions we obtain

$$\mathring{b}_{1,R} = -\frac{f}{3} + \sqrt{\left(\frac{\sigma^z_{g,R}}{\sigma_R}\right)^2 - \frac{4}{45}f^2} \tag{3.30}$$

$$\tilde{b}_{1,R} = -\frac{f}{3} + \sqrt{\frac{\xi^z_{g,R}}{\xi_R} - \frac{4}{45}f^2}. \tag{3.31}$$

Before proceeding further, note that in this thesis I have assumed that the Kaiser linear modeling of redshift space distortions applies on the scales we are interested in. I can now verify this statement by using the LRG synthetic catalogs. To this purpose, and without loss of generality, be $\tilde{b}_{1,R}$ the value of the real-space linear bias parameter estimated from the mock catalogs as $\sqrt{\xi_{g,R}/\xi_R}$. Let's refer to this estimate as to the *true* value of the linear-bias parameter and let's label it as b_1^{th}. In figure 3.10 I compare these measurements against the estimates of the real-space linear bias inferred using eq. (3.31) in three different redshift intervals and for various smoothing scales R. One can see that for $n = 3$, the estimator (3.31) fairly recovers the real-space value of the linear bias parameter. This result lends support to the hypothesis that, at least in the Horizon simulations, and over the range of R scales where we trust the correlator's theory, the distortions induced by large scale peculiar motions are accurately described by the Kaiser model.

Physically, we expect that, whatever is the chosen scale R, if $\sigma_8(0)$ has been consistently determined, then eqs. (3.29), (3.30) and (3.31) all give the same numerical result. Since the linear biasing estimators given in eqs. (3.30) and (3.31) depends on the chosen background cosmology, the correct set of cosmological parameters is thus the very one that makes all the three different linear biasing definitions converge to the same numerical value on all the scale R. In the following Chapter I show how this observation can be exploited to guess the background cosmology. In this Chapter, I use this property to gauge the consistency of my measurements of σ_R, that is to verify that all the different estimators of the linear biasing parameters match only when the analysis is carried out in the proper cosmological background.

In Figure 3.11 I perform this test and show what pathological features do show up when the analysis relies on an improperly chose set of values $(\Omega_M, \Omega_\Lambda)$. The algorithm goes as follows: I estimate σ_R on a given arbitrary scale R (here we chose $R = 16h^{-1}$Mpc), and in four different cosmologies (indicated in each panel of Figure 3.11). I then deduce the value of $\sigma_8(0)$ in each of these four scenarios. In doing this, I implicitly assume that the specific cosmology adopted in order to measure $\sigma_{R=16}$ is the correct one and that, as a consequence, the value of $\sigma_8(0)$ inferred using any other scale R *is identically the same*. I then plug in this value of $\sigma_8(0)$ into eqs. eqs. (3.30) and (3.31) and compare the results

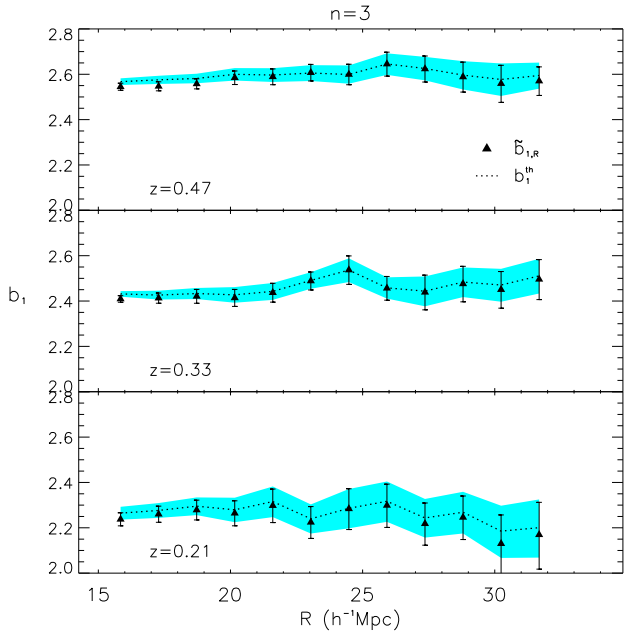

Figure 3.10: Black triangles show the R dependency of the real-space linear bias parameter $\tilde{b}_{1,R}$ estimated from redshift-space mock catalogs using eq. (3.31). Measurements are performed using the correlation length $n = 3$ and in three different redshift bins. These measurements are compared to the real-s[ace linear bias parameter b_1^{th} extracted from real-space mock catalogs using the estimator $\sqrt{\xi_{g,R}/\xi_R}$ (dotted line). The shaded area represents the region of 1-σ uncertainty.

with those obtained via the estimator $b_{1,R}$, i.e. we contrast them against a measurement of the linear biasing parameter that is weakly sensitive to cosmology.

If I analyze the LRG mock catalogs by assuming the simulated set of cosmological parameters ($\Omega_M = 0.26, \Omega_\Lambda = 0.74$, see upper left panel of Figure 3.11) then the estimates of $\mathring{b}_{1,R}$, $\tilde{b}_{1,R}$ and $b_{1,R}$ are consistent between themselves on all the scales R. On the contrary, if I process data by incorrectly assuming a low density (open) background model (upper right panel in Figure 3.11), the different estimations of the linear biasing parameter are not anymore in agreement, even if the estimated value of $\sigma_8(0)$ has not changed (note that the linear power spectrum at present epoch is independent from the amplitude of the cosmological constant and thus insensitive to its variation). In particular, $\mathring{b}_{1,R}$ identically coincides, by definition, with the measure obtained using my estimator $b_{1,R}$ only for $R = 16h^{-1}$Mpc , but the estimates deviate on all the other scales. This discrepancy is due to the fact that, in a "wrong" cosmology, the extrapolated value of $\sigma_8(0)$ may not be independent from the scale R on which σ_R is measured. Or said differently, the equation $\mathring{b}_{1,R}(\sigma_8(0)) = b_{1,R}$ might not have a unique solution $\sigma_8(0)$ for all the scales R.

The effect previously discussed is mostly the consequence of adopting the wrong amplitude for the cosmological constant. It is, however, less pronounced than the discrepancy between the measurements of $\mathring{b}_{1,R}$ and $\tilde{b}_{1,R}$ that arises when the value of Ω_m is poorly guessed. The effects of a wrong choice of the matter density parameter are presented in the lower panels of Figure 3.11. The observed large inconsistency arises essentially from

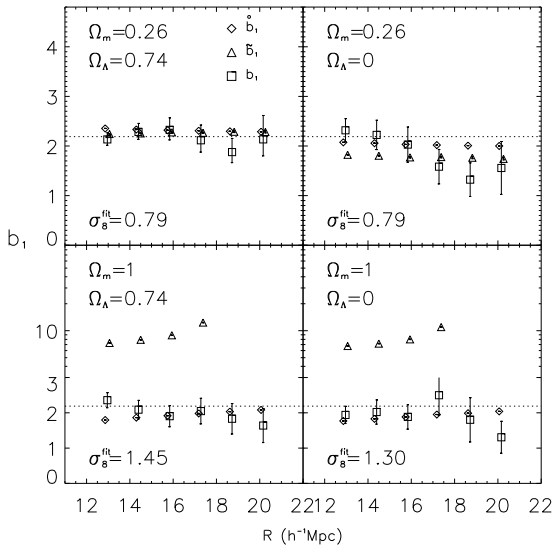

Figure 3.11: Comparison between the values of the linear bias parameter on various scales R for the Horizon simulation. Measurements are obtained using 3 different estimators (cfr. eqs. (3.29), (3.30), and (3.31)) and by analyzing the same redshift data in 4 different cosmological backgrounds as specified in the panels. $\mathring{b}_{1,R}$ and $\tilde{b}_{1,R}$ are estimated on the basis of the best fitting value of $\sigma_8(0)$ indicated in each panel. This last value has been obtained from measurements of σ_R on a scale $R = 16h^{-1}$Mpc and by assuming, arbitrarily, that the linear power spectrum parameters are subject to the following constraints: $h = 0.72$ and $\Omega_b/\Omega_m = 0.17$ for all the cosmological models. Note that in the lower panel, $\tilde{b}_{1,R}$ data are missing on a scales $R = (19, 20)h^{-1}$Mpc. This is due to the fact that the argument of $\sqrt{\xi_{g,R}/\xi_R}$ is negative, i.e. in these extreme cosmological models, the predicted correlation function of the matter density field becomes negative on scales where the corresponding galaxy statistic is still positive. The horizontal dotted line shows the average of the measurements obtained with our estimators $b_{1,R}$ in the *true* cosmological model (i.e. the ΛCDM model of the Horizon simulation) and it is reproduced identically in all the panels. This line graphically helps to highlight the relative insensitivity of our estimator $b_{1,R}$ to the choice of the parameters $(\Omega_M, \Omega_\Lambda)$. Note how different estimates converge to the same value only when the analysis is performed in the *true* cosmological model.

the fact that the zero order spherical Bessel function appearing inside the integral in eq. (4.3) filters in different portions of the signal (i.e. of Δ^2_L) if the characteristic parameters of the linear power spectrum are changed. In other terms, if we consider eq. (4.3) and spuriously overestimate Ω_m, the predicted suppression of power on a scales r is larger than the variation actually seen in the data.

We remark that the estimator of the linear biasing parameter, relying on third-order statistical indicators, is affected by errorbars that are larger than those associated to the classical estimators $\mathring{b}_{1,R}$ and $\tilde{b}_{1,R}$. This imprecision is largely compensated by the fact that my estimator does not depend on any assumption about the nature of the dark matter (i.e. the specific form of the matter power spectrum) and it is almost insensitive also to its abundance (Ω_M). Therefore, by contrasting different estimations of the linear biasing parameter, that is using the diagnostic diagram of Figure 3.11, we can deduce if the scaling of σ_R was inferred in the appropriate cosmological model.

3.6 Accuracy test

The whole strategy to extract the rms of the matter fluctuation from redshift survey data, ultimately relies on the accuracy with which WNLPT predicts the amplitude of the reduced moment $S_{3,R}$ and of the reduced correlator $C_{12,R}$. While the accuracy of theoretical predictions for $S_{3,R}$ has been extensively tested (see Bernardeau et al., 2002, for a review), how well the amplitude of $C_{12,R}$ is reconstructed is not entirely clear, and different claims exists in the literature. (Bernardeau, 1996) found a good agreement between different theoretical ways of computing the reduced 2-point cumulants of order $(1, 2)$ once the LS condition is satisfied (see figure 3 in Bernardeau, 1996, or explicit calculation in appendix C). An opposite conclusion is reached by Gaztañaga, Fosalba & Croft (2002), who compared theoretical predictions against a different set of simulations.

In order to address this issue bypassing the systematic effects induced by halo biasing, it is necessary to estimate $C_{12,R}$ directly from matter particle simulations. Thanks to the courtesy of Enrique Gaztañaga I was given the access to the Marenostrum Institute de Ciències de l'Espai Simulations (hereafter MICE simulations). In particular, to the specific simulation with the best resolution: 4096^3 dark matter particles of typical masse $2.927 \times 10^{10} h^{-1} M_{\odot}$ contained in a simulation box of $3072^3 h^{-3}\mathrm{Mpc}^3$. This numerical experiment was run using the following set of cosmological parameters

$$\begin{cases} \Omega_m = 0.25 \\ \Omega_\Lambda = 0.75 \\ w_o = -1 \\ w_a = 0 \\ h = 70 \\ ns = 0.95 \\ \sigma_8(0) = 0.7 \end{cases} .$$

I had access to two comoving boxes representing the simulation output at $z = 0.5$ and $z = 1.5$. In order to handle the large number of matter particles, I randomly sampled the particles contained in each box with the sampling probability of one over seven hundred. I also analyzed a light cone (stretching up to reshift 1.4) to test for any evolutionary effects into the WNLPT predictions. In Figure (3.12) I show the $C_{12,R}$ statistics as predicted from theory (i.e. WNLPT) and as recovered from simulated data for various correlation ratios $n = 2, 3, 4$ (where and $n = r/R$) and at two different redshifts. These results confirm that, when the correlation length r is small ($n = 2$), the approximated results from WNLPT do not match the fully non-linear results from numerical experiments. However, as soon as $n = 3$, as predicted by Bernardeau (1996), we find substantial agreement between theory and numerical data. Anyway, there seems to be a residual systematic discrepancy of about 5% at redshift 0.5 which could mean that the choice $n = 3$ still poorly represent the LSL condition. This interpretation is supported by the fact that not only the agreement is

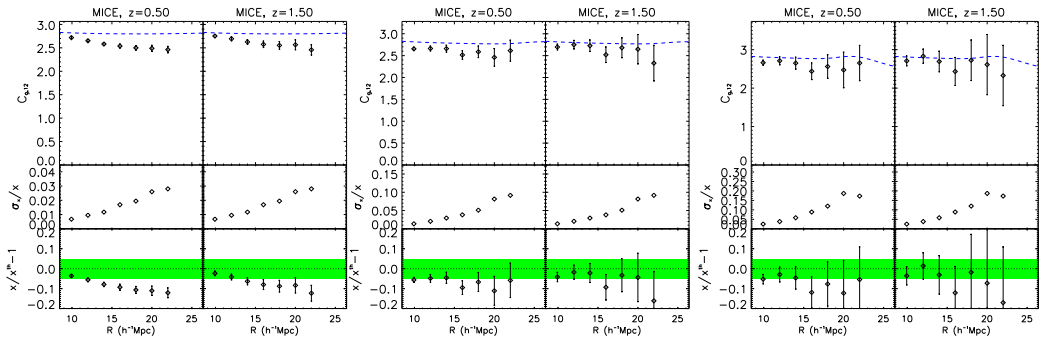

Figure 3.12: $C_{12,R}(nR)$ from MICE simulation (diamonds with error bars) compared to WNLPT prediction (dashed blue line). The intermediate panel shows the relative error of the measurement, obtained by dividing the error on the estimation of $C_{12,R}(nR)$ (here simply indicated as σx by the measured value of $C_{12,R}$ (here simply indicated as x). Error bars on the estimations have been obtained by 64 Jackknife resampling of the complete volume. The lower panel shows the relative deviations between mesurements (x) and theoretical prediction (x_{th}) (the green area represents the region where deviations are less that $\pm 5\%$. The left, central and right panels show results obtained assuming $n = 2, n = 3$, and $n = 4$ respecively.

better for $n = 4$, but also that, for a fixed value of n, the agreement is tighter at high redshift, where non-linear effects are expected to be smaller. A different interpretation of the small residual effect could be that particles in the simulations are still too massive to represent in a fair and unbiased way the continuous perfect fluid of matter. In this case, finite size effects would be responsible for the observed residual shift.

3.7 Conclusions

A key goal that seems to fall within the technical reach of future cosmological experiments is to rule out (or in) eventual infrared modifications of the standard theory of gravity, as possibly manifested by the unexpected growth of cosmic mass in the weakly non-linear, low-curvature, high-redshift regimes. To fulfill this task, it is mandatory to devise sensible observables of the large scale structure of the universe. In this spirit, this paper focuses on the potential of clustering indicators that have been rarely explored in literature , that is the high-order reduced correlators C_{nm} of the 3D mass overdensity field (bu see Szapudi, Szalay & Boschán, 1992; Szapudi et al., 1995, for pioneeristic 2D analyses).

To fully exploit the richness of information contained in this 2-point statistics, whose amplitude is analytically predicted by the weakly non-linear perturbation theory in the large separation limit approximation, I have derived the expressions of the reduced correlators of the smoothed galaxy density field ($C_{nm,g}$) up to order 5. I have found that they preserve both the hierarchical scaling and the factorization properties of the reduced correlators of matter.

Building upon these results I have worked out the explicit expressions for the bias coefficients up to order 4 and a new estimator to measure the rms of the linear matter

fluctuations on a scale R directly from galaxy redshift surveys. The central results of this Chapter, namely the estimator given in eq. (3.22), was tested using artificial galaxy catalogs and shown to recover fairly well the 'hidden' simulated value of σ_R.

Despite the fact that very large survey volume are needed to make the estimation of these observables accurate enough for cosmological purposes, the merit of this approach are evident: *a)* linear biasing is not a parameter that one needs to marginalize over, but a physical parameter that can be estimated in a totally independently way from any assumption about the structure of the linear power spectrum and the value of cosmological parameters. *b)* the real space linear biasing parameter can be measured directly using redshift space observables. *c)* The scaling of σ_R can be inferred directly without imposing any *a-priori* constraint on the eventual non-linear and scale dependent nature of the bias function. *d*) The correlator formalism allows also for a self-consistent test of the coherence of the results obtained concerning the scaling of σ_R, a step further in the direction of making cosmological results not only precise but also accurate.

In this Chapter I have analyzed local simulations with the aim of testing principles and theoretical ingredients on which the proposed strategy relies. Work is already in progress to apply the formalism to the SDSS-dr7 data and to extract the local value of linear matter fluctuations on sensible scales R. The good news is that the next decade holds even greater prospects for growth of the red-shifts data base. Therefore, we also plan to implement the algorithm to mock catalogs simulating future large 3D surveys such as BigBOSS and EUCLID and forecast up to what order and precision the bias coefficients b_i can be estimated, as well as, the figure of merit achievable on σ_8 and on the gravitational growth index γ.

From the theoretical side, valuable insights are expected from a reverse engineering on the proposed test. For example, I show in the next Chapter that if the gravitational model is *a-priori* known then the formalism offers the possibility of narrowing in on the value of fundamental cosmological parameters such as Ω_m and Ω_Λ. Also further work is needed to understand how higher order real-space biasing parameters can be effectively retrieved from redshift space reduced correlators. Finally, interesting possibilities will open up if WNLPT predictions could be extended into the small separation limit where much more statistical power is locked.

Chapter 4

A new cosmic probe: The clustering ratio of galaxies

In this chapter I consider the abundance of dark matter in the universe. This a problem of great observational and interpretative difficulty, but owing to its bearing on fundamental physics, is one of great theoretical significance.

Considerable effort is currently devoted to devising new and sharpening known methods for determining the space density and nature of these elusive particles. So important is the goal that a pioneeristic space mission (EUCLID) will soon be launched to map the large scale distribution of dark matter.

The standard technique to explore the dark sector of the universe, and the workhorse of the EUCLID mission, exploits the baryon acoustic oscillations imprinted in the 3-dimensional distribution of galaxies (BAO probe). I here demonstrate that accurate and precise cosmological information can be extracted from the cell count analysis of the 3D spatial clustering of galaxies once the second-order ratio between one- and two-point moments of the smoothed galaxy density distribution is analyzed.

This new approach is, probably, the simplest ever method that allows to constrain the amount of dark matter via the distribution of luminous galaxies. Despite its simple engineering, it allows to explore with higher resolution the same cosmological parameter space to which the BAO technique gives access.

This probe does not require the calibration of any standard rod, the modeling of the galaxy biasing or redshift distortions and the knowledge of the sample selection function. Its accuracy and precision of this new cosmic probe are demonstrated by an end-to-end analysis of cosmological simulations and by direct implementation to real data.

Using the spectroscopic Sloan Digital Sky Survey (SDSS) data release 7 (DR7) galaxy sample, no cosmic microwave background (CMB) information, no hypotheses on the curvature of the universe (Ω_k) nor on the constant value of the dark energy equation of state (w), we estimate the abundance of matter (Ω_m) with a relative error of 11% (at 68% c.l.). I also present cosmological constraints obtained by applying this probe to high redshift

data from the VIMOS public extragalactic redshift survey (VIPERS). The method may be instrumental in searching for evidences of new physics beyond the standard model of cosmology and in planning future missions such as BigBOSS or EUCLID.

4.1 The method

The challenge of determining the value of the constitutive parameters of the Friedmann equations can be addressed through the development of performant cosmic probes, sophisticated error analysis schemes, and formidable observational programs from both ground and space (Amendola et al., 2012). The Wilkinson microwave anisotropy probe (WMAP), for example, fix with impressive accuracy the parameters of a 'power-law ΛCDM' model (Larson et al., 2011). This is a cosmological model characterized by a flat geometry, a positive dark energy component Ω_X with $w = -1$, and by primordial perturbations that are scalar, Gaussian and adiabatic (Weinberg, 2008). Notwithstanding, other astrophysical probes are needed if we are to constrain deviations from this minimal model (Frieman at al., 2008). Different approaches, being sensitive to different sets of nuisance parameters and to different subsets of the full cosmological parameter set, provides consistency checks, lift parameter degeneracies, enable stronger constraints and, in the end, a safer theoretical interpretation.

I characterize the inhomogeneous distribution of galaxies in terms of the local dimensionless density contrast $\delta_g(\mathbf{x}) \equiv \rho_g(\mathbf{x})/\langle\rho_g(\mathbf{x})\rangle - 1$, where $\rho_g(\mathbf{x})$ is the comoving density of galaxies in configuration space, and where its average is computed over the whole galaxy sample. Since the galaxy distribution is a stochastic point process, a spherical top-hat filter W of radius R is applied to generate a continuous, coarse-grained observable $\delta_{g,R}(\mathbf{x}) = \int \delta(\mathbf{y})W(|\mathbf{x}-\mathbf{y}|/R)d\mathbf{y}$. The 1-point $\kappa_{20,g,R} = \langle\delta^2_{g,R}(\mathbf{x})\rangle_c$ and the 2-point $\kappa_{11,g,R}(\mathbf{r}) = \langle\delta_{g,R}(\mathbf{x})\delta_{g,R}(\mathbf{x}+\mathbf{r})\rangle_c$ cumulant moments are the lowest-order non-zero connected moments of the probability distribution functional (PDF) of the field $\delta_{g,R}(\mathbf{x})$ (Szapudi, Szalay & Boschán, 1992; Bernardeau, 1996; Bel & Marinoni, 2012). If the PDF is stationary and isotropic, the quantity $\kappa_{11,g,R}(\mathbf{r})/\kappa_{20,g,R}$ reduces to the ratio between the correlation function and the variance of the filtered field

$$\eta_{g,R}(r,\mathbf{p}) = \frac{\xi_{g,R}(r,\mathbf{p})}{\sigma^2_{g,R}(\mathbf{p})}, \tag{4.1}$$

where I have explicitly emphasized the dependence of the observable on the set $(\mathbf{p})$ of cosmological parameters. This the case because the LHS term of eq. (4.1) can be estimated from data only once a comoving distance-redshift conversion model is supplied.

It is straightforward to predict the expected value of the second order galaxy clustering ratio $\eta_{g,R}$. On sufficiently large cosmic scales R, matter fluctuations are small and described by the linear (dimensionless) power spectrum $\Delta^2_L = 4\pi A k^{n_s+3}T^2(k)$, where A is a normalization factor, n_s is the primordial spectral index, and $T^2(k)$ is the transfer

function (Eisenstein & Hu, 1998). Accordingly, the amplitudes of the second-order statistics for mass evolve as a function of redshift (z) and scale as $\sigma_R^2(z) = \sigma_8^2(z=0)D^2(z)\mathcal{F}_R$, and $\xi_R(r,z) = \sigma_8(z=0)^2 D^2(z)\mathcal{G}_R(r)$. Both these equations are normalized at the scale $r_8 = 8h^{-1}$Mpc, $D(t)$ represents the linear growing mode (Weinberg, 2008), while the effects of filtering are incorporated in the functions

$$\mathcal{F}_R = \frac{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kR)d\ln k}{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kr_8)d\ln k} \tag{4.2}$$

$$\mathcal{G}_R(r) = \frac{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kR)\frac{\sin(kr)}{kr}d\ln k}{\int_0^{+\infty} \Delta_L^2(k)\hat{W}^2(kr_8)d\ln k} \tag{4.3}$$

where $\hat{W}$ is the Fourier transform of the window function.

By analogy, I define the mass clustering ratio $\eta_R(r) \equiv \xi_R(r)/\sigma_R^2$, which is explicitly given by

$$\eta_R(r) = \frac{\mathcal{G}_R(r)}{\mathcal{F}_R}. \tag{4.4}$$

As long as a linear matter power spectrum is assumed, $\eta_R(r)$ is stable across cosmic time, i.e. it does not evolve as a function of redshift. Moreover, it is insensitive to the amplitude of the matter power spectrum normalization parameter A. Interestingly, it is also unaffected by redshift distortions, and, therefore, independent of their specific modeling, that is the ratio $\eta_R(r)$ between second-order statistics has identical amplitude in both configuration- and real-space. If the only net effect of peculiar velocity is the enhancement of the amplitudes of the density ripples in Fourier space, without any change in their phases or frequencies, then redshift distortions equally contributes to the numerator and denominator of eq. (4.4). For any given R and r, hence, the ratio $\eta_R(r)$ behaves as a cosmic 'standard of clustering'.

The relationship between the mass and galaxy clustering ratios in configuration space follows immediately once we specify how well the overall matter distribution is traced by its luminous subcomponent on a given scale R. By adopting an arbitrary local non-linear biasing scheme, namely $\delta_{g,R}(\mathbf{x}) = \sum_{i=0}^{N}(b_i/i!)\delta_R^i(\mathbf{x})$, and in the limit $\sigma_{g,R}^2 << 1$, I obtain (see previous Chapter)

$$\eta_{g,R}(r) = \eta_R(r), \tag{4.5}$$

an identity that is independent of the specific value of the bias amplitudes b_i. In figure (4.1) I show how the observable $\eta_{g,R}(nR)$ extracted from 160 LasDamas mock catalogs simulating the space distribution of SDSS galaxies (the Oriana lighcones), compares to both the linear and non-linear theoretical predictions for $\eta_R(nR)$. The comparison shows that the agreement between theory and data is extremely good on a wide range of scales, from $n_{min}R_{min} = 24$ up to $n_{max}R_{max} = 120\ h^{-1}$Mpc. In particular, on scales $R > 10h^{-1}$Mpc and $r = 3R$, the approximation (4.5) is good to better than 0.5%, and to better

than 0.1% on scales larger than $15h^{-1}$Mpc (for the same correlation length $r = 3R$).

As for a classical Alcock and Paczynski test (Alcock & Paczynski, 1979), the equivalence expressed by eq. (4.5) holds true if and only if the LHS and RHS are both estimated in the correct cosmology. This is the principle that I exploit to constrain the value of key cosmological parameters. In this respect, we note that the RHS of eq. (4.5) is directly sensitive to the shape (n_s) of the primordial power spectrum of density fluctuations, as well as to the parameters determining the size of the horizon scale at matter-radiation equality (the present-day extrapolated value of the matter ($\Omega_m h^2$) and baryon ($\Omega_b h^2$) density parameters. The possible contribution of massive neutrinos is neglected in this analysis.) By contrast, the LHS term of eq. (4.5) probes the structure of the comoving distance-redshift relation, and it is in principle sensitive to the amount of matter Ω_m and, to Ω_X and w which characterize a wide range of dark energy models. Note that as long as we compute distances in units of h^{-1}Mpc ($h = H_0/100$), the LHS is effectively independent from the value of the hubble parameter H_0.

An additional merit of eq. (4.5) is the neat distinction in the physics that is brought about by its two terms. The LHS is a geometric probe since it is fully specified once the homogeneous expansion rate history of the universe is known. The nature of the RHS, instead, is fundamentally dynamical, since eq. (4.4) can be computed only when a specific gravitational theory is provided. As a result, if the expansion rate of the universe is known from independent evidences, one can directly test whether the assumed gravitational paradigm correctly describe the formation of correlated structures in the universe. Such a strategy, being independent from the specific modeling of redshift space distortions, nicely complements other traditional approaches exploiting this last effect.

I estimate $\sigma^2_{g,R}$ and $\xi_{g,R}(r)$ and correct measurements for shot noise effects according to the method detailed in chapter 2 of this manuscript. Here, I describe how we evaluate $P(\mathbf{p}|\eta_{g,R})$, the likelihood of the unknown set of parameters $\mathbf{p} = (\Omega_m, \Omega_X, w, H_0, \Omega_b h^2, n_s)$, given the actual value of the observable $\eta_{g,R}$. Note that the analysis does not require the specification of further model parameters other than those quoted above.

The posterior likelihood $P(\mathbf{p}|\eta_{g,R})$ is

$$P(\mathbf{p}|\eta_{g,R}) = \frac{L(\eta_{g,R}|\mathbf{p})\pi(\mathbf{p})}{p(\eta_{g,R})}$$

and its value can be estimated once the likelihood of the data L as well as any eventual *a-priori* information about the probability distribution function of the parameters $\mathbf{p}$ are known. To proceed, I further assume that the random variable $\eta_{g,R}$ is distributed as a Gaussian. Data confirms that this is a fair hypothesis, as I show in figure (4.5). Notwithstanding, the likelihood of the estimator is recovered from mock catalogs characterized by a smoothly decreasing density profile (see right panel of figure 4.7). I also tested this assumption by using 40 mock catalogs with a density profile that mimics that of the SDSS DR7 sample. Results are presented in figure (4.6) and confirm the hypothesis that

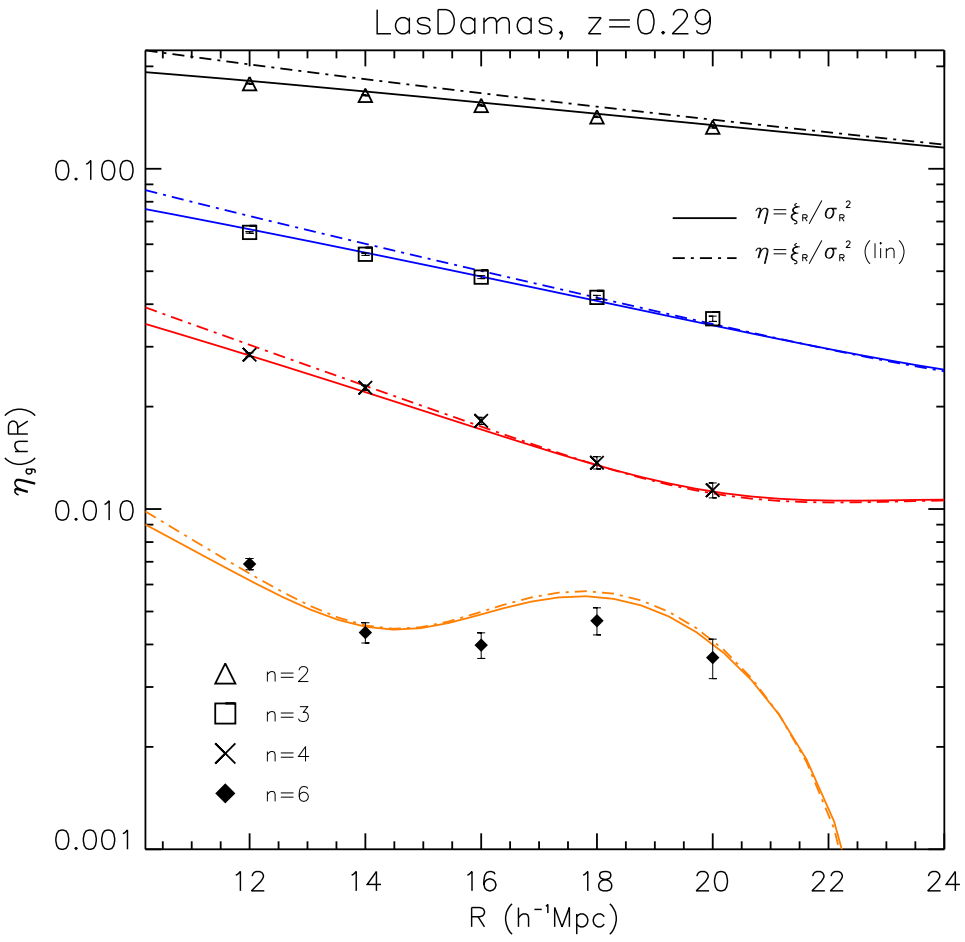

Figure 4.1: Solid lines show the non-linear theoretical prediction for the clustering ration $\eta_R(nR)$. Predictions are obtained using the non-linear prescription of Smith et al. (2003) (conventionally called *halofit*), whereas dot dashed lines display the linear theoretical prediction obtained according to equation (4.4). The different colors correspond the chosen value of the ratio between the correlation length r and the smoothing radius R ($n = r/R$): black, blue, red and orange correspond to $n = (2, 3, 4, 6)$ respectively. The galaxy clustering ratio $\eta_{g,R}(nR)$ extracted from the LasDamas simulations are displayed using triangles, squares, X and filled diamonds (for $n = (2, 3, 4, 6)$ respectively).

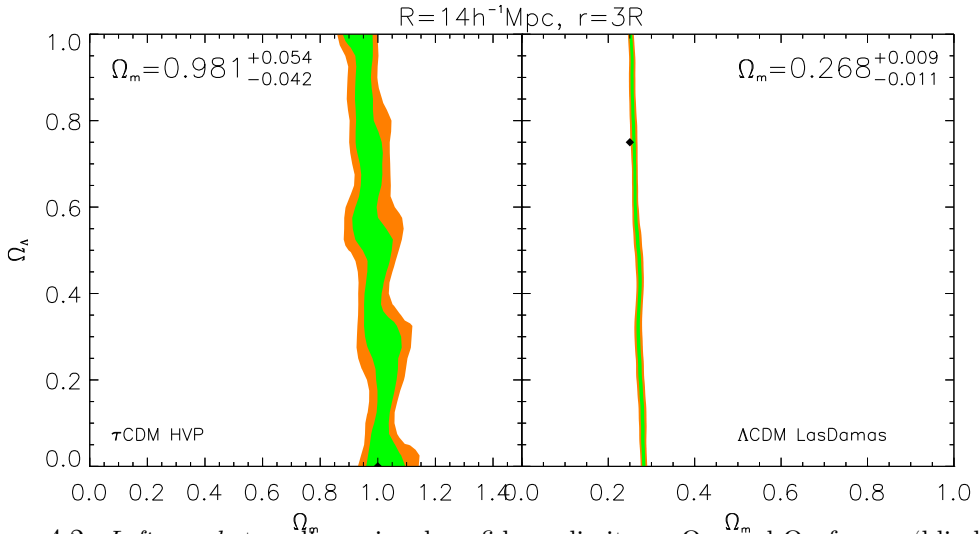

Figure 4.2: *Left panel:* two-dimensional confidence limits on Ω_m and Ω_Λ from a 'blind' analysis of the τCDM HVP simulation Jenkins et al. (2001). This synthetic catalog of clusters of galaxies covers one octant of sky, extends over the redshift interval $0.1 < z < 0.43$, is comprised of $1,000,000$ massive halos with an average space density of $6 \times 10^{-4} h^3 \mathrm{M}pc^{-3}$, and the input cosmological parameters are $\mathbf{p}^i = (1,\ 0,\ -1,\ 21,\ 0,\ 1)$. Contours are plotted for $\mathcal{L}/\mathcal{L}_{min} < 2.3, 6.17$ corresponding to 68 and 95 per cent confidence intervals for a multivariate Gaussian distribution with 2 degrees of freedom. The relevant one-dimensional marginalized constraint is shown in the inset. The clustering ratio $\eta_{g,R}(r, \mathbf{p})$ is estimated using $R = 14h^{-1}\mathrm{M}pc$ and $r = 3R$. For the best fitting cosmology we measure $\eta_{g,R}(r, \mathbf{p}^{best}) = 0.0294 \pm 0.0017$ where the error is evaluated via 30 jackknife resampling of the data, excluding each time a sky area of 36×19.5 deg^2. Dirac delta priors are taken of $\Omega_b h^2$, H_0 and n_s, that are centered on the simulated values. *Right panel:* same as before, but now contours represent the joint likelihood analysis of 40 independent ΛCDM mock catalogs simulating the SDSS LRG sample. Each of these simulations covers 120×45 deg^2, extends over the redshift interval $0.16 < z < 0.43$, is comprised of ~ 60000 galaxies with an average space density of $\sim 8.9 \times 10^{-5} h^3 \mathrm{M}pc^{-3}$ and the input cosmological parameters are $\mathbf{p}^i = (0.25,\ 0.75,\ -1,\ 70,\ 0.0196,\ 1)$. For the best fitting cosmology we find $\eta_{g,R}(r, \mathbf{p}^{best}) = 0.0573 \pm 0.0010$, where the error, computed as the standard deviation of the mean of 40 measurements, includes the contribution from cosmic variance.

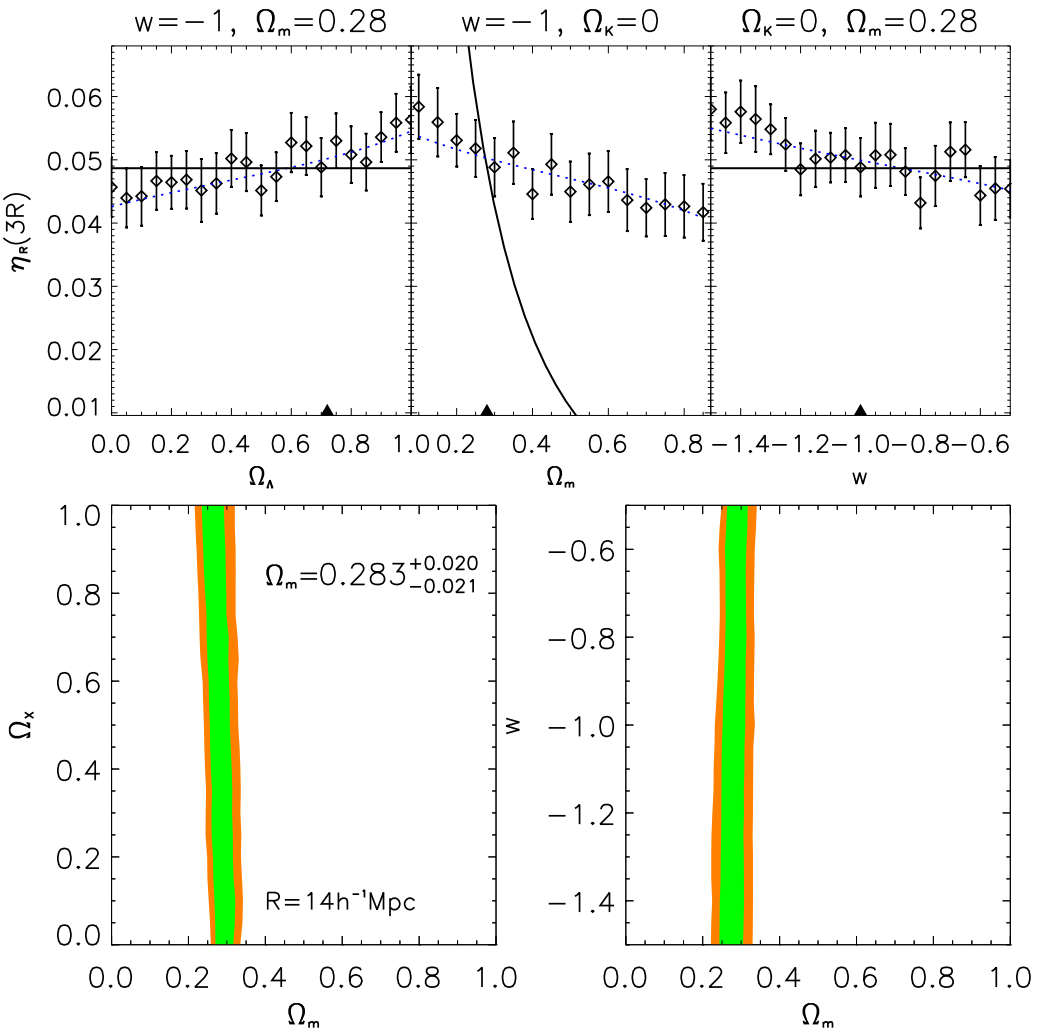

Figure 4.3: *Upper panels:* The galaxy clustering ratio $\eta_{g,R}(r)$ is plotted as a function of Ω_Λ (left), Ω_m (center) and w (right) together with its aproximation (4.7) (dotted blue line) and it is compared to the predicted scaling of the mass clustering ratio $\eta_R(r)$ (solid black line). The *true* cosmological model is the one for which both the values coincides, i.e. the LHS and RHS terms of eq. (4.5) are identical. The observable is computed using the SDSS DR7 sample and assuming $R = 14h^{-1}$Mpc and $r = 3R$. We find $\eta_{g,R}(r, \mathbf{p}^{best}) = 0.0452$, and its associated error $\sigma_\eta = 0.0035$ is computed from 30 jackknife resampling of the data, excluding, each time, a sky area of 12×14 deg^2. This error is in good agreement ($< 10\%$ difference) with the standard deviation displayed by the 40 SDSS-like simulations LasDamas. *Lower panels:* two-dimensional marginalized constraints on a curved XCDM model in which both Ω_X and w are allowed to vary. Gaussian priors are taken of $\Omega_b h^2 = 0.0213 \pm 0.0010$, of $H_0 = 73.8 \pm 2.4$ and of $n_s = 0.96 \pm 0.014$ from BBN Pettini et al. (2008), HST Riess et al. (2011) and WMAP7 Larson et al. (2011) determinations respectively. Contours are plotted for $\mathcal{L}/\mathcal{L}_{min} < 2.3, 6.17$.

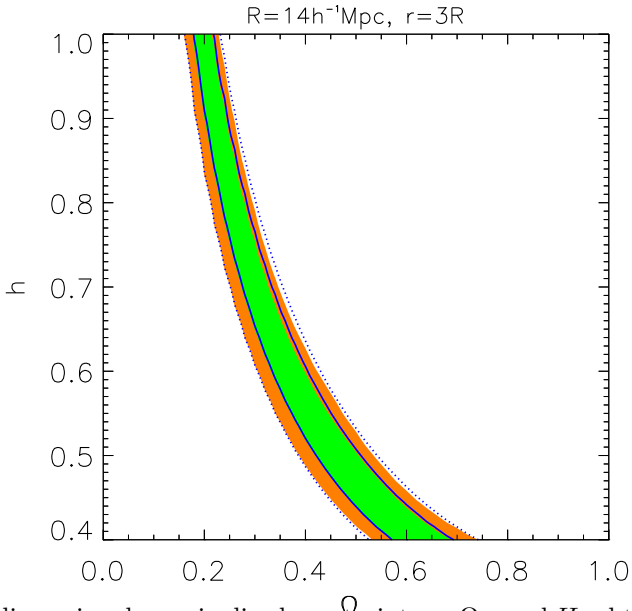

Figure 4.4: Two-dimensional marginalized constraints on Ω_m and H_0 obtained by fitting data with a curved XCDM model. Contours are plotted for $\mathcal{L}/\mathcal{L}_{min} < 2.3, 6.17$ respectively green and orange area whereas solid and dotted line corresponds to same contours obtained using aproximation (4.7). Gaussian priors are taken of $\Omega_b h^2 = 0.0213 \pm 0.0010$ and $n_s = 0.96 \pm 0.014$ from BBN and WMAP determinations respectively.

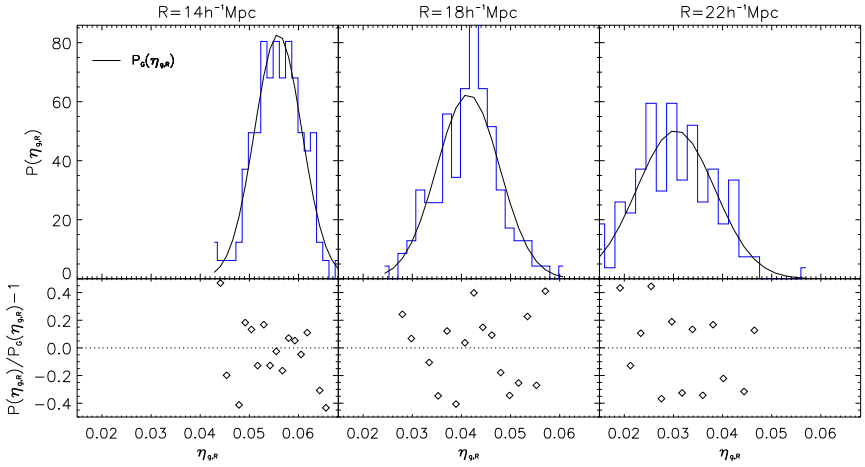

Figure 4.5: *Upper panels:* The blue histogram shows the distribution of the clustering ratio $\eta_{g,R}$ obtained from 160 mock catalogs simulating the SDSS LRG sample. Each panel presents results obtained by using a sdifferent moothing radius R (whose value is shown in the inset), whereas the correlation length is $r = 3R$. The solid line displays the best fitting Gaussian model. *Lower panels:* I show the relative deviation between the Gaussian model and the actual distribution of the data.

the measurements of the clustering ratio are distributed according to a Gaussian law, even when a realistic radial distribution of galaxies is simulated. For the sake of completeness, I show in figure (4.6) that also the variance and the correlator of order $(1,1)$ of the smoothed galaxy density have a Gaussian PDF.

Since the observable $\eta_{g,R}$ is with excellent approximation Gaussian, the most likely set ($\mathbf{p}^{best}$) is the one that minimizes the logarithmic posterior $\mathcal{L} = -\log P$

$$\mathcal{L} = \sum^{N} \log \sigma_{\eta,i} + \frac{\chi^2}{2} - \log \pi + B, \tag{4.6}$$

where $\chi^2 = \sum^N \sigma_{\eta,i}^{-2}(\eta_{g,R} - \eta_R)^2$, where N is the number of estimates of $\eta_{g,R}$, $\sigma_{\eta,i}$ is the error on the observable, and where B is a normalization constant that can be fixed by requiring $\int P d\mathbf{p} = 1$. Since eq. (4.5) is free from look-back time effects, the analysis does not require slicing the sample in arbitrary redshift bins; only one estimation ($N = 1$) of eq. (4.1) is needed across the whole sample volume.

For each comoving distance-redshift model we have recalculated the observable η_g. As a consequence, the posterior $\mathcal{L}$ does not vary smoothly between different models because the number of galaxies counted in any given cell varies from model to model. However, since the computation of the observable $\eta_{g,R}$ takes a limited amount of time, shot-noise is the price we have decided to pay in order to obtain an unbiased likelihood surface. Moreover, we have chosen not to combine constraints from our measurements with the full WMAP likelihood.

4.2 Blind Analysis of numerical simulations

Demonstration of my method under realistic operating conditions is an essential step on the road to real data application. I assess its performances via a 'blind' analysis of mock catalogs that are characterized by widely different sets of expansion rate and power spectrum parameters, mass tracers and radial selection functions. These are the τCDM Hubble Volume Project (HVP) simulation (Jenkins et al., 2001) and the LasDamas simulation (McBride et al., 2009). Clusters mock data (HVP) allow me to check for the presence of insidious algorithmic biases that could arise when training a method to 'recognize' only a fiducial ΛCDM cosmology via a single class of biased tracer of the matter clustering pattern, namely galaxies.

The HVP simulation contains 1 billion dark matter particles with mass $2 \times 10^{12} h^{-1} M_{\odot}$ evolved in a universe containing only dark matter ($\Omega_m = 1$ and $\Omega_k = 0$). The simulation starts at redshift 29 and is characterized by a pure dark matter power spectrum with shape parameter $\Gamma = \Omega_m h = 0.21$ even if the whole simulation is run by choosing $h = 0.5$. By this choice, the τCDM model of HVP simulation mimics the power spectrum of a ΛCDM model without baryons and with reduced matter density parameter $\Omega_m = 0.3$ and

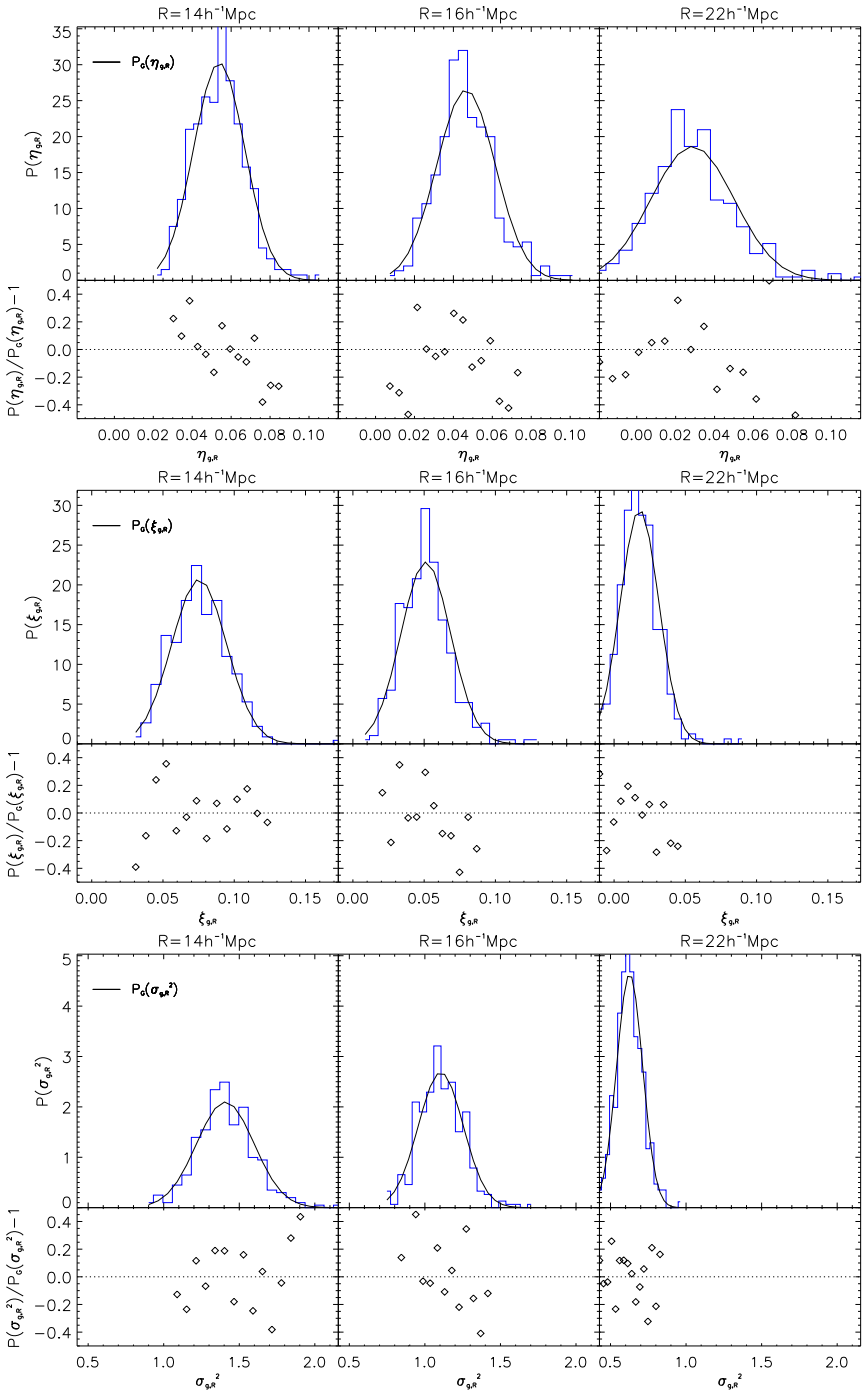

Figure 4.6: Same as in figure (4.5), but for measurements extracted from 40 LasDamas mock catalogs with the same radial distribution of the LRGs in the SDSS DR7 sample. Besides showing the PDF of the clustering ration (upper panels) , I also show the distribution of the measurements of the correlation function (central panels) and of the variance (lower panels) of the density field smoothed on a scale R (shown in the insets).

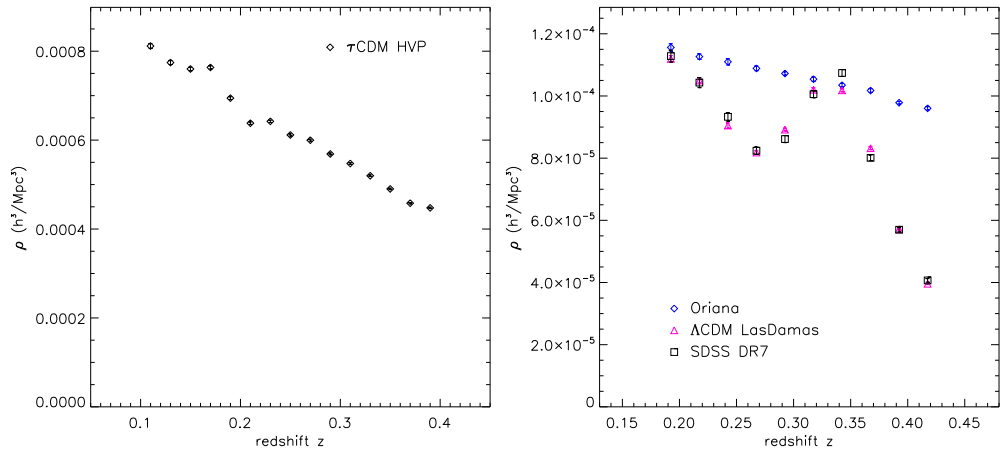

Figure 4.7: Radial density profile of simulations and data used for the clustering ratio test. The left panel shows the redshift scaling of the spatial density of halos in the τCDM HVP light-cone. The right panel shows the density profile of the original Oriana light-cone extracted from the LasDamas simulations (blue diamonds), of the LRG sample in the SDSS DR7 data (black squares) and of the LasDamas subsample that has the same radial density distribution of real data (magenta triangles).

a reduced Hubble constant $h = 0.7$. This tuning of the shape parameter was somewhat physically motivated by assuming the presence of a τ neutrino in the early universe (Bond & Efstathiou, 1991; Efstathiou, Bond & White, 1992). The density profile of the τCDM HVP simulation is given in right panel of figure (4.7).

The left panel of Fig. 4.2 shows that the input value $\Omega_m = 1$ is statistically retrieved and that, as for other clustering probes such as for example the baryon acoustic oscillation technique (Percival et al., 2010; Gaztañaga & Cabre, 2009), the method is fundamentally insensitive to the abundance and nature of dark energy. By a joint analysis of 40 ΛCDM mock catalogs that incorporate all the observing selections of the SDSS luminous red galaxies (LRG) survey, i.e. the Oriana output of the LasDamas simulations, I have verified that the specific SDSS observing biases do not spoil our cosmological inferences. Figure (4.7) compares the radial density profile of the Oriana light cone with the one of the SDSS DR7. One can notice that the density profile of Oriana is quite different from the one observed in SDSS. In order to reproduce the SDSS density profile, we removed randomly objects from the original Oriana light cones to obtain the 40 mock (referred to as ΛCDM LasDamas), to which we applied the clustering ratio test.

The encouraging outcome of this analysis is presented in Fig. 4.2, and shows that data in a volume 40 times larger than that probed by the SDSS LRG sample (still only $\sim 1/4$ of the volume that will be explored by future surveys such as EUCLID (Laureijs et al., 2011)) can constrain Ω_m precisely ($\sim 4\%$) and accurately (the input value is recovered).

4.3 The clustering ratio test applied to SDSS DR7 sample. Cosmological constraints

Turning to real data, the geometry of the SDSS DR7 subsample Abazajian et al. (2009) is dictated by the need of sampling the galaxy distribution with cells of radius R as well as of correlating cell counts on scales r. The inferior limit on R is set by a theoretical constraint. Although not mandatory, the analysis is performed in the linear domain, so to avoid any phenomenological description of the matter power spectrum. The choice of R, is additionally conditioned by the practical requirement of minimizing the shot noise contribution in each cell, i.e. $R > (4\pi\rho/3)^{-1/3}$. In this work I adopt the scale $R = 14h^{-1}$Mpc. The inferior limit $r = 3R$ in the correlation scale, instead, is set to guarantee optimal accuracy in the approximation (4.5). At the opposite end, the larger the values of R and r, and the less informative are the inferred cosmological predictions. The sample that complies with these constraints covers the redshift interval $0.15 < z < 0.43$, a contiguous sky area of 120×45 deg^2, is comprised of $62,652$ LRG galaxies, and has a mean density of $9.2 \times 10^{-5} h^3 \mathrm{M}pc^{-3}$.

Without fixing neither the curvature of the universe (flat prior $0 < |\Omega_k| < 1$) nor the quality of the dark energy component (flat prior $-1/2 < w < -3/2$), but taking (strong) Gaussian priors of $\Omega_b h^2$, H_0 and n_s from big bang nucleosynthesis (BBN, Pettini et al., 2008), Hubble Space Telescope (HST, Riess et al., 2011),and Wilkinson microwave anisotropy probe (WMAP Larson et al., 2011) respectively, data constrain the overall matter abundance with a precision of 7% (see Fig. 4.3, for the sake of completeness, by choosing a smoothing scale of $18/22h^{-1}$Mpc I obtain $0.270^{+0.037}_{-0.027}/0.255 \pm 0.038$.) This figure improves by a factor of two the estimate ($\Omega_m = 0.259 \pm 0.039$ Eisenstein et al., 2005) obtained by analyzing, with the same priors, the large scale clustering (LSC) of the SDSS DR5 sample (which contains 25% less galaxies than DR7). It also improves by $\sim 20\%$ the constraint ($\Omega_m = 0.24^{+0.025}_{-0.024}$) obtained by combining LSC results from the DR7 sample with the full likelihood of the WMAP5 data (Dunkley et al., 2009) and a strong HST prior (Riess et al., 2009). The relative precision achieved with my method compares with that ($\Omega_M = 0.294 \pm 0.023$) obtained by Sanchez et al. (2012) by combining the LSC likelihood from the dr9 sample (containing 5 times more LRG than the DR7 catalog) with the full likelihhod of CMB data (WMAP7, Larson et al., 2011). Interestingly, the best fitting value of Ω_m still remains within the quoted 68% c.l. even when we relax some of the strong priors. With a (weak) flat prior on n_s (in the interval $[0.9, 1.1]$) we obtain $\Omega_m = 0.271^{+0.030}_{-0.031}$, while if I additionaly weaken the prior on $\Omega_b h^2$ (flat in the interval $[0, 0.03]$) we obtain $\Omega_m = 0.255 \pm 0.040$. Although the best fitting value of Ω_m is weakly sensitive to changes in $\Omega_b h^2$ and n_s, Fig. 4.4 shows that Ω_m degenerates with H_0 when the HST prior is removed. If I relax also the strong prior on H_0 (by assuming a flat prior in the interval $[40, 100]$km/s/Mpc), we find $\Omega_m = 0.22^{+0.14}_{-0.04}$. The stability of the best fitting central value Ω_m is of even more interest especially if contrasted to CMB results showing

that it is the combination $\Omega_m h^2$ which is mostly insensitive to the strong prior on the curvature of the universe.

The cosmological information extracted from a cosmic probe is usually efficiently encapsulated in the covariance matrix. Among many advantages, this formalism allows to combine results obtained from a specific method with those obtained from independent probes. The matrix formalism, however, provides us with likelihood contours that poorly reproduce the likelihood contours that I obtained by applying the η test to SDSS data. This is due to the strong, 'non-gaussian' degeneracy between the matter density parameter Ω_m and the reduced Hubble constant h (see figure 4.4). I have therefore devised a different approach in order to ease the combination of my results with those obtained using other cosmological probes.

I simply worked out an empirical formula which returns the measured value of the $\eta_{g,R}$ in any given cosmology. The clustering ratio $\eta_{g,R}(r)$ reconstructed from the SDSS data on a scale $R = 14$ and for $r = 3R$ can be effectively approximated via the fitting formula

$$\eta^{fit}_{g,14}(r) = 0.00712 - 0.0788x + 0.0981x^2 - 0.0538x^3, \tag{4.7}$$

where $x = 1 - \mathrm{f}_\Omega/\mathrm{f}_o$ and the functions f_Ω and f_o are given by

$$\begin{aligned} \mathrm{f}_\Omega = \;& 3w\Omega_X - \Omega_k + \\ & \left\{ w\Omega_X(14+3w) + \tfrac{\Omega_k^2}{2} - 3w\Omega_X\Omega_k - \tfrac{11}{3}\Omega_k \right\} \bar{z} + \\ & \left\{ w\Omega_X[1 - \tfrac{9}{2}(1+w)w] - \tfrac{5}{2}\Omega_k + \tfrac{\Omega_k^2}{2} \right\} \tfrac{\bar{z}^2}{2} \\ \mathrm{f}_o = \;& w_o \left\{ 3 + (14+3w_o)\bar{z} + [1 - \tfrac{9}{2}(1+w_o)w_o \tfrac{\bar{z}^2}{2}] \right\} \end{aligned} \quad , \tag{4.8}$$

where $w_o = -3/2$, $\bar{z} = 0.29$ and $\Omega_k = 1 - \Omega_m - \Omega_X$.

Figure (4.4) shows how well the "non-Gaussian" degeneracy between h and Ω_m is reproduced by inserting this fitting formula in the equation 4.6 and by assuming $\sigma_\eta = 0.004$ in the likelihood analysis (that is a constant value in all cosmologies). The remarkable agreement between the likelihood surface recovered using real data $\eta_{g,R}$ and using the fitting formula $\eta^{fit}_{g,R}$ is also displayed in the upper panel of figure (4.3). However, assuming a constant, cosmology-independent value for σ_η is a somewhat rough approximation which causes a slight over-estimation of the area enclosed by the likelihood contours.

The advantage of probing cosmology via the ratio of second-order, one- and two-point moments of the δ_R distribution is that only a minimum amount of cosmological hypotheses, and no astrophysical nuisance parameters at all, condition the results. Of great interest is the flexibility of the method, which can still be improved along several directions. Owing to its scale-free nature, the precision of the technique could be further improved by adopting a non-linear power spectrum in eqs. (4.2) and (4.3), and by smoothing the over-density field on an even smaller scales R than that adopted throughout this analysis. The only caveat being that mass fluctuations δ_R in configuration and real spaces be simply proportional.

Alternatively, the method could be applied also to large photometric redshift surveys. In this case, predictions of eq. (4.4) must be statistically corrected to account for the line-of-sight distortions introduced when estimating low-resolution distances from photometry. Finally, this probe enlarges the arsenal of methods with which the next generation of redshift surveys such as BigBOSS (Schlegel et al., 2011) and EUCLID (Laureijs et al., 2011) will hunt for new physics by challenging all sectors of the cosmological model.

4.4 The clustering ratio test applied to the VIPERS sample

In this section I present some preliminary results obtained by implementing the clustering ratio test to the VImos Public Extra galactic Redshift Survey (VIPERS). I do not limit myself to present how I applied the test to a deep sample of galaxies. I also digress on how I corrected for the various observational selection effects that affect the VIPERS sample, and in particular for the uneven angular sampling of this survey. The correction method that I will present, in fact, can be implemented in various other applications which involve the cell-count analysis. I begin with a short overview of VIPERS, that allows me to discuss the main difficulties inherent to the application of the cell-counts in this redshift survey and then, I will detail my strategy to de-bias the cell-counts statistics.

4.4.1 Overview of VIPERS

VIPERS is a deep (up to redshift 1.5) spectroscopic survey. Figure (4.9) displays how the redshift distribution and the angular aperture compare with the SDSS survey. Its angular aperture on the sky is shown in figure (4.8), which shows that the surveys is comprised of two disjoint fields of view selected in the W1 and W4 fields of the Canada France Hawaii Telescope Legacy Survey (CFHTLS). The total area covered by VIPERS is about 24 square degrees. Spectra of VIPERS galaxies are the result of nearly 440.5 hours of observation at the very large telescope (VLT).

The location of the two fields of view of VIPERS has been chosen in order to take advantage of the large amount of photometric data already collected by the CFHTLS. The spectroscopic targets of VIPERS have been, indeed, preliminarily selected according to photometric criteria. On one hand, objects photometrically identified as stars are immediately removed, and on the other hand galaxies are targeted according to a color-color selection which minimizes the chance of observing a galaxy at a lower redshift then 0.4. By adopting this strategy, the efficiency of the survey is maximized. In fact, it is not necessary to waste time observing low redshift galaxies, of which the SDSS survey already provide a comprehensive census (see figure 4.9).

Spectra are collected in a very efficient way. With a single telescope pointing, one can target nearly 300 galaxies using the four VIMOS quadrants ($Q1$ to $Q4$). To obtain spectra of these objects, whose magnitude is $I_{AB} < 22.5$, an exposition time of $1h$ is sufficient. A

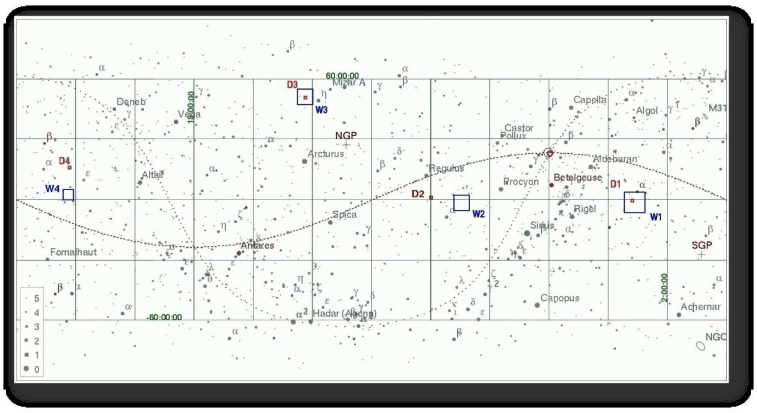

Figure 4.8: Position on the sky of the two VIPERS fields W1 and W4.

graphical illustration of one telescope pointing is displayed in figure (4.10), where one can notice that the four quadrants do not cover a contiguous area. As a consequence some gaps remain (all targets are missing in the gap region). Moreover, the position of the slits (which diffract the light coming from the targets) is optimized by the SPOC (Slit Positioning Optimization Code, see Bottini et al., 2005) algorithm which maximizes the number of observed objects given the position of potential targets. Non negligible and insidious bias arise from such systematic selection effects (gaps + SPOC).

It is therefore of paramount importance to reconstruct the 'true' probability (P_N) of finding N galaxies in random spheres of an arbitrary radius R starting from the observed distribution which is affected by all the VIPERS systematics. Note that correcting the probability distribution is equivalent to correct the complete hierarchy of one-point moments. In figure (4.11) I display the two fields W1 and W4 of VIPERS, they clearly exhibit rectangular patterns due to gap effects, namely the masked region.

We have seen (chapter 1) that a discrete distribution results from the sampling of a continuous stochastic field. Under very general conditions the sampling process is well described in terms of Poissonian selection. However, if the sampling does not satisfy the local Poisson process approximation, the way of correcting the observed moments for shot noise change. Thus it is necessary to investigate if the VIPERS target selection strategy can be described as Poisson sampling.

4.4.2 Mask effects on galaxy sampling

If the regular pattern with which slits can be disposed on a Vimos mask only modifies the conditional sampling probability $P[N|\Lambda]$ in equation (2.6), then the first step is to model the new sampling function. The modeling of the sampling function, although carried out using data in the W1 and W4 fields, must be independent of the field which is sampled. Therefore I analyzed the effects of maskig using using only random distributions, more precisely 50 mock catalogs simulating a random distribution of objects in each field $W1$

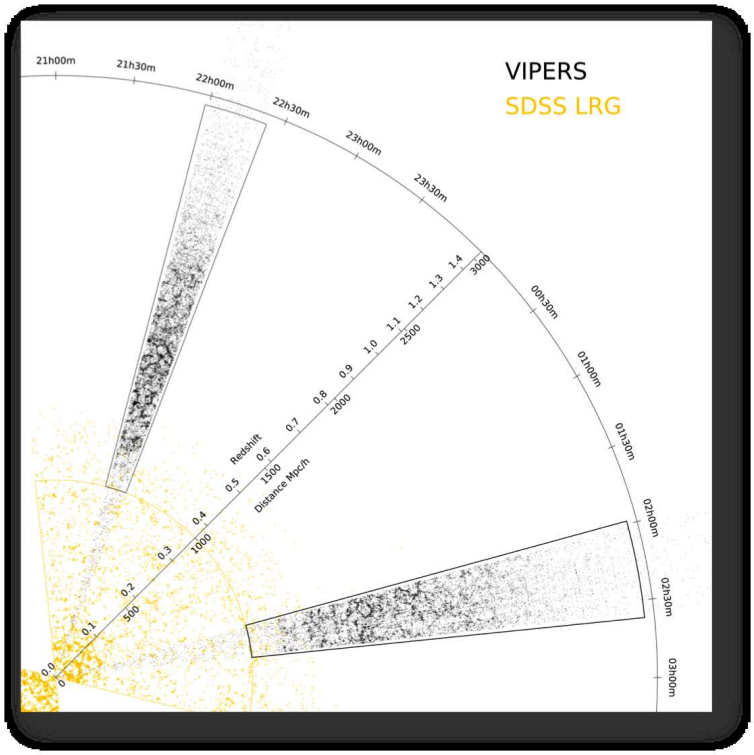

Figure 4.9: Comparison between SDSS (yellow) and VIPERS (black)

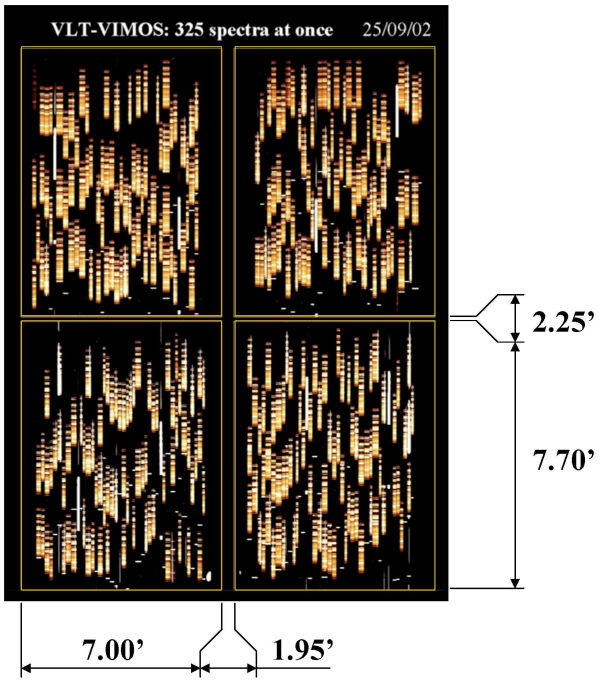

Figure 4.10: VIMOS CCD which can collect 325 spectra with a single telescope pointing. Sizes are in arc-minutes. The presence of empty gaps between the four quadrants (solid yellow rectangles) is clearly visible.

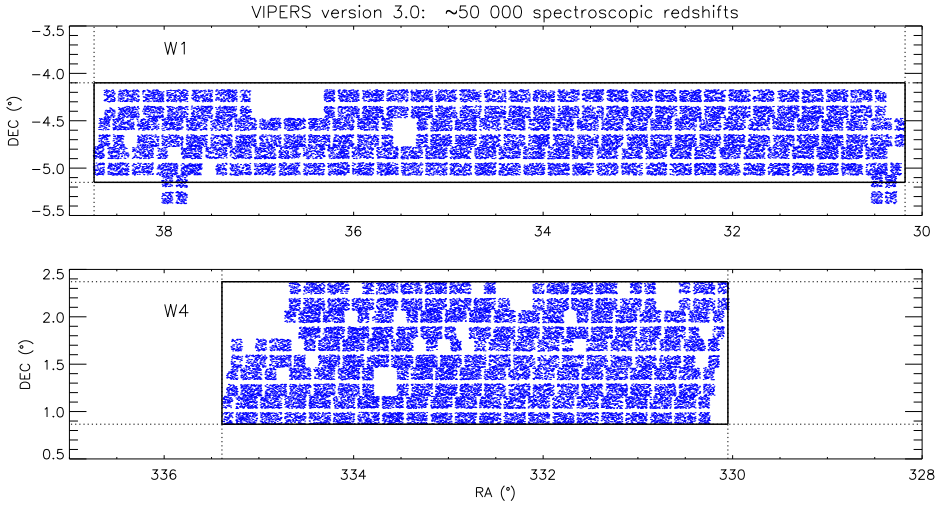

Figure 4.11: Projection on the sky of the version 3.0 of VIPERS data collected in both the W1 (upper panel) and W4 (lower panel) fields. The rectangular pattern induced by gaps is clearly visible in the angular distribution of galaxies (blue points). The solid rectangle represents the effective field of view that we have analyzed with the clustering ratio test.

and $W4$. By definition, when counts-in-cells is applied on an uniform distribution of points it is possible to show that in the limiting case when the size of the survey is much larger then the size of a cell, the predicted counting probability follows a Poisson law

$$P_N = \frac{\bar{N}^N}{N!} e^{-\bar{N}},$$

where $\bar{N}$ is the expected mean number of points in a randomly chosen cell. When no mask is applied on the random mock catalogs (i.e. the local Poisson process can be applied) it is possible to show that the underlying stochastic field associated to an uniform distribution is a Dirac distribution shifted by the quantity $\bar{\Lambda}$ (for a Poisson sampling $\bar{\Lambda} \equiv \bar{N}$)

$$\frac{\bar{N}^N}{N!} e^{-\bar{N}} = \int_0^\infty \frac{\Lambda^N}{N!} e^{-\Lambda} \delta^D(\Lambda - \bar{\Lambda}) d\Lambda.$$

Then, after applying the VIPERS masks on these random distributions, i.e. implementing the VIPERS sampling function, the above expression takes the form

$$P_N = \int_0^\infty P_{mask}[N|\Lambda] \delta^D(\Lambda - \bar{\Lambda}) d\Lambda,$$

which obviously reduces to

$$P_{mask}[N|\bar{\Lambda}] = P_N, \tag{4.9}$$

where P_N is the counting probability of the random distribution after applying masks. Equation (4.9) shows that a random distribution, offers the possibility to measure directly

the sampling function of the masks. Figure (4.12) shows the effect of masks on sampling for both $W1$ and $W4$ fields. It is clear that the two fields are not characterized by the same sampling. As a matter of fact in the $W1$ field we find an excess of low counts that is an observed P_N which is strongly left-skewed. This would suggest the necessity of a non trivial modeling of the sampling function. However the simpler is the model the simpler is the way of finding the correction to apply on the moments. The asymmetric distribution of probability arises because of the strong impact of large gaps (that is bad quadrants, or missing quadrants). To avoid modeling such specific features, it is possible to include in the counting analysis only those spheres for which the fraction of volume included in the unmasked area is greater than a given threshold w_{th}. Figure (4.12) shows the measured sampling functions extracted from the $W1$ and $W4$ fields using only cell counts in spheres for which w_{th} is larger then a given threshold. I also show how the observed distribution of counts is modeled by two theoretical distribution laws: the Poisson law and the negative Binomial distribution. This last distribution

$$P_N^{NegBin} \equiv \frac{\bar{N}^N}{N!(r+\bar{N})^N}\left(1+\frac{\bar{N}}{r}\right)^{-r}\frac{\Gamma(N+r)}{\Gamma(r)}, \tag{4.10}$$

being characterized by 2 parameters, presents a larger number of degrees of freedom, and therefore a higher flexibility, with respect to the simple Poisson law. Note that the value of the r parameter depends on the expected variance σ_N^2 and mean number $\bar{N}$, thus

$$r \equiv \frac{\bar{N}^2}{\sigma_N^2 - \bar{N}}.$$

An additional interesting property of the negative binomial law is that when the r parameter goes to infinity the distribution becomes exactly a Poisson distribution

$$\underset{r\to\infty}{\mathrm{L}im}\ P_N^{NegBin} = P_N^{Poisson} \equiv \frac{\bar{N}^N}{N!}e^{-\bar{N}}. \tag{4.11}$$

Using a negative binomial, I can quantify deviations from a perfect Poisson sampling by means of the r parameter. From figure (4.12) I conclude that the counts probability distribution resulting from selecting cells with $w_{th} = 60\%$ is somewhat better modeled in term of a Negative Binomial.

The next step is to ensure that the value of the r parameter which was fixed using random simulations depends effectively only on the geometry of the mask and not on the properties of the random sample used in the analysis and in particular on the mean number of objects in cells $\bar{\Lambda}$. The easiest way to proceed, is to divide the spatial density of the random mock catalogs by a factor of 2 and 4 (i.e. change the value of $\bar{\Lambda}$) and explore how this change affects the value of r. The results are shown in figure (4.13) where I plot the distributions for $W1$ and $W4$. The best fitting values of the r parameter are reported

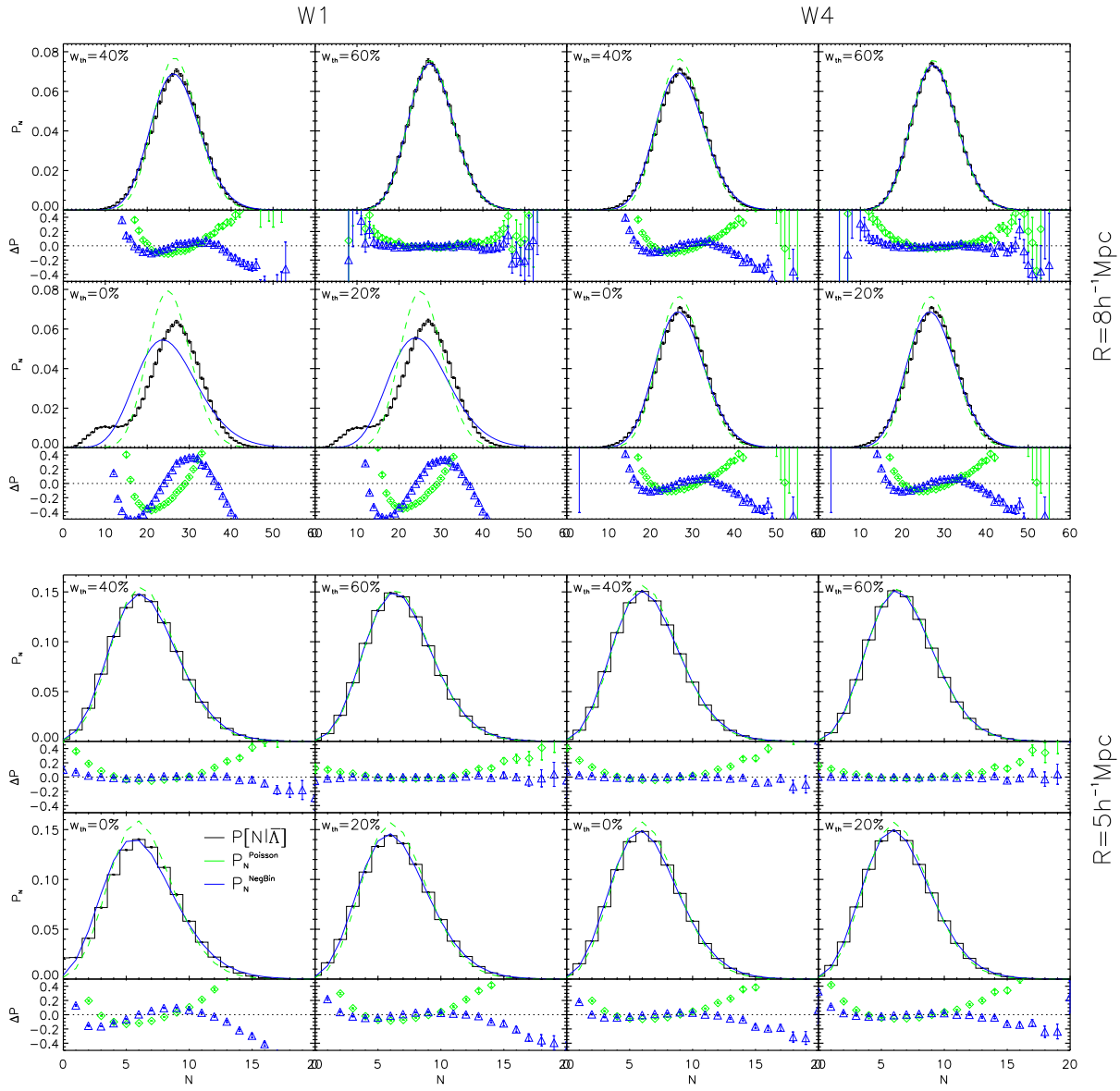

Figure 4.12: Upper panels concern $R = 8h^{-1}$Mpc while lower panels concern $R = 5h^{-1}$Mpc, on the left I describe $W1$ whereas on the right results are shown for $W4$. Each panel concerning a given field and a given radius is decomposed in 4 quadrants which show the measured sampling conditional probability $P[N|\bar{\Lambda}]$ (black histogram), the Poisson modeling (blue solid line) and the Negative Binomial modeling (green dashed line) for different values of the of the threshold w_{th} imposed on the volume fraction of the spheres to select it. Each individual quadrant is composed of two parts, the upper part shows the measured and modeled probability distributions whereas the lower part corresponds to the relative difference between the Poisson and Negative Binomial modeling with respect to the effective measurement of the conditional probability. $\Delta P \equiv P[N|\bar{\Lambda}]/P_N^{model} - 1$ is represented with green diamonds for Poisson modeling and with blue triangles for Negative Binomial. Error bars in measurements are obtained from 50 random distributions which mimic the masked $W1$ and $W4$ fields.

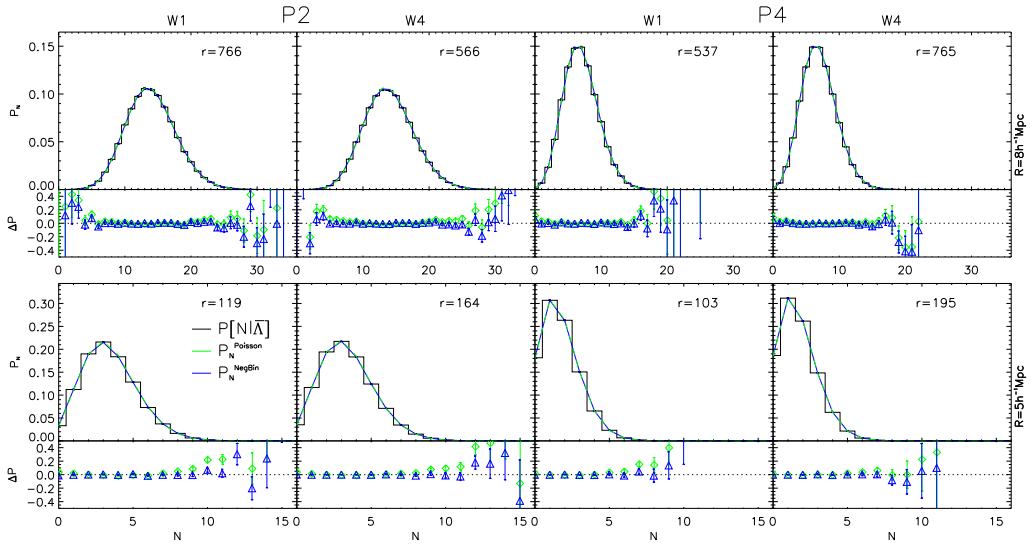

Figure 4.13: The count probability in spheres $R = 8h^{-1}$Mpc (upper panels) and $R = 5h^{-1}$Mpc (lower panels). The color and drawing conventions are the same as in figure (4.12) but all measurements have been done by removing the cells with more then 40% of their volume falling in a gap (i.e. $w_{th} = 60\%$). Left panels show measurements and modeling using data from random catalog (P2) containing half of the objects of the original catalog. The right panels show the same measurements extracted from a catalog (P4) containing a quarter of the object of the original random catalog. Note also that the best fitting parameter r are reported within each individual panel.

in each plot. The $P2$ and $P4$ panels correspond to random catalogs with 2 and 4 times less objects then the original random distributions. These plots show that the value of r depends on $\bar{\Lambda}$. Moreover, he lower the value of $\bar{\Lambda}$, the larger the value of the best fitting r. This means that decreasing the mean number of objects in the original sample (i.e. that unaffected by masks) improves the chances that the sampling function is effectively Poissonian. Since the mean number density of galaxies in the VIPERS catalogs is nearly 4 times lower than that characterizing the original random samples (that is similar to the spatial density of objects in the $P4$ catalog), it is possible to model the VIPERS sampling function in terms of the classical local Poisson process. As a result, the shot noise correction scheme for the $1-$point moments given by equation (1.75) can still be safely applied.

The remaining problem is to infer the moment of the underlying continuous field Λ knowing that the VIPERS observing strategy allows us to measure the redshift for only a fraction (nearly 40%) of the whole population of galaxies with magnitude $I_{AB} < 22.5$. Let N be the random variable characterizing counts in a (parent) population which is not affected by mask sampling or spectroscopic sampling rate and be M the random variable actually observed in a sample. The parameter $\alpha \equiv \frac{\bar{M}}{\bar{N}}$ represents the fraction of data to which the survey give us access. As I have shown in the chapter (1), the cumulant moments of the density contrast ($\langle \delta^n \rangle_c$) must be invariant under the sampling. This means that,

concerning the parent population we can write

$$\frac{\langle \Lambda^n \rangle}{\bar{\Lambda}^n} = \frac{\langle (N)_f^n \rangle}{\bar{N}^n},$$

where, by definition, $\bar{N} \equiv \bar{\Lambda}$. Concerning the sample, we have

$$\frac{\langle \Lambda^n \rangle}{\bar{\Lambda}^n} = \frac{\langle (M)_f^n \rangle}{\bar{M}^n}.$$

which, finally, leads to the formula

$$\langle \Lambda^n \rangle = \frac{\langle (M)_f^n \rangle}{\alpha^n}, \tag{4.12}$$

which allows us to estimate true moments of the continuous stochastic field Λ from the factorial moments of a sample extracted from the parent population.

4.4.3 Reconstructing the count probability P_N

Now that we know the shape of the VIPERS sampling function $P[N|\Lambda]$ as well as the moments $\langle \Lambda^n \rangle$ we must figure out how to compute $P(\Lambda)$ so that, by applying equation 2.6 (see section 1.3.6), we can recover the value of the 'true' count probability distribution P_N. There are different strategies that allow reconstructing a probability density function given its moments. An often adopted approach, consists in using the Edgeworth expansion technique. One assumes a Gaussian distribution and tunes it in order to modify its moments order by order. I tried this first possibility but results were not encouraging. The fact is that the Gaussian distribution offers a poor approximation of the distribution $P(\Lambda)$ that we want to reconstruct. In practice, since the observed P_N is similar to a geometric distribution, the underlying probability density function is roughly given by a Gamma distribution. Because of this I decided to use a Gamma expansion technique, whose virtues have been tested by Gaztañaga, Fosalba & Elizalde (2000). In what follows, however, I exploit results obtained by Mustapha & Dimitrakopoulos (2010).

As suggested by its name, the Gamma expansion method relies on a Gamma distribution ϕ_G with general expression

$$\phi_G(x) = \frac{\theta^{-k}}{\Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}, \tag{4.13}$$

where θ and k are two real parameters. Their value can be set by requiring that the mean $\bar{x}$ and the variance σ_x^2 (of the PDF that we want to reconstruct) are equal to the mean and variance of the Gamma distribution, thus $\theta = \sigma_x^2/\bar{x}$ and $k = \bar{x}^2/\sigma_x^2$. From the definition (4.13) it follows that k must be greater then 1 to allow the Gamma function Γ to be defined. When k is an integer the Gamma function is related to factorial by $\Gamma(k) = (k-1)!$. It means that the distribution must satisfy $\bar{x}^2 \geq \sigma_x^2$; if not, the parameter

k must be set to 1 and the Gamma distribution reduces to an exponential distribution.

Let $P(\Lambda)$ be the probability density that we want to reconstruct. Then the Gamma expansion at order n can be written as

$$P(\Lambda) \simeq P_n^{k-1}(\Lambda) \equiv \sum_{i=0}^{n} c_i L_i^{(k-1)}(\theta\Lambda)\phi_G(\theta\Lambda), \tag{4.14}$$

where $L_i^{(k-1)}$ are the generalized Laguerre polynomials which are defined as

$$L_n^{(k-1)}(z) \equiv \sum_{i=0}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) \frac{\Gamma(k+n)}{\Gamma(k+i)}(-1)^i \frac{z^i}{n!}.$$

Due to their remarkable orthogonal properties, they can be used to tune the moments of the Gamma functions. Indeed, each coefficient c_n in the expression (4.14) can be calculated

$$c_n = \sum_{i=0}^{n} \left(\begin{array}{c} n \\ i \end{array} \right) \frac{\Gamma(k)}{\Gamma(k+i)}(-1)^i \frac{\mu_i}{\theta^i}, \tag{4.15}$$

where by definition $\mu_i \equiv \langle \Lambda^i \rangle$. Note also that by definition $c_0 \equiv 1$ and that c_1 and c_2 are forced to be 0 when we make the mean $\bar{\Lambda}$ and variance σ_Λ^2 equal to the mean and variance of the Gamma distribution (by calculating the corresponding values of the parameters k and θ). Since, as stressed above, a k value lower then 1 can occur, in this case (and this is the case for $R = 5h^{-1}$Mpc) it is necessary to set $k = 1$, a choice that force c_2 to be different from 0.

As the probability density function $P(\Lambda)$ has now an analytical expression which depends on the moments we estimate, it is possible to apply the Poisson process to this distribution and calculate the reconstructed P_N. Using expression (2.6) we have

$$P_N = \int_0^\infty P(\Lambda) \frac{\Lambda^N}{N!} e^{-\Lambda} d\Lambda. \tag{4.16}$$

I verified this scheme to recover the correct value of the count probability P_N using three different mock catalogs simulating the VIPERS survey. These are the *reference* catalog, the *cut* catalog and the *observed* catalog. The *reference* catalogs are light cones (with peculiar velocities included) extracted from the Millennium simulation by implementing the same apparent magnitude cut of VIPERS (i.e. reject object with $I_{AB} \geq 22.5$). The *cut* samples are obtained by applying the VIPERS color selection to objects in the *parent* catalog. The *observed* sample was extracted by running the SPOC algorithm in each VIPERS quadrants in order to obtain a catalog simulating the same VIPERS target selection strategy. In order to be as realistic as possible, I simulated the spectroscopic errors estimated from VIPERS spectral measurements and applied realistic mask obscuring areas of the survey with poor photometric quality. These final (and realistic) mock catalogs are called *test3b* .

Figure (4.14) shows a comparison between the recovered P_N from a *test3b* catalog and

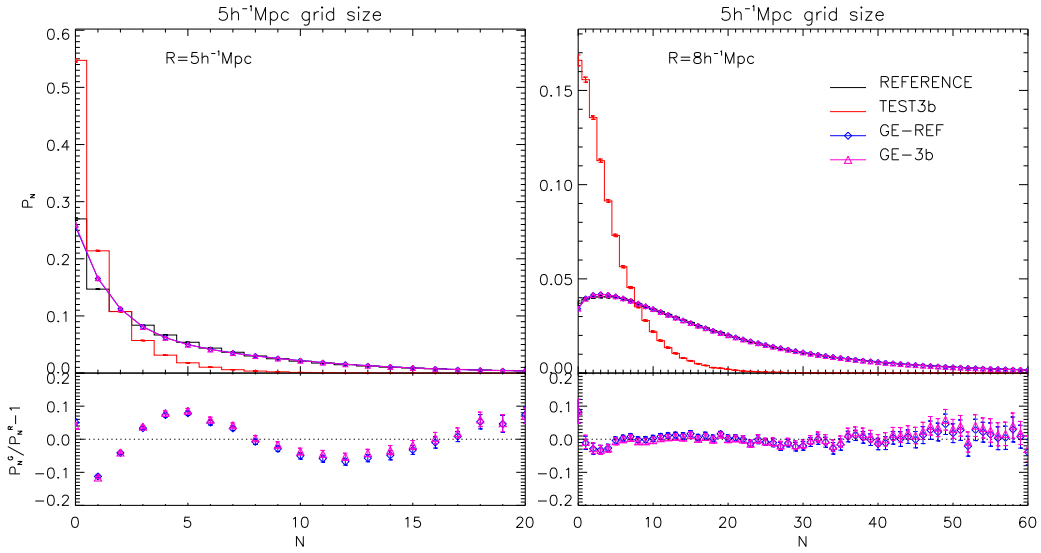

Figure 4.14: *Upper:* black histogram shows the count probability P_N of the *reference* catalog, the red histogram shows the P_N extracted from *test3b* catalog, solid blue line with diamonds shows the result of the Gamma expansion (GE) applied on the *reference* catalog and the solid magenta line with triangles shows GE applied on the *test3b* corrected catalog, for the considered radius $R = 5h^{-1}$Mpc (left) and $R = 8h^{-1}$Mpc (right). *Lower:* Relative difference between the count probabilities P_N reconstructed via the GE applied to the *reference* and to the *test3b* catalogs.

the true P_N extracted from the *reference* catalog. In order to test the accuracy of the reconstructing scheme, I also applied the Gamma expansion to the *reference* P_N. This comparison is shown in figure (4.14). The relative difference between the true P_N and the reconstructed P_N are in good enough agreement (i.e. they differ by less then 10%). In principle, however, the agreement should be bette. In fact, in this preliminary application, I neglected to take into account the radial selection function (i.e. the fact that the spatial density of galaxies decreases as a function of redshift). To test this hypothesis I created a volume limited catalogs (i.e. a sample with a constant density profile) extracted from the *reference* and *test3b* catalogs and I rerun the correction procedure. Results are presented in figure (4.15). Although error bars are larger, data show that the relative difference between the true and the reconstructed P_N show a lower dispersion around the zero value.

4.4.4 Preliminary results on VIPERS data

In this section I present preliminary results obtained by implementing the clustering ratio test with VIPERS data. First, I need to choose the parameters R and r so that the large separation limit is verified. I therefore set $n = r/R = 3$. It means that the size of the cell motifs with which I tessellate the survey volume is about $5R$ ($R+r+R$ is the total size of a motif). As the maximum aperture in declination of VIPERS data is about one degree, the clustering ratio analysis is carried out by smoothing the galaxy density field using a Top-Hat window of radius $R = 5h^{-1}$Mpc. This scale corresponds nearly to the mean inter-galaxy separation in the VIPERS sample. Since the analysis requires computing the

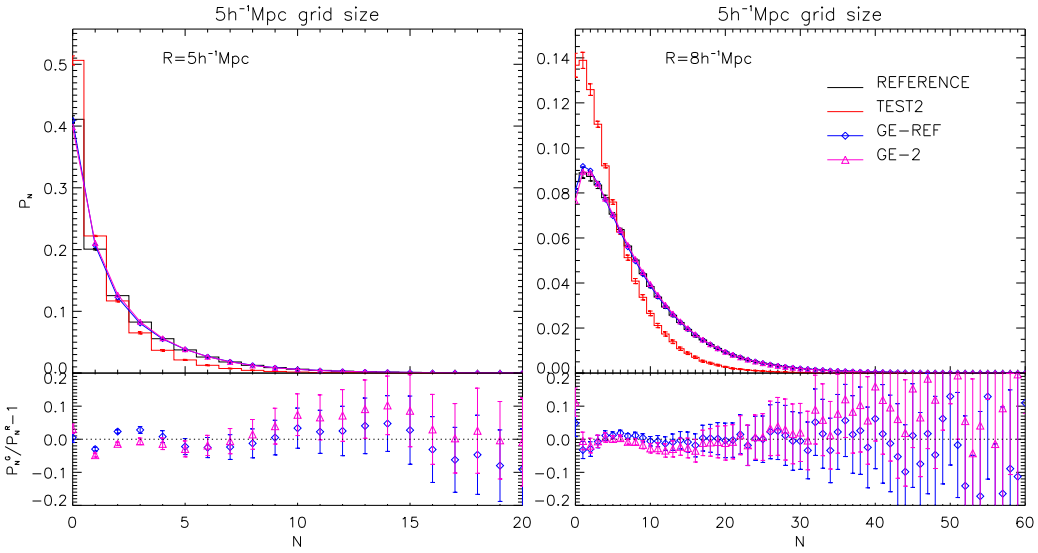

Figure 4.15: Same as in figure (4.14) but applied on volume limited samples extracted from the *reference* and *test2* catalog.

η observable in a large number of cosmological models which satisfy the conditions

$$\begin{aligned} 0 \leq \quad \Omega_m \quad \leq 1 \\ 0 \leq \quad \Omega_\Lambda \quad \leq 1 \end{aligned}$$

The size of the cell motif force us to restrict the analysis to redshifts greater than 0.65. The upper bound is fixed to $z = 1.2$ in order to minimize shot noise effects. Note also that at such small scales (i.e $R = 5h^{-1}$Mpc) the linear power spectrum does not offer a sufficiently accurate description of the matter fluctuation field. As a result, I use here the non linear prescriptions from Smith et al. (2003) in order to predict the value of the clustering ratio in each tested cosmologies.

In this preliminary analysis I consider as free fitting parameters only the matter density Ω_m and the dark energy density Ω_Λ (assuming $w = -1$). I indeed take Dirac priors of the other cosmological parameters (that is $h = 0.738$, $\Omega_b h^2 = 0.021$ and $n_s = 0.96$) even if I do not put any restriction on the spatial curvature of the universe. The encouraging outcome of this analysis are presented in figure (4.16). The left panel confirms that even at high redshift the clustering ratio is weakly sensitive to the value of the cosmological constant in both $W1$ and $W4$ fields. The one-dimensional posterior distribution of the matter density parameter, that I obtained by marginalizing the likelihood over the dark energy density parameter (Ω_Λ), is shown in the right panel of figure 4.16). By taking the contour level corresponding to $\mathcal{L} = \mathcal{L}_{min} + 1$ we find that $\Omega_m = 0.27^{+0.09}_{-0.05}$. This is to our knowledge the most precise (27%) determination of the present day value of Ω_m inferred using the galaxy distribution at redshift $z = 1$.

Even if the smoothing scale adopted to perform the test ($5h^{-1}$Mpc) is rather small in comparison to that ($R = 14h^{-1}$Mpc) used for filtering the galaxy distribution of the SDSS

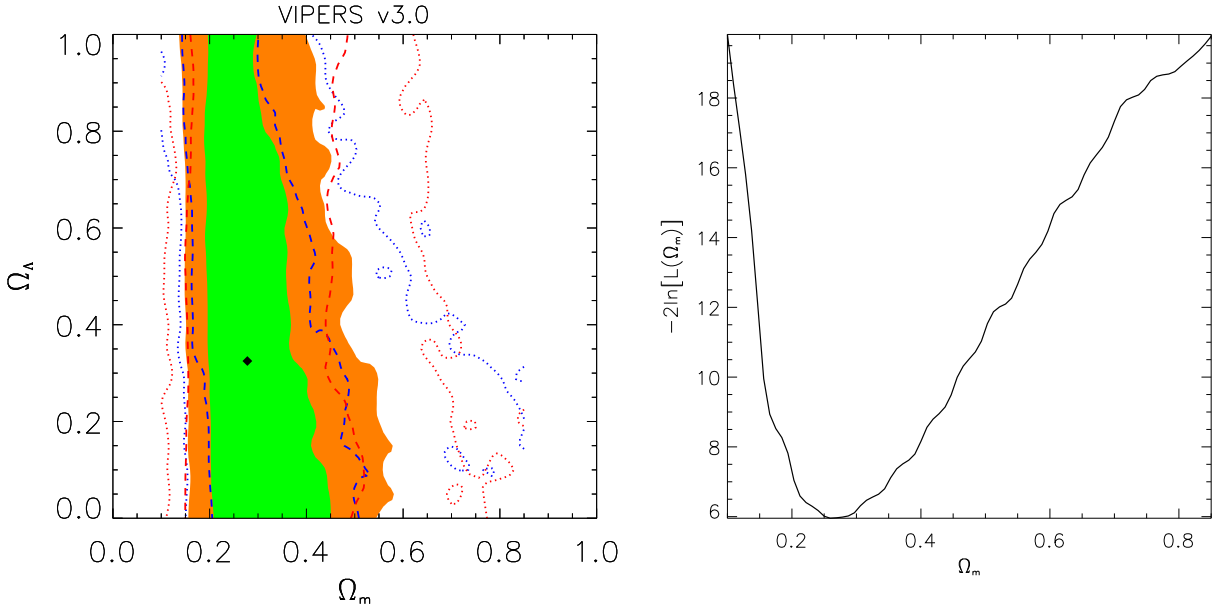

Figure 4.16: *right*: Two-dimensional constraints on a curved universe with a cosmological constant. Dirac priors are taken of $\Omega_b h^2 = 0.021$, of $H_o = 73.8$ and of $n_s = 0.96$. Contours are plotted for $\mathcal{L}/\mathcal{L}_{min} < 2.3, 6.17$ respectively dashed and dotted line for the clustering ratio test applied on $W1$ (blue) and $W4$ (red) independently. Then the green shaded are and the orange area show the $\mathcal{L}/\mathcal{L}_{min} < 2.3, 6.17$ contours levels when combining $W1$ and $W4$. The black diamond materialises the coordinates of the point which maximizes the likelihood probability. *right*: One-dimensional marginalised (on Ω_Λ) posterior distribution of the matter density parameter. The corresponding one-dimensional marginalised constraint ($\mathcal{L} = \mathcal{L}_{min} + 1$) is $\Omega_m = 0.27^{+0.09}_{-0.05}$.

DR7 survey, the resulting constraints on Ω_m inferred at these two widely different cosmic epochs are in good agreement. These results show that the clustering ratio test can be applied also in the non-linear regimes. Indeed, it will be interesting to test its robustness and accuracy on such scales using N-body simulations.

Conclusions and perspectives

The most significant contribution of my PHD work has been devising and implementing competitive cosmological probes that exploit the power of a relatively unexplored statistical quantity, that is the high order, 2−point cumulant moments of the galaxy density field. As a part of this research program I've had the opportunity of deepening my understanding of the dynamical and statical properties of matter perturbations in the weakly non-linear regime. I have acquired a solid expertise in developing routines that perform intensive data computations, and I have been exposed to the challenge of developing theoretical instruments to de-biasing data from observational systematics.

By applying the methods that I contributed to develop to available SDSS data, I have obtained interesting observational constraints on some fundamental cosmological parameters. However, the future research perspectives that will be opened by this preliminary analysis are of even more interest. As a matter of fact, from the theoretical side, work must still be done in order to fully understand robustness and limitations of the weakly non-liner perturbation theory. In particular to tackle the issue of estimating the amplitude of the reduced correlators of order (n, m) and assessing the impact of redshift distortions on these high order 2−point statistics. From the numerical side, we are confronted with the problem of speeding up the computing time required by these algorithms. Not only in order to process in a fast and efficient way high resolution matter particle simulations, but also to prepare the analysis of the large mass of redshift survey data that will soon be available. From the observational side, finally, we have the unique possibility of applying these measuring strategy to deeper and larger redshift samples, and therefore obtaining even more stringent and less degenerate constraints on relevant cosmological parameters. In this perspective, I'm already implementing the tests presented in this manuscript to VIPERS data, and I look forward to applying them using galaxies sampled by the BOSS survey, and in the near future by the BigBOSS and EUCLID surveys.

Bibliography

Abazajian, K. N. et al. 2009, ApJS, 182, 543

Acquaviva, V. & Gawiser, E. 2010 PhRvD, 82, 2001

C. Alcock and B. Paczynski, Nature **281**, 358 (1979)

Amendola, L., Palladino, E. 1999, ApJ, 514, 1

Amendola, L., Polarski, D., Tsujikawa, S., 2007, Phys. Rev. Lett. 98, 1302

Amendola L. et al., 2012, arXiv1206.1225 (2012)

Angulo R.E., Baugh C.M., Lacey C.G., 2008, MNRAS, 387, 921

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

Bagla, J. S., Yadav, J. & Seshadri, T. R. 2005, MNRAS, 390, 829

Bardeen J.M., Bond J.R., Kaiser N., Szaley A.S., 1986, ApJ, 304, 15

Baugh, C. M., Croton, D., Gaztañaga, E. et al. 2004, MNRAS, 351, L44

Bekenstein J. D., 2004, Phys. Rev. D 70, 083509

Bel J., Marinoni C., 2012, MNRAS, 424, 971

Bennett, C. L. et al. 1994, ApJ., 436, 423

Bernardeau F., 1992, ApJ, 392, 1

Bernardeau F., 1994, ApJ, 433, 1

Bernardeau F., 1994, A&A, 291, 697

Bernardeau F., 1996, A&A, 312, 11

Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rept., 367, 1

Bernardeau F., 2008,"Cosmologie, des fondements théoriques aux observations"

Bernardeau F., Brax P., 2011, JCAP, 1106, 19

de Bernardis P, et al. 2000, Nature, 404, 955

Bertacca, D., Bartolo, N., Diaferio, A., Matarrese, S., 2008, JCAP, 10, 23

Bharadwaj, S., Gupta, A. K., Seshadri, T. R. 1999, A&A, 351, 405.

Blake, C. et al., 2011, arXiv:1104.2948

Bolejko, K. & Wyithe, J. S. B. 2009, JCAP, 0902, 020

Bond J.R, Efstathiou G., 1991, Phys. Lett. B, 265, L45

Borgani, S. 1995, Phys. Rep., 251, 1

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

Bottini D. et al. 2005, PASP, 117, 996

Bouchet F.R., Juszkiewicz R., Colombi S., Pellat R., 1992, ApJ, 394, L5

Bouchet F.R., Colombi S., Hivon E., Juszkiewicz R., 1995, A&A, 296, 575

Buchert T., 1989, A&A, 223, 9

Buzzi, A., Marinoni, C., Colafrancesco, S., 2008, JCAP, 11, 001

Caldwell R.R., Dave R., Steinhardt P.J., 1998, Phys. Rev. Lett., 80, 1582

Caldwell R. R., Stebbins A., 2008, Phys. Rev. Lett., 100, 191302

Capozziello, S., Cardone, V. F., Troisi, A. 2005, Phys. Rev. D 71, 043503

Cappi, A., Benoist, C., da Costa, L. N., & Maurogordato, S. 1998 A&A, 335, 779

Catelan P., Lucchin F., Matarrese S., Moscardini L., 1995, MNRAS, 276, 39

Célérier, M.-N. 2007, New Advances in Physics, 1, 29

Chevallier M., Polarski D., 2001, Int. J. Mod. Phys., D10, 213

Clarkson, C., Bassett, B., & Lu, T. H.-C. 2008, Phys. Rev. Lett., 101, 011301

Clifton, T., Ferreira, P. G., & Land, K. 2008, Phys. Rev. Lett., 101, 131302

Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369

Crocce M., Scoccimarro R., 2006, Phys. Rev. D, 73, 63520

Davis, M. 1997 in Critical Dialogs in Cosmology, ed Turok N (Singapore: World Scientific) pp 13

Dubinski J., Kim J., Park C., Humble R., 2004, New Astronomy, 9, 111

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

Dvali, G., Gabadadze, G., Porrati, M. 2000, Phys. Lett. B 485, 208

Efstathiou G., Bond J.R., White S.D.M., 1992, MNRAS, 258

Ehlers J., 1960, Abh Mainz Akad Wiss u Lit (Math Nat Kl); reprinted in English, Gen Rel Grav 25, 1225-1266 (1993)

Einstein A., 1915, SPAW, 844

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

Eisenstein, D. J. et al. 2001, AJ , 122, 2267

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

Ellis, G. F. R. 1976, QJRAS, 16, 245

Feldman, H. A. et al. 2001, Phys, Rev. Lett, 86, 1434

Ferreira P.G., Joyce M., 1998, Phys. Rev., D58, 023503

Ferreira P.G., Joyce M., 1998, Phys. Rev. Lett., 79, 4740

Fosalba P., Gaztañaga E., 1998, MNRAS, 301, 503

Fosalba P., Gaztañaga E., 1998, MNRAS, 301, 535

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

Friedman, A. 1922 Zeitschrift fr Physik, 10, 377

Frieman J. A., Turner M. S., Huterer D., ARA&A **46**, 385 (2008)

Fry J. N., 1984, ApJ, 277, L5

Fry J. N., 1984, ApJ, 279, 499

Fry J.N., 1985, Phys. Letters, 158B, 211

Fry J.N. & Gaztañaga E., 1993, ApJ, 413, 447

Fry J. N., 1994, Phys. Rev. Lett., 73, 2

Gaite, J., Domínguez, A. & Pèrez-Mercader, J. 1999, ApJ, 522: 5

Gaztañaga E., 1994, MNRAS, 268, 913

Gaztañaga E., Frieman, J. A. 1994, ApJ, 437, L13

Gaztañaga E., Fosalba P., 1998, MNRAS, 301, 524

Gaztañaga E., Fosalba P., Elizalde E., 2000, ApJ, 539, 522

Gaztañaga E., Lobo J. A., 2001 ApJ, 548, 47

Gaztañaga E., Fosalba P., Croft R.A.C., 2002, MNRAS, 331, 13

Gaztañaga E., Scoccimarro, R. 2005, MNRAS, 361, 824

Gaztañaga, E., Norberg, P., Baugh, C. M., Croton, D. J., 2005, MNRAS, 364, 620

Gaztanaga E., Cabre A., MNRAS **399**, 1663 (2009)

Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

Giovannelli, R., & Haynes, M. P. 1991, ARA&A, 29, 499

Goodman, J. 1995, Phys. Rev. D, 52, 1821

Goroff M.H., Grinstein B., Rey S.J., Wise M.B., 1986, ApJ, 311, 6

Gregory, P. C. & Condon, J. J. 1991, Astrophys. J. Suppl. Ser., 75, 1011

Groth R.J., Peebles P.J.E., 1975, A&A, 41, 143

Guzzo L. et al., 2008, Nature, 451, 541

Guzzo, L. 1997, New Astronomy, 2, 517

Hamilton A.J.S., 1998, in D.Hamilton (ed.), The evolving Universe, Astrophysics and Space Science Library, 231, 185

Harrison E.R., 1970, Phys. Rev. D, 1, 2726

Haynes, M. P. 1996, PNAS, 93, 14216

Heath D.J., 1977, MNRAS, 179, 351

Heavens, A. F., Jimenez, R. & Maartens, R. 2011, JCAP, 9, 035.

Hivon E., Bouchet F. R., Colombi S., Juszkiewicz R., 1995, A&A, 298, 643

Hogg D.W., 1999,*Distance Measures in Cosmology*, arXiv:astro-ph/9905116v4

Hogg, D. W. 2005, ApJ, 624, 54

Hubble E., 1929, Proceding of the National Academy of Science, 15, 168

Jain B., Bertschinger E., 1994, ApJ, 431, 495

Jain B. & Zhang P., 2008 Phys. Rev. D 78, 063503

A. Jenkins et al., MNRAS, **321**, 372 (2001)

Jia, J. & Zhang, H. B. 2008, JCAP, 12, 002

Juszkiewicz R., Bouchet F.R., Colombi S., 1993, ApJ, 412, L9

Kaiser N., 1987, MNRAS, 227, 1

Kamionkowski M., Buchalter A., 1999, ApJ, 514, 7

Kendall M. G., Stuart A., The advanced theory of statistics, New York, Macmillan, (1983).

Kim J., Park C., Gott III J.R., Dubinski J., 2009, ApJ, 701, 1547

Kim J. & Park C., 2006, ApJ, 639, 600

Komatsu, E., et.al., 2011, ApJS, 192, 18

Kovac K. et al., 2009, arXiv:0910.0004v1 [astro-ph.CO]

Lahav O. et al. 2002, MNRAS, 333, 961

Lahav O., Itoh M., Inagaki S., Suto Y., 1993, ApJ, 402, 387

Lahav O., Lilje P.B., Primack J.R., Rees M.J., 1991, MNRAS, 251, 128

Larson D. et al., ApJS **192**, 16 (2011)

R. Laureijs et al., arXiv:1110.3193 (2011)

Layser D., 1956, ApJ, 61, 1243

Leavitt H.S., Pickering E.C., 1912, Harvard College Observatory Circular, 173, 1

Lemaitre, G. 1931, MNRAS, 91, 483

Lightman A.P., Schechter P.L., 1990, ApJ Suppl. Ser., 74, 831

Linder E.V., 2003, Phys. Rev. Lett, 95, 141301

Linder E.V., Jenkins A., 2003, MNRAS, 346, 573

Linder E.V., 2005, Phys. Rev. D, 72, 43529

Lue A., Scoccimarro R., & Starkman G., 2004 Phys. Rev. D 69 044005

Makino J., Sasaki M., Suto Y., 1992, Phys. Rev. D, 46, 585

Marinoni C. et al., 2005, A&A, 442, 801

Marinoni, C. et al. 2008, A&A, 487, 7

Marinoni, C., Buzzi A., 2010, Nature, 468, 539

Martinez, V. J., Lopez-Mart, B. Pons-Bordera M.-J. 2001, ApJ, 554, 5

Martinez, V. J., Saar, E., 2002, Statistics of galaxy distribution, Chapman & Hal

Matarrese S., Lucchin F., Bonometto S. A., 1986, ApJ, 310, L21

McBride, C. K. et al. 2009, AAS, 41, 253

Meeron, E. 1957 J. Chem. Phys., 27, 1238

Milgrom M., 1983, ApJ. 270, 371

Moutarde F., Alimi J.-M., Bouchet F.R., Pellat R., Ramani A., 1991, ApJ, 382, 377

Multamaki T., Gaztañaga E., & Manera M., 2003, ApJ, 344, 761

Mustapha H., Dimitrakopoulos R., 2010, C&M, 60, 2178

Pan, J. & Coles, P. 2001, MNRAS, 318, 51

Pan J., Szapudi I., 2005, MNRAS, 362, 1363

Peacock, J. A. et al., 2001, Nature, 410, 169

Peebles P. J. E., 1980, The Large-Scale Structure of the Universe (Princeton: Princeton Univ. Press)

W. J. Percival et al. 2010 MNRAS **401**, 2148 (2010)

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

Pettini M. et al., MNRAS **391**, 1499 (2008)

Piazza F., Marinoni C., 2003, Phys. Rev. Lett. 91, 1301

Pietronero, L., Montuori, M. & Sylos-Labini, F. 1997, in Critical Dialogs in Cosmology, ed Turok N (Singapore: World Scientific) pp 24

Ratra B., Peebles P.J.E., 1988, Phys. Rev., D37, 3406

Reyes R., Mandelbaum R., Seljak U., Baldauf T., Gunn J.E., Lombriser L., Smith R.E., 2010, Nature, 464, 256

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

A. G. Riess et al. **699**, 539 (2009).

Riess A. et al., 2011, ApJ, 730, 119

Robertson, H. P. 1929, PNAS, 15, 822

Romano, A. E. 2007, Phys. Rev. D, 76, 103525

A. G. Sanchez et al. arXiv:1203.6616 (2012)

Scoccimarro, R., et.al., 1998, MNRAS, 299, 1097

Schlegel D. et al., 2011, arXiv1106.1706

Schuecker P., Bohringer H., Collins C. A., Guzzo L., 2003, A&A, 398, 867

Scrimgeour, M. I. et al. 2012, MNRAS submitted, arXiv:1205.6812

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

Slipher V.M., 1912, Lowell Observatory Bulletin, 1, 26

Smith R. E., Peacock J. A., Jenkins A., White S. D. M., Frenk C. S., Pearce F. R., Thomas P. A. Efstathiou G., Couchman H. M. P., 2003, MNRAS, 341, 1311

Song Y.-S., Percival W. J., 2009, JCAP, 10, 004

Spergel D.N., 2007, ApJ, 170, 377

Stoeger, W. R., & Maartens, R., Ellis, G. F. R. 1995, ApJ, 443, 1

Strauss, M. A. et al. 1992, ApJ, 385, 421

Strauss, M. A. et al. 2002, ApJ, 124, 1810

Sylos-Labini, F. et al. 2009, A&A, 508, 17

Sylos-Labini, F. et al. 2009, A&A, 496, 7

Szapudi I., Szalay A.S., Boschán P., 1992, ApJ, 390, 350

Szapudi I., Szalay A.S., 1993, ApJ, 408, 43

Szapudi I., Dalton G.B., Efstathiou G., Szalay A.S., 1995, ApJ, 444, 520

Szapudi I., 1998, MNRAS, 300, L35

Tegmark M. et al., 2006, Phys. Rev. D, 74, 3507

Uzan, J. P., Clarkson, C. & Ellis, G. F. R. 2008, Phys. Rev. Lett., 100, 191303

Uzan J. P., Gen. Relativ. Gravit. (special issue on lensing), Jetzer P., Mellier Y. & Perlick V., eds

Verde L. et al., 2002, MNRAS, 335, 432

Walker, A. G. 1936, Proc. London Math. Soc., 42, 90

Wang L., Steinhardt P.J., 1998, ApJ, 508, 483

Wetterich C., 1988, Nucl. Phys., B302, 668

Weinberg, S. 1972, Gravitation & Cosmology, Wiley & Son, New York

Weinberg S., Cosmology, Oxford University Press (2008)

Wu, K. K. S., Lahav, O., & Rees M. J. 1999, Nature, 397, 225

Yadav, J., Bharadwaj, S., Pandey, B. & Seshadri, T. R. 2005, MNRAS, 364, 601

Zel'dovich Ya.B., 1968, Soviet Physics Uspekhi, 11, 381

Zel'dovich Ya.B., 1970, A&A, 5, 84

Zhang P., Liguori M., Bean R., Dodelson S., 2007, Phys. Rev. Lett. 99, 141302

Zhang, P., & Stebbins, A. 2011, Phys. Rev. Lett., 107, 041301

Zlatev I., Wang L., Steinhardt P.J., 1999, Phys. Rev. Lett., 82, 896

Zwicky F., 1933, Helretica Physica Acta, 6, 110

Appendix A

Second order perturbation theory

The determination of the time evolution of the amplitude of an arbitrary matter overdensity δ in an arbitrary cosmological background is a problem that can be solved only by in a perturbative way. The linear solution of this problem was already presented in chapter 1. In this section, using an original formalism, I calculate the second-order correction. Let me first introduce some definitions and notations. The spatial Fourier transform $\vec{f_k}$ of a vectorial function $\vec{f}$ that depends on position $\vec{x}$ is

$$\vec{f_k} = \mathcal{F}\left[\vec{f}(\vec{x})\right] = \frac{1}{(2\pi)^3}\int \vec{f}(\vec{x}) e^{i\vec{k}\cdot\vec{x}} d^3\vec{x},$$

and the inverse Fourier transform is

$$\vec{f}(\vec{x}) = \int \vec{f_k} e^{-i\vec{k}\cdot\vec{x}} d^3\vec{k}.$$

I take the Fourier transform of the LHS of equation (1.19)

$$\mathcal{F}\left[\frac{\partial\delta}{\partial\tau}\right] = \frac{\partial\delta_k}{\partial\tau}$$

$$\mathcal{F}\left[\theta\right] = \theta_k.$$

Since the RHS is non linear, it takes more steps to express it in Fourier space. I first expand the divergence

$$\vec{\nabla}.(\delta\vec{v}) = (\vec{\nabla}\delta).\vec{v} + \delta\vec{\nabla}.\vec{v},$$

and, then , I take the Fourier transform of each term

$$\vec{\nabla}.\vec{v} = \int -i\vec{k}_1.\vec{v}_{k_1} e^{-i\vec{k}_1.\vec{x}} d^3\vec{k}_1$$

and

$$\vec{\nabla}.\delta = \int -i\vec{k}_2.\delta_{k_2} e^{-i\vec{k}_2.\vec{x}} d^3\vec{k}_2$$

so

$$\vec{\nabla}.(\delta\vec{v}) = -\iint i\left(\vec{k}_1+\vec{k}_2\right).\vec{v}_{k_1}\delta_{k_2}e^{-i(\vec{k}_1+\vec{k}_2).\vec{x}}d^3\vec{k}_1d^3\vec{k}_2.$$

Since $\vec{\nabla}.\vec{v}=\theta$, I obtain

$$\vec{v}_{k_1} = i\frac{\vec{k}_1}{{k_1}^2}\theta_{k_1},$$

from which I deduce

$$\vec{\nabla}.(\delta\vec{v}) = \iint\left(\vec{k}_1+\vec{k}_2\right).\frac{\vec{k}_1}{{k_1}^2}\theta_{k_1}\delta_{k_2}e^{-i(\vec{k}_1+\vec{k}_2).\vec{x}}d^3\vec{k}_1d^3\vec{k}_2$$

so that, finally, I obtain

$$\mathcal{F}\left[\vec{\nabla}.(\delta\vec{v})\right] = \frac{1}{(2\pi)^3}\iiint\left(\vec{k}_1+\vec{k}_2\right).\frac{\vec{k}_1}{{k_1}^2}\theta_{k_1}\delta_{k_2}e^{i[\vec{k}-(\vec{k}_1+\vec{k}_2)].\vec{x}}d^3\vec{k}_1d^3\vec{k}_2d^3\vec{x}.$$

By defining $\vec{k}_{12}=\vec{k}_1+\vec{k}_2$, I can re-express the previous equation as

$$\mathcal{F}\left[\vec{\nabla}.(\delta\vec{v})\right] = \iint\vec{k}_{12}.\frac{\vec{k}_1}{{k_1}^2}\theta_{k_1}\delta_{k_2}\left(\frac{1}{(2\pi)^3}\int e^{i(\vec{k}-\vec{k}_{12}).\vec{x}}d^3\vec{x}\right)d^3\vec{k}_1d^3\vec{k}_2$$

where one can recognize the Dirac delta.

$$\delta^D(\vec{k}-\vec{k}_{12}) = \frac{1}{(2\pi)^3}\int e^{i(\vec{k}-\vec{k}_{12}).\vec{x}}d^3\vec{x},$$

The Fourier transform of equation (1.19) is therefore

$$\frac{\partial\delta_k}{\partial\tau}+\theta_k = -\iint\vec{k}_{12}.\frac{\vec{k}_1}{{k_1}^2}\theta_{k_1}\delta_{k_2}\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_1d^3\vec{k}_2.$$

It is convenient to define $\Theta_k=-\frac{\theta_k}{\mathcal{H}}$ where $\mathcal{H}$ is the Hubble parameter is conformal time ($\mathcal{H}\partial\tau=\partial\ln(a)$), and where a is the scale factor of the universe. I obtain the non linear continuity equation of the matter perturbations

$$\frac{\partial\delta_k}{\partial\ln a}-\Theta_k = -\iint\alpha\left(\vec{k}_1,\vec{k}_2\right)\Theta_{k_1}\delta_{k_2}\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_1d^3\vec{k}_2, \tag{A.1}$$

where

$$\alpha\left(\vec{k}_1,\vec{k}_2\right) = \vec{k}_{12}.\frac{\vec{k}_1}{{k_1}^2}.$$

Now I take the Fourier transform of equation (1.20). Since

$$\vec{v}(\vec{x}) \equiv \int\vec{v}_{k_1}e^{-i\vec{k}_1.\vec{x}}d^3\vec{k}_1,$$

it follows that

$$\vec{v}.\vec{\nabla} = \int \vec{v}_{k_1}.(-i\vec{k}_2)e^{-i(\vec{k}_1+\vec{k}_2).\vec{x}}d^3\vec{k}_1 d^3\vec{k}_2$$

from which I deduce that

$$\left(\vec{v}.\vec{\nabla}\right)\vec{v} = -i\iint \left(\vec{v}_{k_1}.\vec{k}_2\right)\vec{v}_{k_2}e^{-i(\vec{k}_1+\vec{k}_2).\vec{x}}d^3\vec{k}_1 d^3\vec{k}_2,$$

and finally

$$\vec{\nabla}.\left[\left(\vec{v}.\vec{\nabla}\right)\vec{v}\right] = -i\iint -i\left(\vec{k}_1+\vec{k}_2\right).\vec{v}_{k_2}\vec{v}_{k_1}.\vec{k}_2 e^{-i(\vec{k}_1+\vec{k}_2).\vec{x}}d^3\vec{k}_1 d^3\vec{k}_2.$$

Since

$$\begin{cases} \vec{v}_{k_1} = i\frac{\vec{k}_1}{{k_1}^2}\theta_{k_1} \\ \\ \vec{v}_{k_2} = i\frac{\vec{k}_2}{{k_2}^2}\theta_{k_2} \end{cases},$$

where, by definition, $\vec{k}_{12} = \vec{k}_1 + \vec{k}_2$, it follows that

$$\vec{\nabla}.\left[\left(\vec{v}.\vec{\nabla}\right)\vec{v}\right] = \iint \frac{\vec{k}_{12}.\vec{k}_2}{{k_2}^2}\frac{\vec{k}_1.\vec{k}_2}{{k_1}^2}\theta_{k_1}\theta_{k_2}e^{-i\vec{k}_{12}.\vec{x}}d^3\vec{k}_1 d^3\vec{k}_2.$$

Finally, the Fourier transform of equation (1.20) is given by

$$\frac{\partial\theta_k}{\partial\tau} + \mathcal{H}\theta_k + \frac{3}{2}\Omega_m\mathcal{H}^2\delta_k = -\frac{1}{(2\pi)^3}\iiint \beta\left(\vec{k}_1,\vec{k}_2\right)\theta_{k_1}\theta_{k_2}e^{i(\vec{k}-\vec{k}_{12}).\vec{x}}d^3\vec{k}_1 d^3\vec{k}_2 d^3\vec{x},$$

where

$$\beta\left(\vec{k}_1,\vec{k}_2\right) \equiv \frac{\vec{k}_{12}.\vec{k}_2}{{k_2}^2}\frac{\vec{k}_1.\vec{k}_2}{{k_1}^2}.$$

By performing a first integration in configuration space $\vec{x}$ I obtain

$$\frac{\partial\theta_k}{\partial\tau} + \mathcal{H}\theta_k + \frac{3}{2}\mathcal{H}^2\delta_k = -\iint \beta\left(\vec{k}_1,\vec{k}_2\right)\theta_{k_1}\theta_{k_2}\left(\int\frac{1}{(2\pi)^3}e^{i(\vec{k}-\vec{k}_{12}).\vec{x}}d^3\vec{x}\right)d^3\vec{k}_1 d^3\vec{k}_2,$$

a result that can be expressed in the following form

$$\frac{1}{\mathcal{H}^2}\frac{\partial\theta_k}{\partial\tau} + \frac{\theta_k}{\mathcal{H}} + \frac{3}{2}\delta_k = -\iint \beta\left(\vec{k}_1,\vec{k}_2\right)\frac{\theta_{k_1}}{\mathcal{H}}\frac{\theta_{k_2}}{\mathcal{H}}\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1 d^3\vec{k}_2.$$

By deriving with respect to conformal time τ the expression

$$\theta_k = \mathcal{H}\Theta_k,$$

I obtain

$$\frac{\partial\theta_k}{\partial\tau} = -\Theta_k\frac{\partial\mathcal{H}_k}{\partial\tau} - \mathcal{H}\frac{\partial\Theta_k}{\partial\tau}.$$

and, therefore

$$\frac{\partial\theta_k}{\partial\tau} = -\Theta_k \frac{\frac{\partial\mathcal{H}_k}{\partial\tau}}{\mathcal{H}^2} - \frac{\partial\Theta_k}{\mathcal{H}\partial\tau}$$

In order to simplify the previous expression it is necessary to express the term $\frac{\frac{\partial\mathcal{H}_k}{\partial\tau}}{\mathcal{H}^2}$, this can be done using the FL equations in conformal time, for a generically curved, 2-component Universe (matter + dark energy with time evolving equation of state). This leads to

$$\frac{\frac{\partial\mathcal{H}_k}{\partial\tau}}{\mathcal{H}^2} = -\frac{1}{2}\left\{(1+3w)\Omega_X + \Omega_m\right\},$$

so that I finally get

$$\frac{\partial\Theta_k}{\partial\ln a} + (2-(1+3w)\Omega_X - \Omega_m)\frac{\Theta_k}{2} - \frac{3}{2}\Omega_m\delta_k = \iint \beta\left(\vec{k}_1,\vec{k}_2\right)\Theta_{k_1}\Theta_{k_2}\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1 d^3\vec{k}_2. \tag{A.2}$$

By combining equations (A.1) and (A.2) I obtain the following system,

$$\begin{cases} \frac{\partial\delta_k}{\partial\ln a} - \Theta_k = -\iint \alpha\left(\vec{k}_1,\vec{k}_2\right)\Theta_{k_1}\delta_{k_2}\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_1 d^3\vec{k}_2 \\ \frac{\partial\Theta_k}{\partial\ln a} + (2-(1+3w)\Omega_X - \Omega_m)\frac{\Theta_k}{2} - \frac{3}{2}\Omega_m\delta_k = \iint \beta\left(\vec{k}_1,\vec{k}_2\right)\Theta_{k_1}\Theta_{k_2}\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1 d^3\vec{k}_2 \end{cases},$$

where non linear terms have been moved to the RHS. It is then convenient to turn the RHS into a symmetric form. Since $\vec{k}_1$ and $\vec{k}_2$ are integration variables, I can permute them in equations (A.1) and (A.2). For (A.1) I obtain

$$\begin{cases} \frac{\partial\delta_k}{\partial\ln a} - \Theta_k = -\iint \alpha\left(\vec{k}_1,\vec{k}_2\right)\Theta_{k_1}\delta_{k_2}\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_1 d^3\vec{k}_2 \\ \frac{\partial\delta_k}{\partial\ln a} - \Theta_k = -\iint \alpha\left(\vec{k}_2,\vec{k}_1\right)\Theta_{k_2}\delta_{k_1}\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_2 d^3\vec{k}_1 \end{cases}$$

and, by taking the half sum, I obtain

$$\frac{\partial\delta_k}{\partial\ln a} - \Theta_k = -\iint \left[\frac{\alpha(\vec{k}_1,\vec{k}_2)\Theta_{k_1}\delta_{k_2} + \alpha(\vec{k}_2,\vec{k}_1)\Theta_{k_2}\delta_{k_1}}{2}\right]\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_1 d^3\vec{k}_2. \tag{A.3}$$

I apply the same trick to equation (A.2)

$$\frac{\partial\Theta_k}{\partial\ln a} + (1+3w\Omega_X + \Omega_m)\frac{\Theta_k}{2} - \frac{3}{2}\Omega_m\delta_k = \iint \left[\frac{\beta(\vec{k}_1,\vec{k}_2)+\beta(\vec{k}_2,\vec{k}_1)}{2}\right]\Theta_{k_1}\Theta_{k_2}\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1 d^3\vec{k}_2, \tag{A.4}$$

where I define

$$\beta^s_{12}\left(\vec{k}_1,\vec{k}_2\right) = \frac{\beta\left(\vec{k}_1,\vec{k}_2\right)+\beta\left(\vec{k}_2,\vec{k}_1\right)}{2},$$

which gives, In terms of the wave vectors $\vec{k}_1$, $\vec{k}_2$, $\vec{k}_{12}$,

$$\beta_s\left(\vec{k}_1,\vec{k}_2\right)=\frac{1}{2}\frac{{k_{12}}^2(\vec{k}_1.\vec{k}_2)}{{k_1}^2{k_2}^2}.$$

We finally obtain

$$\begin{cases}\frac{\partial\delta_k}{\partial\ln a}-\Theta_k=-\iint\frac{1}{2}\left[\frac{\alpha_{12}\Theta_{k_1}\delta_{k_2}+\alpha_{21}\Theta_{k_2}\delta_{k_1}}{2}\right]\delta^D(\vec{k}-\vec{k}_{12})d^3\vec{k}_1d^3\vec{k}_2\\ \frac{\partial\Theta_k}{\partial\ln a}+\epsilon\Theta_k-\frac{3}{2}\Omega_m\delta_k=\iint\beta_{12}^s\left(\vec{k}_1,\vec{k}_2\right)\Theta_{k_1}\Theta_{k_2}\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1d^3\vec{k}_2\end{cases},\tag{A.5}$$

where $\alpha_{ij}\equiv\alpha\left(\vec{k}_i,\vec{k}_j\right)$ and where $\epsilon\equiv 2(1+\mathcal{H}'\mathcal{H}^2)=2+(1+3w)\Omega_X+\Omega_m$. Note that this system can be solved analytically, if the coefficients on the LHS do not depend on time, which occurs only if $\Omega_m=1$ and $\Omega_X=0$, i.e. in the case of an Einstein-de Sitter Universe (e.g. Moutarde, 1991; Goroff, 1986; Makino, 1992; Jain & Bertschinger, 1994; Bernardeau et al., 2002). However it is possible to exploit a different change of variable to obtain an approximate analytical solution. The variable μ defined as $\mu=\Theta/f$, where $f\equiv\mathrm{d}\ln D/\mathrm{d}\ln a$ is the linear growth rate, allows transforming the derivative with respect to the scale factor into derivatives with respect to the growth factor. Thus

$$\Theta=f\mu$$

and if I take the derivative with respect to $\ln a$

$$\frac{\mathrm{d}\Theta}{\mathrm{d}\ln a}=\frac{\mathrm{d}f}{\mathrm{d}\ln a}\Theta+f\frac{\mathrm{d}\Theta}{\mathrm{d}\ln a}$$

and replace $f\mathrm{d}\ln a=\mathrm{d}\ln D_1$ I obtain

$$\frac{\mathrm{d}\Theta}{\mathrm{d}\ln a}=\frac{\mathrm{d}f}{\mathrm{d}\ln a}\Theta+f^2\frac{\mathrm{d}\Theta}{\mathrm{d}\ln D_1}$$

It is now necessary to express $\frac{\mathrm{d}f}{\mathrm{d}\ln a}$ in a more convenient way. This can be done by using the first order perturbation equation which applies to the growth factor in conformal time

$$D_1''+\mathcal{H}D_1'-\frac{3}{2}\Omega_m\mathcal{H}^2D_1=0,\tag{A.6}$$

and changing variable in order to obtain the corresponding equation for the growth rate. As ${D_1}'\equiv\mathrm{d}D_1/\mathrm{d}\tau$ and $\mathcal{H}\mathrm{d}\tau=\mathrm{d}\ln a$ it follows that ${D_1}'=\mathcal{H}fD_1$ and that

$$D_1''=(\mathcal{H}'f+f'\mathcal{H})D_1+\mathcal{H}^2f^2D_1,$$

By inserting these expressions of D_1' and D_1'' into eq. (A.6) and by dividing it with respect

to the Hubble rate in confomal time $\mathcal{H}$ I obtain

$$D_1 \left\{ \frac{\mathcal{H}'}{\mathcal{H}^2} f + \frac{\mathrm{d}f}{\mathrm{d}\ln a} + f^2 + f - \frac{3}{2}\Omega_m \right\} = 0.$$

and, therefore, a differential equation for the growth rate f (since the growing mode is different from 0 for all times),

$$\frac{\mathrm{d}f}{\mathrm{d}\ln a} = \frac{3}{2}\Omega_m - \left(\frac{\mathcal{H}'}{\mathcal{H}^2} + 1\right) - f^2.$$

Accordingly, the system of equations (A.5) becomes, in terms of the new variable μ,

$$\left\{ \begin{array}{l} \frac{\partial \delta_k}{\partial \ln D_1} - \mu_k = -\iint \frac{1}{2}\left[\frac{\alpha_{12}\mu_{k_1}\delta_{k_2} + \alpha_{21}\mu_{k_2}\delta_{k_1}}{2}\right] \delta^D(\vec{k} - \vec{k}_{12}) d^3\vec{k}_1 d^3\vec{k}_2 \\ \frac{\partial \mu_k}{\partial \ln D_1} + \left(\frac{3}{2}\frac{\Omega_m}{f^2} - 1\right)\mu_k - \frac{3}{2}\frac{\Omega_m}{f^2}\delta_k = \iint \beta_{12}^s\left(\vec{k}_1, \vec{k}_2\right) \mu_{k_1}\mu_{k_2}\delta^D\left(\vec{k} - \vec{k}_{12}\right) d^3\vec{k}_1 d^3\vec{k}_2 \end{array} \right. ,$$

It is standard to assume that the solutions of this non linear system can be expressed as a sum where each contribution is sorted by order of magnitude (i.e. $X^{(n)} \gg X^{(n+1)}$)

$$\delta(\vec{x}, t) = \sum_{n=1}^{\infty} \delta^{(n)}(\vec{x}, t) \tag{A.7}$$

$$\theta(\vec{x}, t) = \sum_{n=1}^{\infty} \theta^{(n)}(\vec{x}, t), \tag{A.8}$$

which can be expressed in Fourier space. Indeed, let's suppose that $\frac{\Omega_m}{f^2} \simeq 1$, which is a fairly good approximation, since it has been shown that, for a wide range of cosmologies, $f \simeq \Omega_m^{0.55}$. So it becomes possible to look for separable (e.g. Bernardeau et al., 2002) solutions of the form

$$\delta_k(t) = \sum_{n=1}^{\infty} D_n(t)\delta_k^{(n)} \tag{A.9}$$

$$\mu_k(t) = -\mathcal{H}f \sum_{n=1}^{\infty} E_n(t)\theta_k^{(n)}, \tag{A.10}$$

where the time dependency of each order (n) is encoded in the functions D_n and E_n which are, by definition, scale independent. The factorized time dependent part of the divergence of the velocity field (i.e. $-\mathcal{H}f$) guarantees that, at linear order, $E_1 = D_1$. The following calculation is made using the formalism introduced by Crocce & Scoccimarro (2006) Let's represent the solutions as a 2 dimensional vector, which leads to the compact representation

$$\overrightarrow{\Psi}(\vec{k}_1) = \begin{bmatrix} \delta_{k_1} \\ \mu_{k_1} \end{bmatrix}$$

where $\Psi_i\left(\vec{k}_1\right)$ is the i^{th} component of vector $\overrightarrow{\Psi}(\vec{k}_1)$, that is

$$\begin{cases} \Psi_1\left(\vec{k}_1\right) = \delta_{k_1} \\ \Psi_2\left(\vec{k}_1\right) = \mu_{k_1} \end{cases}$$

Let's start by transforming the LHS using a matrix representation

$$\begin{bmatrix} \frac{\partial \delta_k}{\partial \ln D_1} - \mu_k \\ \frac{\partial \mu_k}{\partial \ln D_1} + \frac{1}{2}\Theta_k - \frac{3}{2}\delta_k \end{bmatrix} = \frac{\partial}{\partial \ln a}\overrightarrow{\Psi}(\vec{k}) + \Omega \times \overrightarrow{\Psi}(\vec{k}),$$

where

$$\Omega = \begin{bmatrix} 0 & -1 \\ \frac{-3}{2} & \frac{1}{2} \end{bmatrix}.$$

An element of the Ω matrix is identified by Ω_i^j, which allows us to write

$$\frac{\partial}{\partial \ln a}\overrightarrow{\Psi}(\vec{k}) + \Omega \times \overrightarrow{\Psi}(\vec{k}) \Leftrightarrow \frac{\partial}{\partial \ln a}\Psi_i(\vec{k}) + \Omega_i^j \Psi_j(\vec{k}),$$

where

$$\Omega_i^j \Psi_j = \sum_j \Omega_{i,j}\Psi_j.$$

The RHS is somewhat more complicate. It necessary to define two matrix γ_1 and γ_2 as

$$\gamma_1 = \begin{bmatrix} 0 & \frac{\alpha(\vec{k}_2,\vec{k}_1)}{2}\delta^D\left(\vec{k}-\vec{k}_{12}\right) \\ \frac{\alpha(\vec{k}_1,\vec{k}_2)}{2}\delta^D\left(\vec{k}-\vec{k}_{12}\right) & 0 \end{bmatrix},$$

and

$$\gamma_2 = \begin{bmatrix} 0 & 0 \\ 0 & \beta_s\left(\vec{k}_1,\vec{k}_2\right)\delta^D\left(\vec{k}-\vec{k}_{12}\right) \end{bmatrix}.$$

Therefore, the full non linear system can be written in a quite simple way

$$\frac{\partial}{\partial \ln a}\Psi_i(\vec{k}) + \Omega_i^j \Psi_j(\vec{k}) = \iint \gamma_i^{jl}\Psi_j(\vec{k}_1)\Psi_l(\vec{k}_2)d^3\vec{k}_1 d^3\vec{k}_2 \tag{A.11}$$

We now look for solutions of the type $\Psi_i(\vec{k},\tau)$,

$$\Psi_i(\vec{k},\tau) = \Psi_i^{(1)}(\vec{k}) + \Psi_i^{(2)}(\vec{k}) + ... + \Psi_i^{(n)}(\vec{k}) = \sum_{n=1}^{\infty}\Psi_i^{(n)}(\vec{k})$$

Let's see what happen when I insert this perturbative solution in (A.11). Since

$$\Psi_j(\vec{k}_1) = \sum_{n=1}^{\infty}\Psi_j^{(n)}(\vec{k}_1),$$

and

$$\Psi_l(\vec{k}_2) = \sum_{m=1}^{\infty} \Psi_l^{(m)}(\vec{k}_2).$$

their product is

$$\Psi_j(\vec{k}_1)\Psi_l(\vec{k}_2) = \sum_{n=1}^{\infty}\left[\Psi_j^{(n)}(\vec{k}_1)\sum_{m=1}^{\infty}\left(a^m\Psi_l^{(m)}(\vec{k}_2)\right)\right],$$

which can be written as a double sum

$$\Psi_j(\vec{k}_1)\Psi_l(\vec{k}_2) = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\Psi_j^{(n)}(\vec{k}_1)\Psi_l^{(m)}(\vec{k}_2).$$

Then it is convenient to shift indexes as $\lambda = n + m$

$$\Psi_j(\vec{k}_1)\Psi_l(\vec{k}_2) = \sum_{\lambda=2}^{\infty}\sum_{m=1}^{\lambda-1}\Psi_j^{(\lambda-m)}(\vec{k}_1)\Psi_l^{(m)}(\vec{k}_2),$$

and define $\Psi_j^{(0)}(\vec{k}_1) = 0$. By setting $\lambda = n$ I obtain

$$\Psi_j(\vec{k}_1)\Psi_l(\vec{k}_2) = \sum_{n=1}^{\infty}\sum_{m=1}^{n-1}\Psi_j^{(n-m)}(\vec{k}_1)\Psi_l^{(m)}(\vec{k}_2).$$

Now it is necessary to express the derivative $\frac{\partial}{\partial \ln a}\Psi_i(\vec{k})$. To this purpose it is convenient to define the vector $\Phi(D_1)$ such that

$$\Phi^{(n)} \equiv \begin{bmatrix} D_n \\ E_n \end{bmatrix}.$$

It follows that

$$\begin{aligned}
\Psi_i(\vec{k},\tau) &= \sum_{n=1}^{\infty}\Phi_i^{(n)}\Psi_i^{(n)}(\vec{k}) \\
\tfrac{\partial}{\partial D_1}\Psi_i(\vec{k},\tau) &= \sum_{n=1}^{\infty}\frac{\partial}{\partial D_1}\Phi_i^{(n)}\Psi_i^{(n)}(\vec{k}) \\
D_1\tfrac{\partial}{\partial D_1}\Psi_i(\vec{k},\tau) &= \sum_{n=1}^{\infty}D_1\frac{\partial}{\partial D_1}\Phi_i^{(n)}\Psi_i^{(n)}(\vec{k}) \\
\tfrac{\partial}{\partial \ln D_1}\Psi_i(\vec{k},\tau) &= \sum_{n=1}^{\infty}\mathrm{I}_{\mathrm{i}}^{\mathrm{j}}\Phi_j^{(n)}\Psi_j^{(n)}(\vec{k})
\end{aligned}$$

where

$$\mathrm{I} = \begin{bmatrix} f_n & 0 \\ 0 & g_n \end{bmatrix}$$

and where $f_n = \frac{\mathrm{d}\ln D_n}{\mathrm{d}\ln D_1}$ and $g_n = \frac{\mathrm{d}\ln E_n}{\mathrm{d}\ln D_1}$. By replacing these terms in equation (A.11) I obtain

$$\sum_{n=1}^{\infty}\left(\mathrm{I}_i^j + \Omega_i^j\right)\Psi_j^{(n)}(\vec{k}) = \sum_{n=1}^{\infty}\iint\gamma_i^{jl}\sum_{m=1}^{n-1}\Psi_j^{(n-m)}(\vec{k}_1)\Psi_l^{(m)}(\vec{k}_2)d^3\vec{k}_1 d^3\vec{k}_2,$$

which can be simplified as

$$\mathrm{A}^{-1}{}_i^j \Psi_j^{(n)}(\vec{k}) = \sum_{m=1}^{n-1} \iint \gamma_i^{jl} \Psi_j^{(n-m)}(\vec{k}_1) \Psi_l^{(m)}(\vec{k}_2) d^3\vec{k}_1 d^3\vec{k}_2, \tag{A.12}$$

where

$$\mathrm{A}^{-1} = \mathrm{I} + \Omega = \begin{bmatrix} f_n & -1 \\ -\frac{3}{2} & g_n + \frac{1}{2} \end{bmatrix}.$$

To solve this system of differential equations we need to invert A^{-1}. As a consequence its determinant must be non zero. On the other hand it is possible to show that the only possible solution is obtained for $D_n = E_n = D_1^n$, therefore $g_n = f_n = n$. Thus the determinant is expressed as

$$\begin{aligned} det(\mathrm{A}^{-1}) &= f_n(g_n + \tfrac{1}{2}) - \tfrac{3}{2} \\ &= n^2 + n\tfrac{1}{2} - \tfrac{3}{2} \\ &= \tfrac{1}{2}\left(2n^2 + n - 3\right) \\ &= \tfrac{1}{2}(n-1)(2n+3). \end{aligned}$$

As a result, since $n \geq 2$ the conditions $n \neq 0$ and $n \neq -3/2$ are satisfied and the matrix A^{-1} can be inverted. It leads to

$$\mathrm{A} = \frac{1}{(n-1)(2n+3)} \begin{bmatrix} 2n+1 & 2 \\ 3 & 2 \end{bmatrix},$$

thus the complete hierarchy of equations can be solved using a matrix representation when $f \simeq \Omega_m^{0.5}$ (which corresponds nearly to the Einstein-de Sitter solution). In other terms by substituting D_1 by a one recover the general solution for the universe containing only matter and zero curvature ($\Omega_m = 1$). The solution up to a given oder must be derived in a recursive way, by using equation (A.12). For instance, the second order solution can be calculated by setting $n = 2$, which leads to

$$\Psi_i^{(2)}(\vec{k}) = \mathrm{A}_i^k \iint \gamma_k^{jl} \Psi_j^{(1)}(\vec{k}_1) \Psi_l^{(1)}(\vec{k}_2) d^3\vec{k}_1 d^3\vec{k}_2,$$

where

$$\mathrm{A} = \frac{1}{7} \begin{bmatrix} 5 & 2 \\ 3 & 4 \end{bmatrix}.$$

The solution can therefore be represented as

$$\begin{bmatrix} \delta_k^{(2)} \\ \theta_k^{(2)} \end{bmatrix} = \mathrm{A} \times \begin{bmatrix} \iint \gamma_1^{jl} \Psi_j^{(1)}(\vec{k}_1) \Psi_l^{(1)}(\vec{k}_2) d^3\vec{k}_1 d^3\vec{k}_2 \\ \iint \gamma_2^{jl} \Psi_j^{(1)}(\vec{k}_1) \Psi_l^{(1)}(\vec{k}_2) d^3\vec{k}_1 d^3\vec{k}_2. \end{bmatrix},$$

If I consider only matter perturbations I obtain

$$7\delta_k^{(2)} = 5\iint \gamma_1^{jl}\Psi_j^{(1)}(\vec{k}_1)\Psi_l^{(1)}(\vec{k}_2)d^3\vec{k}_1 d^3\vec{k}_2 + 2\iint \gamma_2^{jl}\Psi_j^{(1)}(\vec{k}_1)\Psi_l^{(1)}(\vec{k}_2)d^3\vec{k}_1 d^3\vec{k}_2.$$

By substituting the corresponding matrix γ_1 and γ_2 we get

$$\delta_k^{(2)} = \frac{1}{7}\iint \left[\frac{5}{2}\alpha(\vec{k}_1,\vec{k}_2)\theta_{k_1}^{(1)}\delta_{k_2}^{(1)} + \frac{5}{2}\alpha(\vec{k}_2,\vec{k}_1)\theta_{k_2}^{(1)}\delta_{k_1}^{(1)} + 2\beta_s(\vec{k}_1,\vec{k}_2)\theta_{k_1}^{(1)}\theta_{k_2}^{(1)}\right]\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1 d^3\vec{k}_2.$$

Since, at linear order, $\theta_k = \delta_k$ I obtain

$$\delta_k^{(2)} = \frac{1}{7}\iint F_2\left(\vec{k}_1,\vec{k}_2\right)\delta_{k_1}^{(1)}\delta_{k_2}^{(1)}\delta^D\left(\vec{k}-\vec{k}_{12}\right)d^3\vec{k}_1 d^3\vec{k}_2,$$

where

$$F_2\left(\vec{k}_1,\vec{k}_2\right) = \frac{5}{2}\alpha(\vec{k}_1,\vec{k}_2) + \frac{5}{2}\alpha(\vec{k}_2,\vec{k}_1) + 2\beta_s(\vec{k}_1,\vec{k}_2),$$

is called the second order perturbation theory kernel. This kernel , that can be expressed in terms of the wave vectors $\vec{k}_1$ et $\vec{k}_2$ as

$$F_2\left(\vec{k}_1,\vec{k}_2\right) = \frac{5}{7} + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right) + \frac{2}{7}\left(\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\right)^2.$$

displays how non-linear effect are generated by the mixing of wave vectors, a phenomenon usually referred to as mode coupling.

Now I briefly outline how one can address the the problem of finding the 'true' non-linear solution, that is the solutions of the system (A.5) when we take into account the time dependency of the coefficients ϵ and Ω_m. Following the same matrix formalism previously introduced it follows that

$$\frac{\mathrm{d}D_n}{\mathrm{d}\ln D_1}\delta_k^{(n)} - E_n\theta_k^{(n)} = \int \mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\Delta_n(\vec{k}_1,\vec{k}_2)\delta^D(\vec{k}-\vec{k}_{12}) \quad \text{(A.13)}$$

$$\frac{\mathrm{d}E_n}{\mathrm{d}\ln D_1}\theta_k^{(n)} + [\frac{3}{2}\frac{\Omega_m}{f^2} - 1]E_n\theta_k^{(n)} + \frac{3}{2}\frac{\Omega_m}{f^2}D_n\delta_k^{(n)} = \int \mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\Sigma_n(\vec{k}_1,\vec{k}_2)\delta^D(\vec{k}-\vec{k}_{12}), \quad \text{(A.14)}$$

where

$$\Delta_n(\vec{k}_1,\vec{k}_2) \equiv \frac{1}{2}\sum_{m=1}^{n-1}\left\{\alpha_{12}E_{n-m}D_m\theta_{k_1}^{(n-m)}\delta_{k_2}^{(m)} + \alpha_{21}E_mD_{n-m}\theta_{k_2}^{(n-m)}\delta_{k_1}^{(m)}\right\},$$

and

$$\Sigma_n(\vec{k}_1,\vec{k}_2) \equiv \beta_{12}^s\sum_{m=1}^{n-1}E_{n-m}E_m\theta_{k_1}^{(n-m)}\theta_{k_2}^{(m)}.$$

This system of equation don't have any solution, in this case, the guess on the space and time separability of the solutions is no longer useful.

Specifically, I show the calculation for the second order of matter fluctuations. Therefore we are left with the system of equations

$$\frac{\mathrm{d}\delta_k^{(2)}}{\mathrm{d}\ln D_1} - \theta_k^{(2)} = D_1^2 \iint \mathrm{d}^3\vec{k}_1 \mathrm{d}^3\vec{k}_2 \frac{\delta_{k_1}^{(1)}\delta_{k_2}^{(1)}}{2} \left\{\alpha_{12} + \alpha_{21}\right\} \delta^D(\vec{k} - \vec{k}_{12}) \quad \text{(A.15)}$$

$$\frac{\mathrm{d}\theta_k^{(2)}}{\mathrm{d}\ln D_1} + [\frac{3}{2}\frac{\Omega_m}{f^2} - 1]\theta_k^{(2)} + \frac{3}{2}\frac{\Omega_m}{f^2}\delta_k^{(2)} = D_1^2 \iint \mathrm{d}^3\vec{k}_1 \mathrm{d}^3\vec{k}_2 \delta_{k_1}^{(1)}\delta_{k_2}^{(1)}\beta_{12}^s \delta^D(\vec{k} - \vec{k}_{12}), \quad \text{(A.16)}$$

where $\delta_k^{(2)}$ and $\theta_k^{(2)}$ depend on both k-modes and D_1. This system can be written in a concise way as

$$\frac{\mathrm{d}\delta_k^{(2)}}{\mathrm{d}\ln D_1} - \theta_k^{(2)} = D_1^2 A_k \quad \text{(A.17)}$$

$$\frac{\mathrm{d}\theta_k^{(2)}}{\mathrm{d}\ln D_1} + [\epsilon - 1]\theta_k^{(2)} + \epsilon\delta_k^{(2)} = D_1^2 B_k, \quad \text{(A.18)}$$

where $\epsilon \equiv \frac{3}{2}\frac{\Omega_m}{f^2}$,

$$A_k \equiv \iint \mathrm{d}^3\vec{k}_1 \mathrm{d}^3\vec{k}_2 \frac{\delta_{k_1}^{(1)}\delta_{k_2}^{(1)}}{2} \left\{\alpha_{12} + \alpha_{21}\right\} \delta^D(\vec{k} - \vec{k}_{12}),$$

and

$$B_k \equiv \iint \mathrm{d}^3\vec{k}_1 \mathrm{d}^3\vec{k}_2 \delta_{k_1}^{(1)}\delta_{k_2}^{(1)}\beta_{12}^s \delta^D(\vec{k} - \vec{k}_{12}).$$

By deriving eq. (A.17) I obtain $\frac{\mathrm{d}^2\delta_k^{(2)}}{\mathrm{d}\ln^2 D_1} - \frac{\mathrm{d}\theta_k^{(2)}}{\mathrm{d}\ln D_1} = 2D_1^2 A_k$ from which it is possible to express the second order differential equation of matter fluctuations as

$$\frac{\mathrm{d}^2\delta_k^{(2)}}{\mathrm{d}\ln^2 D_1} + [\epsilon - 1]\frac{\mathrm{d}\delta_k^{(2)}}{\mathrm{d}\ln D_1} - \epsilon\delta_k^{(2)} = D_1^2 \left\{B_k + [1 + \epsilon]A_k\right\}.$$

This is a non-homogeneous, linear, second order differential equation with time varying coefficients. Note that, since the homogeneous equation is, by construction, the equation ruling the evolution of linear fluctuations, we already know that its general solution is proportional to the growth factor D_1. As a consequence, I need only to find a particular solution. Since the RHS is decomposed in two terms proportional to A_k and B_k respectively, I look for particular solution characterized by the same decomposition. Moreover, since the growth factor D_1^2 appears as a factor in the LHS, and since I derive with respect to the logarithm of D_1, it is also possible to factorize the solution with D_1^2, thus I set

$$\delta_k^{(2)} = D_1^2 \left\{g_B B_k + g_A A_k\right\},$$

and by deriving it two times with respect to $\ln D_1$ and by replacing the result in the above equation I obtain two differential equations, one for g_A and another for g_B

$$g_B^{''} + (\epsilon + 3)g_B^{'} + (\epsilon + 2)g_B = 1 \tag{A.19}$$

$$g_A^{''} + (\epsilon + 3)g_A^{'} + (\epsilon + 2)g_A = 1 + \epsilon, \tag{A.20}$$

where the prime indicates the drivative with respect to the logarithm of the linear growth factor D_1. One can remark that these two equations are not independent because by substituting $g_A = 1 - g_B$ in eq. A.20 one recovers equation eq. A.20. As a consequence, the general second order solution that we are looking for, can be written as

$$\delta_k^{(2)} = D_1^2 \left\{ gB_k + (1-g)A_k \right\},$$

where $g \equiv g_B$. After some algebra this solution can be expressed as

$$\delta_k^{(2)} = \iint F_2\left(\vec{k}_1, \vec{k}_2\right) \delta_{k_1}^{(1)} \delta_{k_2}^{(1)} \delta^D\left(\vec{k} - \vec{k}_{12}\right) d^3\vec{k}_1 d^3\vec{k}_2, \tag{A.21}$$

where

$$F_2\left(\vec{k}_1, \vec{k}_2\right) = (1-g) + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\left(\frac{k_2}{k_1} + \frac{k_1}{k_2}\right) + g\left(\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\right)^2, \tag{A.22}$$

a term that is usually called the second order perturbation theory kernel of the matter field.

It is interesting to compare the expression that I obtained with results from other approaches in the literature. In (Catelan et al., 1995) the second order PT kernel depends on a function B such as

$$F_2\left(\vec{k}_1, \vec{k}_2\right) = (1-B) + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\left(\frac{k_2}{k_1} + \frac{k_1}{k_2}\right) + B\left(\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\right)^2.$$

It is obvious that the function g that I introduced can be identified with the function B. Furthermore, Bouchet (1992) parameterized the second order PT kernel in terms of the the κ function

$$F_2\left(\vec{k}_1, \vec{k}_2\right) = \left(\frac{1}{2} + \kappa\right) + \frac{1}{2}\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\left(\frac{k_2}{k_1} + \frac{k_1}{k_2}\right) + \left(\frac{1}{2} - \kappa\right)\left(\frac{\vec{k}_1.\vec{k}_2}{k_1 k_2}\right)^2.$$

As a consequence, we see that $\kappa + g = 1/2$. Finally, by comparing my result (cfr. eq.

(1.39)) with those of Kamionkowski & Buchalter (1999)

$$F_2\left(\vec{k}_1,\vec{k}_2\right)=\frac{1}{2}\left\{(1+\mu)+\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\left(\frac{k_2}{k_1}+\frac{k_1}{k_2}\right)+(1-\mu)\left(\frac{\vec{k}_1.\vec{k}_2}{k_1k_2}\right)^2\right\},$$

I conclude that $\mu+2g=1$. These multiple definition of the second order growth factor obtained following different analytical approaches lead exactly to the same expression for the second order perturbation theory kernel F_2. This convergence makes the use of the perturbation theory kernel extremely robust and reliable.

Appendix B

2-point moments with respect to 2-point cumulants

In this Appendix I list the relations between two-point moments of the density contrast $\mu_{nm} \equiv \langle \delta^n(\vec{x}_1)\delta^m(\vec{x}_2)\rangle$ as a function of the one- and two-point cumulant moments $\kappa_n \equiv \langle \delta^n\rangle_c$ and $\kappa_{nm} \equiv \langle \delta^n(\vec{x}_1)\delta^m(\vec{x}_2)\rangle_c$:

$$
\begin{aligned}
\mu_{1,1} &= \kappa_{1,1} \\
\\
\mu_{12} &= \kappa_{12} \\
\\
\mu_{13} &= \kappa_{13} + 3\,\kappa_{11}\kappa_2 \\
\mu_{22} &= \kappa_{22} + 2\,{\kappa_{11}}^2 + {\kappa_2}^2 \\
\\
\mu_{14} &= \kappa_{14} + 6\,\kappa_{12}\kappa_2 + 4\,\kappa_{11}\kappa_3 \\
\mu_{23} &= 3\,\kappa_{12}\kappa_2 + 6\,\kappa_{12}\kappa_{11} + \kappa_3\kappa_2 + \kappa_{23} \\
\\
\mu_{15} &= 15\,\kappa_{11}{\kappa_2}^2 + \kappa_{15} + 5\,\kappa_{11}\kappa_4 + 10\,\kappa_{12}\kappa_3 + 10\,\kappa_{13}\kappa_2 \\
\mu_{24} &= 8\,\kappa_{13}\kappa_{11} + 6\,{\kappa_{12}}^2 + \kappa_{24} + 3\,{\kappa_2}^3 + 4\,\kappa_{12}\kappa_3 + 6\,\kappa_{22}\kappa_2 + 12\,{\kappa_{11}}^2\kappa_2 + \kappa_4\kappa_2 \\
\mu_{33} &= 9\,\kappa_{11}{\kappa_2}^2 + 9\,{\kappa_{12}}^2 + 6\,{\kappa_{11}}^3 + \kappa_{33} + 6\,\kappa_{13}\kappa_2 + {\kappa_3}^2 + 9\,\kappa_{22}\kappa_{11} \\
\\
\mu_{16} &= 6\,\kappa_{11}\kappa_5 + \kappa_{16} + 60\,\kappa_{11}\kappa_3\kappa_2 + 20\,\kappa_{13}\kappa_3 + 15\,\kappa_{14}\kappa_2 + 45\,\kappa_{12}{\kappa_2}^2 + 15\,\kappa_{12}\kappa_4 \\
\mu_{25} &= 20\,\kappa_{13}\kappa_{12} + 20\,\kappa_3{\kappa_{11}}^2 + 60\,\kappa_{12}\kappa_{11}\kappa_2 + \kappa_{25} + 15\,\kappa_{12}{\kappa_2}^2 + 10\,\kappa_{23}\kappa_2 + \kappa_5\kappa_2 \\
&\quad +10\,\kappa_3{\kappa_2}^2 + 5\,\kappa_{12}\kappa_4 + 10\,\kappa_{14}\kappa_{11} + 10\,\kappa_{22}\kappa_3 \\
\mu_{34} &= 12\,\kappa_{13}\kappa_{12} + 36\,\kappa_{12}\kappa_{11}\kappa_2 + \kappa_{34} + 12\,\kappa_{11}\kappa_3\kappa_2 + 4\,\kappa_{13}\kappa_3 + 3\,\kappa_{14}\kappa_2 + 18\,\kappa_{12}{\kappa_2}^2 \\
&\quad +6\,\kappa_{23}\kappa_2 + 3\,\kappa_3{\kappa_2}^2 + 18\,\kappa_{12}\kappa_{22} + 12\,\kappa_{23}\kappa_{11} + 36\,{\kappa_{11}}^2\kappa_{12} + \kappa_3\kappa_4
\end{aligned}
$$

These relations are valid for any discrete or continuous field, as long as the field is defined via its density contrast $\delta(\vec{x}) \equiv \frac{\lambda(\vec{x})}{\lambda} - 1$.

Appendix C

Reduced correlators and cumulants of the smoothed matter field

Cumulants $S_{N,R}$ and correlators $C_{nm,R}$ of the density contrast of the matter field smoothed with a spherical Top-Hat window function have already been calculated in the weakly non linear regime (Bernardeau, 1994a,b, 1996). Here I briefly reproduce the calculation of the reduced correlor $C_{12,R}$.

To this purpose, it is necessary to express the third order 2-point cumulant moment of the smoothed density contrast of the matter field $\langle\delta(\vec{x}_1)\delta^2(\vec{x}_2)\rangle$. In Chapter 1, I have shown that in the weakly non-linear regime the matter density contrast can be expressed as

$$\delta(\vec{k},\tau) = D(\tau)\delta^{(1)}(\vec{k}) + D^2(\tau)\delta^{(2)}(\vec{k}), \tag{C.1}$$

where

$$\delta^{(2)}(\vec{k}) = \iint F_2\left(\vec{q}_1,\vec{q}_2\right)\delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{q}_2)\delta^D\left(\vec{k}-\vec{q}_1-\vec{q}_2\right)d^3\vec{q}_1 d^3\vec{q}_2,$$

and where

$$F_2\left(\vec{q}_1,\vec{q}_2\right) = \frac{5}{7} + \frac{1}{2}\frac{\vec{q}_1\cdot\vec{q}_2}{q_1 q_2}\left(\frac{q_2}{q_1}+\frac{q_1}{q_2}\right) + \frac{2}{7}\left(\frac{\vec{q}_1\cdot\vec{q}_2}{q_1 q_2}\right)^2.$$

Multiplying equation (C.1) by the Fourier transform of the spherical Top-Hat window function $\hat{W}(kR)$ and taking its inverse Fourier transform I get

$$\delta_R(\vec{x},\tau) = \underbrace{D(\tau)\delta_R^{(1)}(\vec{x})}_{\equiv\delta_1(\vec{x})} + \underbrace{D^2(t)\delta_R^{(2)}(\vec{x})}_{\equiv\delta_2(\vec{x})}, \tag{C.2}$$

where

$$\delta_R^{(1)}(\vec{x}) = \int \delta^{(1)}(\vec{k})\hat{W}(kR)e^{i\vec{k}\cdot\vec{x}}\mathrm{d}^3\vec{k}, \tag{C.3}$$

and where

$$\delta_R^{(2)}(\vec{x}) = \iint F_2\left(\vec{q}_1, \vec{q}_2\right) \delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{q}_2)\hat{W}[|\vec{q}_1 + \vec{q}_2|R]e^{i(\vec{q}_1+\vec{q}_2)\cdot\vec{x}}\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2. \tag{C.4}$$

The smoothed density contrast (eq. C.2) can be taken at two positions $(\vec{x}_1)$ and $(\vec{x}_2)$, thus

$$\delta_R(\vec{x}_1) = \delta_1(\vec{x}_1) + \delta_2(\vec{x}_1),$$

and

$$\delta_R^2(\vec{x}_2) = \delta_1^2(\vec{x}_2) + \delta_2^2(\vec{x}_2) + 2\delta_1(\vec{x}_2)\delta_2(\vec{x}_2),$$

where for simplicity I drop the explicit dependence on time. By neglecting terms of order strictly greater than two, I can express the product

$$\delta_R(\vec{x}_1)\delta_R^2(\vec{x}_2) \simeq \delta_1(\vec{x}_1)\delta_1^2(\vec{x}_2) + \delta_2(\vec{x}_1)\delta_1^2(\vec{x}_2) + 2\delta_1(\vec{x}_1)\delta_1(\vec{x}_2)\delta_2(\vec{x}_2).$$

And, by taking its ensemble average, I get

$$\langle\delta_R(\vec{x}_1)\delta_R^2(\vec{x}_2)\rangle \simeq \langle\delta_1(\vec{x}_1)\delta_1^2(\vec{x}_2)\rangle + 2\langle\delta_1(\vec{x}_1)\delta_1(\vec{x}_2)\delta_2(\vec{x}_2)\rangle + \langle\delta_2(\vec{x}_1)\delta_1^2(\vec{x}_2)\rangle,$$

Since initial conditions of the linear order are assumed to be drawn from a multivariate Gaussian distribution it results that $\langle\delta_1(\vec{x}_1)\delta_1^2(\vec{x}_2)\rangle = 0$, thus the third order 2-point cumulant can be expressed at the leading order as a function of two contributions

$$\langle\delta_R(\vec{x}_1)\delta_R^2(\vec{x}_2)\rangle \simeq \underbrace{2\langle\delta_1(\vec{x}_1)\delta_1(\vec{x}_2)\delta_2(\vec{x}_2)\rangle}_{I} + \underbrace{\langle\delta_2(\vec{x}_1)\delta_1^2(\vec{x}_2)\rangle}_{J}. \tag{C.5}$$

Given the expressions of the first and the second order terms (eq. C.3 and C.4) I can express I as

$$\begin{aligned} I = 2D^4(t)\int \quad & F_2(\vec{q}_1, \vec{q}_2)\hat{W}(k_1R)\hat{W}(k_2R)\hat{W}[q_{12}R]\langle\delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{q}_2)\delta^{(1)}(\vec{k}_1)\delta^{(1)}(\vec{k}_2)\rangle \\ & \times e^{i(\vec{k}_1\cdot\vec{x}_1+\vec{k}_2\cdot\vec{x}_2+\vec{q}_{12}\cdot\vec{x}_2)}\mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2, \end{aligned}$$

where $\vec{q}_{12} = \vec{q}_1 + \vec{q}_2$. Since the 4-point moments (in the above expression) involves only the initial perturbation field (which is Gaussian), the Wick theorem (Bernardeau, 2008) applies and the 4-point moments can be expressed as a sum of products of the 2-point moments

$$\begin{aligned} \langle\delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{q}_2)\delta^{(1)}(\vec{k}_1)\delta^{(1)}(\vec{k}_2)\rangle = \quad & \langle\delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{q}_2)\rangle\langle\delta^{(1)}(\vec{k}_1)\delta^{(1)}(\vec{k}_2)\rangle + \\ & \langle\delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{k}_1)\rangle\langle\delta^{(1)}(\vec{q}_2)\delta^{(1)}(\vec{k}_2)\rangle + \\ & \langle\delta^{(1)}(\vec{q}_1)\delta^{(1)}(\vec{k}_2)\rangle\langle\delta^{(1)}(\vec{q}_2)\delta^{(1)}(\vec{k}_1)\rangle. \end{aligned}$$

It results that $I = I_0 + I_1 + I_2$ where

$$I_0 = 2D^4(t) \int \begin{array}{l} F_2(\vec{q}_1, \vec{q}_2)\hat{W}(k_1R)\hat{W}(k_2R)\hat{W}[q_{12}R]\delta_D(\vec{q}_1 + \vec{q}_2)P(\vec{q}_1)\delta_D(\vec{k}_1 + \vec{k}_2)P(\vec{k}_1) \\ \times e^{i(\vec{k}_1\cdot\vec{x}_1 + \vec{k}_2\cdot\vec{x}_2 + \vec{q}_{12}\cdot\vec{x}_2)}\mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2, \end{array}$$

$$I_1 = 2D^4(t) \int \begin{array}{l} F_2(\vec{q}_1, \vec{q}_2)\hat{W}(k_1R)\hat{W}(k_2R)\hat{W}[q_{12}R]\delta_D(\vec{q}_1 + \vec{k}_1)P(\vec{q}_1)\delta_D(\vec{q}_2 + \vec{k}_2)P(\vec{q}_2) \\ \times e^{i(\vec{k}_1\cdot\vec{x}_1 + \vec{k}_2\cdot\vec{x}_2 + \vec{q}_{12}\cdot\vec{x}_2)}\mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2, \end{array}$$

and

$$I_2 = 2D^4(t) \int \begin{array}{l} F_2(\vec{q}_1, \vec{q}_2)\hat{W}(k_1R)\hat{W}(k_2R)\hat{W}[q_{12}R]\delta_D(\vec{q}_1 + \vec{k}_1)P(\vec{q}_1)\delta_D(\vec{q}_2 + \vec{k}_2)P(\vec{q}_2) \\ \times e^{i(\vec{k}_1\cdot\vec{x}_1 + \vec{k}_2\cdot\vec{x}_2 + \vec{q}_{12}\cdot\vec{x}_2)}\mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2, \end{array}$$

where I used $\langle\delta^{(1)}(\vec{k}_1)\delta^{(1)}(\vec{k}_2)\rangle = \delta_D(\vec{k}_1 + \vec{k}_2)P(\vec{k}_1)$ and P is the initial power spectrum. One can remark that by exchanging $\vec{q}_1$ and $\vec{q}_2$ in I_2 we recover I_1 thus $I = I_0 + I_{12}$ where

$$I_{12} = 4D^4(t) \int \begin{array}{l} F_2(\vec{q}_1, \vec{q}_2)\hat{W}(k_1R)\hat{W}(k_2R)\hat{W}[q_{12}R]\delta_D(\vec{q}_1 + \vec{k}_1)P(\vec{q}_1)\delta_D(\vec{q}_2 + \vec{k}_2)P(\vec{q}_2) \\ \times e^{i(\vec{k}_1\cdot\vec{x}_1 + \vec{k}_2\cdot\vec{x}_2 + \vec{q}_{12}\cdot\vec{x}_2)}\mathrm{d}^3\vec{k}_1\mathrm{d}^3\vec{k}_2\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2. \end{array}$$

Moreover, in I_0 the Dirac distribution imposes that $\vec{q}_1 = -\vec{q}_2$. Since the second order perturbation theory kernel can be expressed as

$$F_2\left(\vec{q}_1, \vec{q}_2\right) = \frac{1}{2}\left(1 + \frac{\vec{q}_1\cdot\vec{q}_2}{q_1^2}\right) + \frac{1}{2}\left(1 + \frac{\vec{q}_1\cdot\vec{q}_2}{q_2^2}\right) - \frac{2}{7}\left[1 - \left(\frac{\vec{q}_1\cdot\vec{q}_2}{q_1q_2}\right)^2\right], \tag{C.6}$$

it is obvious that $F_2(\vec{q}_1, -\vec{q}_1) = 0$, as a result $I_0 = 0$ thus

$$I = 4D^4(t)\int F_2(\vec{q}_1, \vec{q}_2)\hat{W}(q_1R)\hat{W}(q_2R)\hat{W}[q_{12}R]P(q_1)P(q_2)e^{i\vec{q}_1\cdot\vec{r}}\mathrm{d}^3\vec{q}_1\mathrm{d}^3\vec{q}_2,$$

where I used the fact that $\int \delta_D(\vec{q}_1 + \vec{k}_1)\hat{W}(k_1R)e^{i\vec{k}_1\cdot\vec{x}_1}\mathrm{d}^3\vec{k}_1 = \hat{W}(q_1R)e^{-i\vec{q}_1\cdot\vec{x}_1}$. According to (C.6) the expression I can be split into three contributions $I = I' + I'' + I'''$ where

$$I' = 2D^4(t)\int \mathrm{d}^3\vec{q}_1P(q_1)\hat{W}(q_1R)e^{i\vec{q}_1\cdot\vec{r}}\int \mathrm{d}^3\vec{q}_2P(q_2)\hat{W}(q_2R)\hat{W}[q_{12}R]\left(1 + \frac{\vec{q}_1\cdot\vec{q}_2}{q_1^2}\right),$$

$$I'' = 2D^4(t)\int \mathrm{d}^3\vec{q}_1P(q_1)\hat{W}(q_1R)e^{i\vec{q}_1\cdot\vec{r}}\int \mathrm{d}^3\vec{q}_2P(q_2)\hat{W}(q_2R)\hat{W}[q_{12}R]\left(1 + \frac{\vec{q}_1\cdot\vec{q}_2}{q_2^2}\right),$$

and

$$I''' = -\frac{8}{7}D^4(t)\int \mathrm{d}^3\vec{q}_1P(q_1)\hat{W}(q_1R)e^{i\vec{q}_1\cdot\vec{r}}\int \mathrm{d}^3\vec{q}_2P(q_2)\hat{W}(q_2R)\hat{W}[q_{12}R]\left[1 - \left(\frac{\vec{q}_1\cdot\vec{q}_2}{q_1q_2}\right)^2\right].$$

In spherical coordinates (when q_1 is chosen as the azimuthal axis) the vector $\vec{q}_2$ can be split

into its radial (q_2) and angular (θ_2, ϕ_2) parts and the properties of the spherical Top-Hat window (Bernardeau, 1994a) can be expressed as

$$\int \mathrm{d}^2\Omega_2 \hat{W}[q_{12}R] \left(1 + \frac{\vec{q}_1 \cdot \vec{q}_2}{q_1^2}\right) = 4\pi \hat{W}(q_1 R) j_0(q_2 R), \tag{C.7}$$

and

$$\int \mathrm{d}^2\Omega_2 \hat{W}[q_{12}R] \left[1 - \left(\frac{\vec{q}_1 \cdot \vec{q}_2}{q_1 q_2}\right)^2\right] = 4\pi \frac{2}{3} \hat{W}(q_1 R) \hat{W}(q_2 R), \tag{C.8}$$

where $\mathrm{d}^2\Omega = \sin(\theta_2)\mathrm{d}\theta_2\mathrm{d}\phi_2$ and j_0 is the spherical Bessel function of order zero ($j_0(x) \equiv \sin(x)/x$). Using the identity (C.7) one can express I' and I'' in a simpler form

$$I' = 2 \underbrace{D^2(t) \int \Delta_{q_1} \hat{W}^2(q_1 R) j_0(q_1 r) \mathrm{d}\ln q_1}_{=\xi_R(r)} \times D^2(t) \int \Delta_{q_2} \hat{W}(q_2 R) j_0(q_2 R) \mathrm{d}\ln q_2,$$

and

$$I'' = 2D^2(t) \int \Delta_{q_1} \hat{W}(q_1 R) j_0(q_1 R) j_0(q_1 r) \mathrm{d}\ln q_1 \times \underbrace{D^2(t) \int \Delta_{q_2} \hat{W}^2(q_2 R) \mathrm{d}\ln q_2}_{=\sigma_R^2},$$

where I have integrated over the angular part of the vector $\vec{q}_1$ and where $\Delta_q \equiv 4\pi q^3 P(q)$ is the dimensionless power spectrum. In order to simplify both expressions I' and I'' it is convenient to use the logarithmic derivative of the 2-point correlation $\xi_R(r)$

$$\beta_R(r) \equiv \frac{\mathrm{d}\ln \xi_R(r)}{\mathrm{d}\ln R}. \tag{C.9}$$

Using the definition (C.9) and the fact that $\frac{\mathrm{d}\hat{W}(x)}{\mathrm{d}\ln x} = 3[j_0(x) - \hat{W}(x))]$ one can show that

$$D^2(t) \int \Delta_{q_1} \hat{W}(q_1 R) j_0(q_1 R) j_0(q_1 r) \mathrm{d}\ln q_1 = \xi_R(r) \left[\frac{\beta_R(r)}{6} + 1\right].$$

Thus I' and I'' can be expressed respectively as

$$I' = \xi_R(r)\sigma_R^2 \left[2 + \frac{\gamma_R}{3}\right], \tag{C.10}$$

and

$$I'' = \xi_R(r)\sigma_R^2 \left[2 + \frac{\beta_R(r)}{3}\right], \tag{C.11}$$

where $\gamma_R \equiv \mathrm{d}\ln\sigma_R^2/\mathrm{d}\ln R = \beta_R(0)$. Then using the identity (C.8) one can express

$$I''' = -\frac{16}{21}\xi_R(r)\sigma_R^2, \tag{C.12}$$

and by combining equations (C.10), (C.11) and (C.12) we obtain

$$I = \xi_R(r)\sigma_R^2 \left(\frac{68}{21} + \frac{\gamma_R}{3} + \frac{\beta_R(r)}{3}\right).$$

It is not possible, however, to express the term J in equation (C.5) according to the procedure outlined above. This is due to the presence of a factor $e^{i\vec{q}_2\cdot\vec{r}}$ in the integral representation of J that makes impossible the use of the identities (C.7) and (C.8) when we integrate over $\vec{q}_2$. It is possible to integrate numerically the term J (see Bernardeau, 1996) and to show that in the large separation limit ($r \geq 3R$) it can be neglected. On the other hand, by assuming that in the large separation limit $\hat{W}(q_{12}R) \simeq \hat{W}(q_1R)\hat{W}(q_2R)$ (cf. Bernardeau, 1996) it is possible to express

$$J \simeq 2\left\{1 - \frac{\beta_R'(r)}{3}\Upsilon_R(r) - \frac{4}{21}\left[2 - \Upsilon_R(r)\right]\Upsilon_R(r)\right\}\xi_R(r)^2,$$

where $\beta_R'(r) = \mathrm{d}\ln\xi_R(r)/\mathrm{d}\ln r$ and $\Upsilon = \bar{\xi}_R(r)/\xi_R(r)$, and $\bar{\xi}_R(r) = 3\int_0^r r'^2\xi_R(r')\mathrm{d}r'/r^3$. The term J is proportional to the square of the 2-point correlation function whereas the term I is proportional to the variance times the 2-point correlation function. Since in the large separation limit we have that $\xi_R(r) \ll \sigma_R^2$, it follows that the term J can be neglected. Thus we finally obtain that

$$\langle\delta_R(\vec{x})\delta_R^2(\vec{x}+\vec{r})\rangle = \left(\frac{68}{21} + \frac{\gamma_R}{3} + \frac{\beta_R(r)}{3}\right)\xi_R(r)\sigma_R^2. \tag{C.13}$$

If the distance distance r is least greater or equal to $3R$, the third order 2-point cumulant is proportional to the product of the variance and of the 2-point correlation function. This coefficient of proportionality is the reduced correlator of order $C_{12,R}$. However when taking the limiting case $R \to 0$ (i.e. no smoothing is applied to the field) the expression of J is exact and the third order reduced correlator is expressed as

$$C_{12} = \frac{68}{21} + 2\left\{1 - \frac{\beta'(r)}{3}\Upsilon(r) - \frac{4}{21}\left[2 - \Upsilon(r)\right]\Upsilon(r)\right\}\eta(r),$$

where $\eta(r) \equiv \xi(r)/\sigma^2$. In practice, for a realistic power spectrum, and when the matter overdensity is not smoothed the clustering ratio $\eta(r)$ turns out to be zero, which leads to $C_{12} = 68/21$.

Appendix D

Moment generating function

Here I show how to construct the moment generating function. Let's consider an arbitrary function $M(t)$ that is continuous and differentiable, let's expand it in Taylor serie

$$M(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n M(t)}{dt^n} \right|_{t=0} t^n,$$

and let's define

$$\left. \frac{d^n M(t)}{dt^n} \right|_{t=0} \equiv \langle N^n \rangle.$$

As a result,

$$M(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \langle N^n \rangle t^n,$$

however

$$\langle N^n \rangle = \sum_{N=0}^{\infty} N^n P_N,$$

which leads to

$$M(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{N=0}^{\infty} N^n P_N t^n.$$

It follows that

$$M(t) = \sum_{N=0}^{\infty} P_N \sum_{n=0}^{\infty} \frac{1}{n!} (Nt)^n,$$

where we can idendity the Taylor expansion of the exponential

$$M(t) = \sum_{N=0}^{\infty} P_N e^{Nt}.$$

Which can be finally expressed as

$$M(t) = \langle e^{Nt} \rangle.$$

Appendix E

Generating function of factorial moments

Here I show how to construct the factorial generating function. Let's consider an arbitrary function $M(t)$ that is continuous and differentiable. Let's expand it in Taylor serie

$$F(t) = \sum_{n=0}^{\infty} \frac{1}{n!} \left. \frac{d^n F(t)}{dt^n} \right|_{t=0} t^n$$

where, by definition,

$$\left. \frac{d^n F(t)}{dt^n} \right|_{t=0} \equiv \langle (N)_f^n \rangle.$$

It follows

$$F(t) = \left\langle (1+t)^N \right\rangle,$$

and, as a result, I get

$$F(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \langle (N)_f^n \rangle.$$

However,

$$\langle (N)_f^n \rangle = \langle N(N-1)(N-2)...(N-n+1) \rangle = \left\langle \prod_{k=0}^{n-1} (N-k) \right\rangle,$$

which can be written as

$$\langle (N)_f^n \rangle = \sum_{N=0}^{\infty} P_N \prod_{k=0}^{n-1} (N-k).$$

By replacing it into the Taylor expansion, one obtains

$$F(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \sum_{N=0}^{\infty} P_N \prod_{k=0}^{n-1} (N-k),$$

and thus

$$F(t) = \sum_{N=0}^{\infty} P_N \sum_{n=0}^{\infty} \frac{t^n}{n!} \prod_{k=0}^{n-1} (N-k).$$

where we identify the Taylor expansion of $(1+t)^N$. As a result

$$F(t) = \sum_{N=0}^{\infty} P_N (1+t)^N = \left\langle (1+t)^N \right\rangle .$$

Appendix F

Higher order galaxy 2-point cumulant moments

In this appendix I present the 2−point cumulant moments of order (n, m) (up to order 7) of the galaxy density field as a function of the analogous statistics of the mass density field.

$$
\begin{aligned}
\kappa_{g,11} &= b_1{}^2\kappa_{11}+b_1{}^2\left(c_3+C_{12}c_2\right)\kappa_{11}\,\kappa_2+1/2\,b_1{}^2c_2{}^2\kappa_{11}{}^2\\
\kappa_{g,12} &= b_1{}^3\left(C_{12}+2\,c_2\right)\kappa_{11}\,\kappa_2+b_1{}^3\left(5/2\,c_3C_{12}+2\,c_2{}^2C_{12}+3\,c_3c_2+c_4+1/2\,c_2C_{22}+c_2{}^2S_3+c_3S_3+c_2C_{13}\right)\kappa_{11}\,\kappa_2{}^2\\
&\quad +b_1{}^3\left(c_2{}^3+1/2\,c_4+3\,c_2{}^2C_{12}+2\,c_3c_2+c_3C_{12}\right)\kappa_{11}{}^2\kappa_2+b_1{}^3c_3c_2\kappa_{11}{}^3\\
\kappa_{g,13} &= b_1{}^4\left(6\,c_2{}^2+C_{13}+3\,c_2S_3+3\,c_3+6\,c_2C_{12}\right)\kappa_{11}\,\kappa_2{}^2+b_1{}^4\left(3\,C_{12}c_2+6\,c_2{}^2\right)\kappa_{11}{}^2\kappa_2+b_1{}^4c_3\kappa_{11}{}^3+b_1{}^4\left(3/2\,c_2C_{14}\right.\\
&\quad +9/2\,C_{12}c_3S_3+12\,c_2{}^2c_3+6\,c_2{}^3S_3+\tfrac{45}{2}\,c_2c_3C_{12}+9\,c_2c_4+3/2\,c_3S_4+9/2\,c_4S_3+3/2\,c_5+15/2\,c_3{}^2+3\,c_2{}^2S_4\\
&\quad \left.+3\,c_2{}^2C_{22}+15/2\,c_2{}^2C_{12}S_3+9/2\,c_4C_{12}+6\,c_2{}^2C_{13}+\tfrac{45}{2}\,c_3S_3c_2+5\,c_3C_{13}+6\,c_2{}^3C_{12}+1/2\,c_2C_{23}\right)\kappa_{11}\,\kappa_2{}^3\\
&\quad +b_1{}^4\left(9/2\,c_2{}^2C_{12}{}^2+6\,c_2{}^2C_{13}+3/2\,c_3C_{22}+\tfrac{39}{2}\,c_2c_3C_{12}+3/2\,c_4C_{12}+9/2\,c_3S_3c_2+3\,c_2{}^4+18\,c_2{}^3C_{12}+3/2\,c_3C_{12}{}^2\right.\\
&\quad \left.+15/2\,c_2c_4+18\,c_2{}^2c_3+6\,c_2{}^3S_3\right)\kappa_{11}{}^2\kappa_2{}^2+b_1{}^4\left(9/2\,c_3{}^2+1/2\,c_5+3/2\,c_4C_{12}+9\,c_2c_3C_{12}+6\,c_2{}^2c_3\right)\kappa_{11}{}^3\kappa_2\\
&\quad +3/2\,b_1{}^4c_2c_4\kappa_{11}{}^4\\
\kappa_{g,22} &= b_1{}^4\left(4\,c_2C_{12}+4\,c_2{}^2+C_{22}\right)\kappa_{11}\,\kappa_2{}^2+b_1{}^4\left(4\,c_3+8\,C_{12}c_2+4\,c_2{}^2\right)\kappa_{11}{}^2\kappa_2+4\,b_1{}^4c_2{}^2\kappa_{11}{}^3+b_1{}^4\left(2\,c_2{}^2C_{22}\right.\\
&\quad +2\,c_2{}^2C_{12}S_3+4\,c_3C_{22}+4\,c_2{}^3S_3+2\,c_2C_{23}+12\,c_2c_3C_{12}+2\,C_{12}c_3S_3+8\,c_2{}^2c_3+4\,c_2{}^2C_{13}+4\,c_2c_4+4\,c_3S_3c_2\\
&\quad \left.+4\,c_2{}^3C_{12}+2\,c_4C_{12}\right)\kappa_{11}\,\kappa_2{}^3+b_1{}^4\left(2\,c_2{}^4+6\,c_3S_3c_2+8\,c_2{}^2C_{22}+2\,c_4S_3+4\,c_3C_{12}{}^2+10\,c_2{}^2C_{12}{}^2+8\,c_4C_{12}+10\,c_3{}^2\right.\\
&\quad \left.+2\,c_5+36\,c_2c_3C_{12}+8\,c_2{}^2c_3+4\,c_2{}^2C_{13}+12\,c_2{}^3C_{12}+4\,c_3C_{13}+6\,c_2c_4\right)\kappa_{11}{}^2\kappa_2{}^2+b_1{}^4\left(8\,c_2c_4+16\,c_2{}^3C_{12}\right.\\
&\quad \left.+20\,c_2c_3C_{12}+16\,c_2{}^2c_3\right)\kappa_{11}{}^3\kappa_2+b_1{}^4\left(2c_3{}^2+c_2{}^4+4\,c_2{}^2c_3\right)\kappa_{11}{}^4\\
\kappa_{g,14} &= b_1{}^5\left(C_{14}+24\,c_2{}^3+12\,c_3S_3+4\,c_4+12\,c_2C_{12}S_3+36\,c_2{}^2S_3+12\,c_2C_{13}+36\,c_2{}^2C_{12}+4\,c_2S_4+12\,c_3C_{12}\right.\\
&\quad \left.+36\,c_3c_2\right)\kappa_{11}\,\kappa_2{}^3+b_1{}^5\left(36\,c_2{}^2C_{12}+4\,c_2C_{13}+3\,c_2C_{12}{}^2+36\,c_2{}^3+12\,c_3c_2+12\,c_2{}^2S_3\right)\kappa_{11}{}^2\kappa_2{}^2+b_1{}^5\left(6\,c_3C_{12}\right.\\
&\quad \left.+12\,c_3c_2\right)\kappa_{11}{}^3\kappa_2+b_1{}^5c_4\kappa_{11}{}^4+b_1{}^5\left(42\,c_2c_3S_3{}^2+12\,c_3S_3C_{13}+8\,c_3C_{12}S_4+2\,c_6+90\,c_2{}^3C_{12}S_3+24\,c_2{}^2C_{13}S_3\right.\\
&\quad +20\,c_2{}^2C_{12}S_4+6\,c_2{}^2C_{22}S_3+180\,c_3c_2{}^2C_{12}+84\,c_2c_3C_{13}+282\,c_2{}^2c_3S_3+6\,c_2c_3C_{22}+62\,c_2c_3S_4+74\,c_2c_4C_{12}\\
&\quad +114\,c_2c_4S_3+24\,c_4S_3C_{12}+168\,c_2c_3S_3C_{12}+8\,c_5C_{12}+24\,c_2c_5+12\,c_5S_3+12\,c_4C_{13}+8\,c_4S_4+6\,c_4S_3{}^2\\
&\quad \left.+72\,c_4c_2{}^2+44\,c_3c_4+54\,c_3{}^2C_{12}+84\,c_3{}^2S_3+114\,c_3{}^2c_2+60\,c_2{}^3c_3+17/2\,c_3C_{14}+2\,c_3S_5+1/2\,c_2C_{2,4}+2\,c_2C_{1,5}\right.
\end{aligned}
$$

$$
\begin{aligned}
&+6\,{c_2}^2 S_5+18\,{c_2}^3 C_{22}+24\,{c_2}^3 {S_3}^2+6\,{c_2}^2 C_{23}+36\,{c_2}^3 C_{13}+36\,{c_2}^4 S_3+18\,{c_2}^2 C_{14}+24\,{c_2}^4 C_{12}+36\,{c_2}^3 S_4\big){\kappa_{11}}\,{\kappa_2}^4\\
&+{b_1}^5\big(2\,c_3 C_{12} C_{13}+3\,c_3 C_{22} C_{12}+12\,{c_2}^5+20\,{c_2}^2 C_{12} C_{13}+72\,{c_2}^3 C_{12} S_3+270\,c_3 {c_2}^2 C_{12}+42\,c_2 c_3 {C_{12}}^2+32\,c_2 c_3 C_{13}\\
&+156\,{c_2}^2 c_3 S_3+18\,c_2 c_3 C_{22}+8\,c_2 c_3 S_4+54\,c_2 c_4 C_{12}+30\,c_2 c_4 S_3+48\,c_2 c_3 S_3 C_{12}+8\,c_2 c_5+2\,c_4 C_{13}+3/2\,c_4 {C_{12}}^2\\
&90\,c_4 {c_2}^2+6\,c_3 c_4+12\,{c_3}^2 C_{12}+48\,{c_3}^2 c_2+144\,{c_2}^3 c_3+2\,c_3 C_{23}+72\,{c_2}^3 C_{13}+54\,{c_2}^3 {C_{12}}^2+72\,{c_2}^4 S_3+10\,{c_2}^2 C_{14}\\
&+108\,{c_2}^4 C_{12}+18\,{c_2}^3 S_4\big){\kappa_{11}}^2{\kappa_2}^3+{b_1}^5\big(108\,c_3 {c_2}^2 C_{12}+30\,c_2 c_3 {C_{12}}^2+20\,c_2 c_3 C_{13}+48\,{c_3}^2 C_{12}+12\,{c_3}^2 S_3\\
&+18\,c_2 c_4 C_{12}+6\,c_2 c_5+78\,{c_3}^2 c_2+12\,c_3 c_4+36\,{c_2}^3 c_3+24\,{c_2}^2 c_3 S_3+3\,c_4 C_{22}+3\,c_5 C_{12}+6\,c_4 {C_{12}}^2\big){\kappa_{11}}^3{\kappa_2}^2+\\
&{b_1}^5\big(2\,c_5 C_{12}+1/2\,c_6+20\,c_2 c_4 C_{12}+18\,c_4 {c_2}^2+8\,c_3 c_4\big){\kappa_{11}}^4\kappa_2+2\,{b_1}^5 c_2 c_5 {\kappa_{11}}^5\\
\kappa_{g,23} \;=\;& {b_1}^5\big(C_{23}+2\,c_2 C_{13}+6\,c_2 C_{22}+6\,{c_2}^2 S_3+3\,c_3 C_{12}+18\,{c_2}^2 C_{12}+6\,c_3 c_2+3\,c_2 C_{12} S_3+12\,{c_2}^3\big){\kappa_{11}}\,{\kappa_2}^3\\
&+{b_1}^5\big(6\,c_2 C_{22}+18\,c_3 C_{12}+12\,c_2 {C_{12}}^2+3\,c_3 S_3+6\,c_2 C_{13}+6\,{c_2}^2 S_3+18\,{c_2}^3+54\,{c_2}^2 C_{12}+3\,c_4+30\,c_3 c_2\big){\kappa_{11}}^2{\kappa_2}^2\\
&+{b_1}^5\big(18\,c_3 c_2+24\,{c_2}^3+2\,c_4+36\,{c_2}^2 C_{12}+6\,c_3 C_{12}\big){\kappa_{11}}^3\kappa_2+6\,{b_1}^5 c_2 c_3 {\kappa_{11}}^4+6\,{b_1}^5 {c_2}^3 {\kappa_{11}}^4+{b_1}^5\big(c_2 C_{33}\\
&+3\,c_2 c_3 {S_3}^2+c_3 S_3 C_{13}+3/2\,c_3 C_{12} S_4+27\,{c_2}^3 C_{12} S_3+4\,{c_2}^2 C_{13} S_3+3\,{c_2}^2 C_{12} S_4+6\,{c_2}^2 C_{22} S_3+72\,c_3 {c_2}^2 C_{12}\\
&14\,c_2 c_3 C_{13}+57\,{c_2}^2 c_3 S_3+30\,c_2 c_3 C_{22}+3\,c_2 c_3 S_4+24\,c_2 c_4 C_{12}+12\,c_2 c_4 S_3+9/2\,c_4 S_3 C_{12}+42\,c_2 c_3 S_3 C_{12}\\
&+9/2\,c_3 S_3 C_{22}+3/2\,c_5 C_{12}+3\,c_2 c_5+c_4 C_{13}+9/2\,c_4 C_{22}+24\,c_4 {c_2}^2+3\,c_3 c_4+12\,{c_3}^2 C_{12}+3\,{c_3}^2 S_3+18\,{c_3}^2 c_2+\\
&30\,{c_2}^3 c_3+13/2\,c_3 C_{23}+3/2\,c_2 C_{2,4}+9\,{c_2}^3 C_{22}+3\,{c_2}^3 {S_3}^2+13\,{c_2}^2 C_{23}+18\,{c_2}^3 C_{13}+18\,{c_2}^4 S_3+3\,{c_2}^2 C_{14}+\\
&12\,{c_2}^4 C_{12}+6\,{c_2}^3 S_4\big){\kappa_{11}}\,{\kappa_2}^4+{b_1}^5\big(12\,c_3 C_{12} C_{13}+6\,c_3 C_{22} C_{12}+3/2\,c_6+6\,{c_2}^5+30\,{c_2}^2 C_{12} C_{13}+42\,{c_2}^3 C_{12} S_3\\
&+279\,c_3 {c_2}^2 C_{12}+108\,c_2 c_3 {C_{12}}^2+72\,c_2 c_3 C_{13}+84\,{c_2}^2 c_3 S_3+45\,c_2 c_3 C_{22}+9\,c_2 c_3 S_4+108\,c_2 c_4 C_{12}+36\,c_2 c_4 S_3\\
&+12\,c_4 S_3 C_{12}+72\,c_2 c_3 S_3 C_{12}+27\,{c_2}^2 C_{22} C_{12}+12\,c_5 C_{12}+18\,c_2 c_5+9/2\,c_5 S_3+9\,c_4 C_{13}+3/2\,c_4 S_4+6\,c_4 C_{22}\\
&+15\,c_4 {C_{12}}^2+45\,c_4 {c_2}^2+33\,c_3 c_4+81\,{c_3}^2 C_{12}+33\,{c_3}^2 S_3+90\,{c_3}^2 c_2+48\,{c_2}^3 c_3+3\,c_3 C_{23}+3\,c_3 C_{14}+48\,{c_2}^3 C_{22}\\
&+18\,{c_2}^2 C_{23}+36\,{c_2}^3 C_{13}+69\,{c_2}^3 {C_{12}}^2+12\,{c_2}^4 S_3+6\,{c_2}^2 C_{14}+54\,{c_2}^4 C_{12}+3\,{c_2}^3 S_4\big){\kappa_{11}}^2{\kappa_2}^3+{b_1}^5\big(6\,c_5 C_{12}\\
&+6\,c_4 {C_{12}}^2+27\,c_2 c_3 C_{22}+13\,c_2 c_5+c_6+96\,c_2 c_4 C_{12}+99\,{c_2}^3 {C_{12}}^2+12\,{c_2}^4 S_3+102\,{c_2}^3 c_3+51\,{c_3}^2 C_{12}\\
&+22\,c_3 c_4+42\,{c_2}^2 c_3 S_3+360\,c_3 {c_2}^2 C_{12}+c_5 S_3+105\,c_2 c_3 {C_{12}}^2+96\,{c_2}^4 C_{12}+69\,{c_3}^2 c_2+3\,{c_3}^2 S_3+13\,c_2 c_4 S_3\\
&+78\,c_4 {c_2}^2+24\,{c_2}^3 C_{22}+30\,c_2 c_3 C_{13}+36\,{c_2}^3 C_{13}+3\,c_4 C_{13}\big){\kappa_{11}}^3{\kappa_2}^2+{b_1}^5\big(6\,{c_2}^5+21\,c_2 c_4 C_{12}\\
&+126\,c_3 {c_2}^2 C_{12}+54\,{c_3}^2 c_2+27\,{c_3}^2 C_{12}+27\,c_4 {c_2}^2+36\,{c_2}^4 C_{12}+6\,c_2 c_5+48\,{c_2}^3 c_3+9\,c_3 c_4\big){\kappa_{11}}^4\kappa_2\\
&+12\,{b_1}^5 {c_2}^3 c_3 {\kappa_{11}}^5+3\,{b_1}^5 c_3 c_4 {\kappa_{11}}^5+9\,{b_1}^5 {c_3}^2 c_2 {\kappa_{11}}^5+6\,{b_1}^5 {c_2}^2 c_4 {\kappa_{11}}^5
\end{aligned}
$$

List of publications

P1 - *Second-order matter fluctuations via higher order galaxy correlators*, J. Bel, C. Marinoni, 2012, **MNRAS**, 424, 971

P2 - *The scale of cosmic isotropy*, C. Marinoni, J. Bel, A. Buzzi, 2012, **JCAP**, 75, 512

P3 - *Determination of the abundance of cosmic matter via the cell count moments of the galaxy distribution*, J. Bel, C. Marinoni **Phys. Rev. Lett.** submitted

P4- *The power spectrum from the angular distribution of galaxies in the CFHTLS-Wide fields at redshift* ~ 0.7, Granett, B. R.; Guzzo, L.; Coupon, J.; Arnouts, S.; Hudelot, P.; Ilbert, O.; McCracken, H. J.; Mellier, Y.; Adami, C.; Bel, J.; Bolzonella, M.; Bottini, D.; Cappi, A.; Cucciati, O.; de la Torre, S.; Franzetti, P.; Fritz, A.; Garilli, B.; Iovino, A.; Krywult, J.; Le Brun, V.; Le Fevre, O.; Maccagni, D.; Malek, K.; Marulli, F.; Meneux, B.; Paioro, L.; Polletta, M.; Pollo, A.; Scodeggio, M.; Schlagenhaufer, H.; Tasca, L.; Tojeiro, R.; Vergani, D.; Zanichelli, A., 2012, **MNRAS**, 421, 451

P5 - *The VIMOS Public Extragalactic Redshift Survey (VIPERS): spectral classification through Principal Component Analysis*, Marchetti, A.; Granett, B. R.; Guzzo, L.; Fritz, A.; Garilli, B.; Scodeggio, M.; Abbas, U.; Adami, C.; Arnouts, S.; Bolzonella, M.; Bottini, D.; Cappi, A.; Coupon, J.; Cucciati, O.; De Lucia, G.; de la Torre, S.; Franzetti, P.; Fumana, M.; Ilbert, O.; Iovino, A.; Krywult, J.; Le Brun, V.; Le Fevre, O.; Maccagni, D.; Malek, K.; Marulli, F.; McCracken, H. J.; Meneux, B.; Paioro, L.; Polletta, M.; Pollo, A.; Schlagenhaufer, H.; Tasca, L.; Tojeiro, R.; Vergani, D.; Zanichelli, A.; Bel, J.; Bersanelli, M.; Blaizot, J.; Branchini, E.; Burden, A.; Davidzon, I.; Di Porto, C.; Guennou, L.; Marinoni, C.; Mellier, Y.; Moscardini, L.; Nichol, R. C.; Peacock, J. A.; Percival, W. J.; Phleps, S.; Schimd, C.; Wolk, M.; Zamorani, G., 2012, **MNRAS**, ArXiv : 1207.4374