UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA
Dipartimento di Scienza e Alta Tecnologia
Dottorato in Astronomia e Astrofisica
XXVII ciclo

Estimating the galaxy power spectrum of the VIPERS galaxy distribution

Stefano Rota

Relatore: Prof. Luigi Guzzo
Correlatore: Dr. Julien Bel, Dr. Ben Granett

Anno Accademico 2013-2014

Estimating the galaxy power spectrum of the VIPERS galaxy distribution

Abstract

The power spectrum is a statistic that describes the clustering of matter in Fourier space. The shape and amplitude of the power are directly predicted by theory and direct measurements of the power can constrain cosmological parameters.

The large size of recent surveys has opened new windows in the observation of secondary features of the galaxy power spectrum such as the Baryonic Acoustic Oscillations (BAO), which used as standard ruler can infer dark energy properties, and Redshift-Space Distortions (RSD), which thanks to the anisotropic observed clustering in redshift space can discriminate between different theories of gravity.

In this thesis, we apply the power spectrum statistic to the VIMOS Public Extragalactic Redshift Survey (VIPERS) distribution of galaxies in order to study the underlying clustering of matter.

VIPERS is an ongoing spectroscopic survey, composed of two fields, W1 and W4, with the aim of mapping the spatial distribution of galaxies within a large volume of the $z \sim 1$ Universe. The main properties of VIPERS are the high redshift range surveyed $0.4 < z < 1.2$ and the high sampling rate at those redshifts.

Its main goal is the measurement of the anisotropy of the galaxy clustering in redshift space in order to constrain the growth rate of structure at a mean redshift $\langle z \rangle \sim 0.8$, never reached by other surveys.

Even the measure of the galaxy power spectrum at these redshifts is something new. Our goal is to estimate the clustering of matter in Fourier space constraining the cosmological quantities that mainly affect the power spectrum shape such as the matter density and the baryonic fraction. The approach used to estimate the galaxy power spectrum is quite classical and has been used in past surveys such as the Two Degree Field Galaxy Redshift Survey (2dFGRS).

We describe the power spectrum estimator using a novel simulation approach called "Γ"-simulation, including the main properties of a real survey such as the cone-like geometry and the density gradient function of the redshift.

The main factor that alters our power spectrum measurement is the window function, due to the limited size of the sample, which is not negligible even for a box geometry. We show a method to correctly take into account the window function effects including it with a three dimensional convolution with the input model of the simulation.

To analyse the potential of the VIPERS sample for extracting cosmological information, we use a set of VIPERS mock catalogues drawn from the MultiDark simulation. We analyse all the selection function effects of the survey, such as the sampling strategy and the decreasing mean density, in our estimation and study their effect on the recovered power spectrum.

The VIPERS window function is problematic due to the small angular coverage, strongly suppressing the clustering in Fourier space at large scales and completely damping the BAO signal. We measure the window function of VIPERS which will be used to perform

a three dimensional convolution with the theoretical model generated from the tested cosmology.

RSD also determine the observed clustering and, in the simplest case, their impact could be divided in two regimes. Small scale effects due to the high non-linear velocity of galaxies inside virialised structures are corrected with a simple velocity dispersion model, an empirical prediction of the behaviour of the clustering in redshift space. Large scale effects due to the linear in-fall of galaxies in high density peaks, which increase the observed clustering of matter, are included with the Kaiser term dependent on the linear growth rate of matter. We include these two terms before applying the 3D convolution of the model with the window function.

Some tests are performed on VIPERS mock catalogues to recover the known cosmological information of the simulation with a chi-square technique and, in all cases, we were able to extract the input cosmology with lower systematics errors than statistical ones. Finally we apply this approach to the data dividing the full survey into four subsamples, two redshift bins, $0.6 < \mathrm{z}_1 < 0.9$ and $0.9 < \mathrm{z}_2 < 1.1$, for each field. Despite the different window functions in each subsample due to small differences in volume and in geometry, the four measurements are consistent with each other. This is due to the fact that, at large scales, statistical fluctuations related to cosmic variance are much bigger with respect to small differences in the window. Even the overall amplitude is similar between the two redshift bins because the lower amplitude expected for the higher redshift power spectrum (because structures were less compact in the past) is compensated by the higher bias factor.

We compare the measured power spectrum from data with some theoretical models in order to extract the baryonic fraction and the amount of matter density in the Universe. The probability contours display evidence of a well-known degeneracy previously observed in other surveys (such as 2dFGRS) between the two recovered parameters in the determination of the overall shape of the matter power spectrum. Increasing the matter density of the Universe in fact shifts the power spectrum toward higher modes enhancing the clustering at the scales sampled by VIPERS. In this case to compensate the increase of power, the good fit is given by models with high baryonic fraction necessary to suppress the power at those scales. Vice-versa a low-Ω_M Universe does not need a high baryonic fraction. This degeneracy is broken combining the two redshift bins with a joint likelihood.

Finally, we provide an estimation of the matter density fixing all the other cosmological parameters to the best known values. We evaluate the matter density to be $\Omega_\mathrm{M} = 0.272^{+0.027}_{-0.031}$ in perfect agreement with Planck (once we rescale the Hubble factor from the Hubble Space Telescope used for the estimation of Ω_M to be consistent with the Planck one) and with the matter density estimated from the VIPERS Public Data Release-1 measurements of the clustering ratio.

Contents

Chapter 1

Introduction

1.1 The background Universe

The Copernican principle states that the Universe is homogeneous and isotropic, i.e. it has the same properties at each position and in each direction we observe it. Under these assumptions, solving the Einstein's gravity equations with the Friedmann-Lemaitre-Robertson-Walker (FLRW) metric, one can obtain the Friedman equations, which govern the expansion or contraction of the Universe and its acceleration

$$\begin{aligned} \left(\frac{\dot{\mathrm{a}}}{\mathrm{a}}\right)^2 &= \frac{8\pi \mathrm{G}}{3}\rho - \frac{\mathrm{k}}{\mathrm{a}^2} \\ \frac{\ddot{\mathrm{a}}}{\mathrm{a}} &= -\frac{4\pi \mathrm{G}}{3}(\rho + 3\mathrm{p}) \ , \end{aligned} \tag{1.1}$$

where p and ρ are respectively the pressure and density of the main components of the Universe, $\mathrm{k} = 0, 1, -1$ for flat, close or open geometry and a(t) the scale factor dependent on time.
Galaxies are observed to have a recession velocity proportional to their distance from us, following the Hubble equation,

$$\mathrm{v} = \mathrm{H}_0 \, \mathrm{r} \ . \tag{1.2}$$

Since through the Copernican principle we are not in a preferred position in the Universe, we infer that galaxies are apparently receding from each other meaning that the Universe is expanding. The Hubble parameter H, that at the current epoch coincides with the Hubble constant H_0, is defined as $\mathrm{H} = \dot{\mathrm{a}}/\mathrm{a}$ and parameterises the expansion rate of the Universe. The scale factor a(t) describes the dynamics of the Universe and its time dependence depends on the main density energies. Densities ρ_i are expressed with respect to $\Omega_\mathrm{i} = \rho_\mathrm{i}/\rho_\mathrm{c}$, where ρ_c, is the critical density defined as $\rho_\mathrm{c} = 3\,\mathrm{H}^2/8\pi\mathrm{G}$. Assuming that pressure and energy density of each component are related by an equation of state $\mathrm{p} = \mathrm{w}\rho$ parameterised by w, the evolution of densities follows $\rho_\mathrm{i} \propto \mathrm{a}^{-3(1+\mathrm{w}_\mathrm{i})}$.

Over the last 20 years, the ΛCDM scenario emerged as the standard cosmological model supported by independent observations as Cosmic Microwave Background (CMB) anisotropy (Hinshaw et al., 2013; Planck Collaboration et al., 2013), large-scale structures (Bernardeau et al., 2002) and the accelerating expansion of the Universe inferred with distant supernovae (Riess et al., 1998; Perlmutter et al., 1999).
In this picture, the main component of the matter density of the Universe is the Cold Dark Matter (CDM), which contributes to $\sim 25\,\%$ of the total energy density and have a determinant role in structure formation. Dark matter does not emit nor absorb light or other electromagnetic radiation at any significant level and its existence has been predicted to explain some gravitational effects in rotation curves of disc galaxies (Rubin and Ford, 1970), dynamics of the hot gas in clusters (Zwicky, 1937) and weak lensing (Bartelmann and Schneider, 2001).
The standard matter, baryons, is not sufficient to explain all these observations due to its lower density $\sim 5\%$. Since overall matter, dark matter and baryons, is composed by non-relativistic particles with $\mathrm{w} = 0$, the evolution of its energy density is $\rho_{\mathrm{M}} \propto \mathrm{a}^{-3}$ meaning that, due to the expansion of the Universe, it was higher in the past when the Universe was smaller.
Another contribution to the total energy density is given by radiation, composed of relativistic particles and photons with a non-null pressure $\mathrm{w} = 1/3$. Its energy density scales, differently with respect to the matter case, as $\rho_{\mathrm{R}} \propto \mathrm{a}^{-4}$: given this different dependence on a, although radiation is today negligible ($\Omega_{\mathrm{R}} \sim 10^{-4}$), there was a time, called the radiation era, when it was dominant.
The present day relation between the total density and the curvature is $\Omega_{\mathrm{tot}} - 1 = \frac{\mathrm{k}}{\mathrm{H}_0^2}$. Due to the discovery and analysis of the CMB, relic of the initial Universe, we now know that the Universe is not far from being flat, i.e. $\mathrm{k} \approx 0$, implying $\Omega_{\mathrm{tot}} \sim 1$. Summing the radiation and matter contributions to the total energy density, we obtain a value of the order of $\sim 30\%$ meaning that we are missing the majority of the density contributing to the geometry.

In the last decade of the 20th century, it became evident that dark and ordinary matter were insufficient to describe accurately a variety of cosmological observations. It was at the end of that decade that the relation between luminosity and distance of Type 1a supernovae (SNIa) revealed that in the Universe about 70% of the total energy density comes from an extra component, which dominates the evolution of the Universe causing the cosmic expansion to accelerate. The simplest model for dark energy is a cosmological constant Λ with $\mathrm{w} = -1$ whose density is constant in time. This means that the dark energy component was negligible in the past where the dynamics of the Universe was dominated by radiation and then by matter.

The study of the clustering of matter is important to infer cosmological quantities and analyse the evolution of the Universe. To reconstruct the underlying distribution of matter, we study the distribution of galaxies, which interact gravitationally with CDM. The distance of a galaxy can be measured from its redshift, $\mathrm{z} = \lambda_{\mathrm{obs}}/\lambda_{\mathrm{em}} - 1$: a photon emitted with wavelength λ_{em} will be observed with wavelength λ_{obs}, shifted toward redder region of the electromagnetic spectrum due to the stretching of space caused by the expansion of the Universe. Since the wavelength of the light emitted by

the galaxy scales as $\lambda \propto \mathrm{a(t)}$, the redshift could be related directly to the scale factor of the Universe at the time $\mathrm{t_{em}}$ when the light was emitted $\mathrm{z = a(t_{obs})/a(t_{em}) - 1}$.
Thanks to this, the observed redshift of a galaxy, once we assume a cosmology, can be associated to its comoving distance. Measuring the redshift of a large number of objects allows us to reconstruct the three dimensional distribution of galaxies, which could be associated to the clustering of matter predicted by theory. Statistical tools are needed to describe the clustering of galaxies (and matter) such as the power spectrum and the correlation function.

1.2 Power spectrum statistics

Present-day structures of matter have formed by the gravitational amplification of small density perturbations. These primordial perturbations are thought to be quantum fluctuations in the early Universe that have been brought to cosmological scales by inflation (Guth, 1981). Under the effect of gravity, density perturbations grow during the evolution of the Universe forming the structures we observe today in the cosmic web.
The density contrast which measures under- and over- densities with respect to the mean density of the Universe $\bar{\rho}$ at the position $\mathbf{x}$ is defined as

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \; . \tag{1.3}$$

Assuming the field $\delta(\mathbf{x})$ to follow a Gaussian probability distribution, at least at early times, the statistical distribution of matter is completely described by the two-point correlation function of this field. By definition, the mean value of the matter density field is $\langle \delta(\mathbf{x}) \rangle \equiv 0$. The covariance in configuration space is called "correlation function"

$$\xi(\mathbf{x}_1, \mathbf{x}_2) \equiv \langle \delta(\mathbf{x}_1)\, \delta(\mathbf{x}_2) \rangle \; . \tag{1.4}$$

The cosmological principle states that the Universe, on large enough scales, is homogeneous and isotropic being the properties of the Universe the same for all observers. From statistical homogeneity and isotropy, we have that

$$\xi(\mathbf{x}_1, \mathbf{x}_2) = \xi(\mathbf{x}_1 - \mathbf{x}_2) = \xi(|\mathbf{x}_1 - \mathbf{x}_2|) \; . \tag{1.5}$$

The correlation function, in terms of a discrete process, can be better understood defining it as the measure of the excess probability, relative to a Poisson distribution, of finding two galaxies within two volume elements $\mathrm{dV_1}$ and $\mathrm{dV_2}$ separated by a distance r

$$\mathrm{dP_{12} = \bar{n}[1 + \xi(r)]dV_1 dV_2} \; , \tag{1.6}$$

where $\bar{\mathrm{n}}$ is the mean number density of galaxies. Therefore, $\xi(\mathrm{r})$ measures the clustering in excess ($\xi(\mathrm{r}) > 0$) or in defect ($\xi(\mathrm{r}) < 0$) compared with a random distribution, for which $\xi(\mathrm{r}) = 0$. Another way to study statistically the distribution of matter is in Fourier space, analysing the evolution of different wave-modes. The two-point correlation function of the density field in Fourier space is defined as

$$\mathrm{P}(\mathbf{k}_1, \mathbf{k}_2) \equiv \langle \delta(\mathbf{k}_1)\, \delta^*(\mathbf{k}_2) \rangle \; . \tag{1.7}$$

We define the Fourier Transform of a function f(x) from configuration space to the Fourier one (and vice-versa) as

$$\begin{aligned} \mathrm{f(x)} &= \int \mathrm{f(k)\, e^{-ik\cdot x} \frac{dk}{2\pi}} \\ \mathrm{f(k)} &= \int \mathrm{f(x)\, e^{ik\cdot x}\, dx}\ . \end{aligned} \tag{1.8}$$

Applying the 3D conversion of the density contrast in Fourier space $\delta(\mathbf{k}) = \int \delta(\mathbf{x})\, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\, \mathrm{d}^3\mathrm{x}$ to 1.7, we obtain

$$\begin{aligned} &\int \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\rangle \mathrm{e}^{\mathrm{i}(\mathbf{k}_1\cdot\mathbf{x}_1 - \mathbf{k}_2\cdot\mathbf{x}_2)}\, \mathrm{d}^3\mathrm{x}_1\, \mathrm{d}^3\mathrm{x}_2 \xrightarrow{\mathbf{x}_2=\mathbf{x}_1+\mathbf{r}} \\ \int \mathrm{e}^{-\mathrm{i}(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{x}_1} &\int \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_1+\mathbf{r})\rangle \mathrm{e}^{-\mathrm{i}\mathbf{k}_2\cdot\mathbf{r}}\, \mathrm{d}^3\mathrm{r}\, \mathrm{d}^3\mathrm{x}_1 \xrightarrow{\text{homogeneity}} \\ &\int \xi(\mathbf{r}) \mathrm{e}^{-\mathrm{i}\mathbf{k_2}\cdot\mathbf{r}}\, \delta^{\mathrm{D}}(\mathbf{k}_2 - \mathbf{k}_1)(2\pi)^3\, \mathrm{d}^3\mathrm{r} \to \\ &\mathrm{P}(\mathbf{k}_2)\, \delta^{\mathrm{D}}(\mathbf{k}_2 - \mathbf{k}_1)(2\pi)^3\ , \end{aligned} \tag{1.9}$$

where, in the last equality, we define the power spectrum as the Fourier counterpart of the correlation function. δ_{D} is the Dirac delta function, showing that modes are uncorrelated in Fourier space. Given the isotropy of the Universe, the power spectrum depends only on the modulus of the vector **k** and not on its direction,

$$\langle \delta(\mathbf{k}_1)\delta^*(\mathbf{k}_2)\rangle = (2\pi)^3\, \delta^{\mathrm{D}}(\mathbf{k}_1 - \mathbf{k}_2)\, \mathrm{P}(|\mathbf{k}_1|)\ , \tag{1.10}$$

The correlation function and the power spectrum encode the same information being a Fourier pair:

$$\begin{aligned} \xi(\mathrm{r}) &= \int \mathrm{P(k)\, e^{-i\mathbf{k}\cdot\mathbf{x}} \frac{d^3k}{(2\pi)^3}} \\ \mathrm{P(k)} &= \int \xi(\mathrm{r})\, \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\, \mathrm{d}^3\mathrm{x}\ . \end{aligned} \tag{1.11}$$

One of the advantage of using the power spectrum analysis, in principle, is that, from equation 1.10, modes are uncorrelated leading to a diagonal covariance matrix. In practice this is not true for real surveys where the limited size of the sample leads to a mixing of modes due to a window function effect. Further, thank to the use of the FFT algorithm, the estimation of the power spectrum from a galaxy sample is faster than the one of the correlation function. The importance of the power spectrum is due to the fact that the linear perturbation theory predicts directly its shape: it encodes information about the formation of the primordial fluctuations, and especially about how these are modified according to the matter content of the Universe.

1.3 Cosmology with P(k)

In the ΛCDM standard cosmology, the dark energy component dominates the dynamic of the Universe only at late epochs when structures have already formed. During the matter dominated era, the high energy density of CDM with respect to the other energy components makes it the major responsible for the evolution of fluctuations at different scales modelling the shape of the power spectrum.

Although baryons contribute to only 5 % of the total energy density, their effect in the evolution of structures is not negligible leaving some features in the power spectrum. The concordance ΛCDM cosmological model is described by six free parameters, which completely determine the shape and amplitude of the matter power spectrum at all epochs: the Hubble parameter today, $\mathrm{H_0}$ (with h the dimensionless Hubble constant, $\mathrm{h = H_0/100\,[km\,s^{-1}Mpc^{-1}]}$), the physical baryon density, $\Omega_\mathrm{B}\mathrm{h}^2$, the total matter density, $\Omega_\mathrm{M}\mathrm{h}^2$, the optical depth to reionization, τ, the amplitude of the primordial scalar density perturbations on scales of $\mathrm{k_0 = 0.05\,h\,Mpc^{-1}}$, $\mathrm{A_s}$, and the spectral index of the primordial power spectrum of density perturbations, $\mathrm{n_s}$. Current analyses are trying to refine this vision of the Universe reducing the statistical errors on cosmological parameters. Extensions of this cosmological model include extra parameters as curvature of the Universe Ω_k, an equation of state of the dark energy, w different from $\mathrm{w} = -1$, the mass of the neutrinos, $\sum \mathrm{m}_\nu$, the scale dependence of the spectral index of the primordial power spectrum, $\mathrm{n_{run}}(= \frac{\mathrm{d\ln ns}}{\mathrm{d\ln k}})$, and the amplitude of the primordial tensor perturbations, r.

1.3.1 Evolution of P(k)

Structures we observe today have formed by gravitational amplification of small primordial inhomogeneities. Many inflationary models predict that the primordial power spectrum $\mathrm{P_{prim}}$ obeys a power law with amplitude $\mathrm{A_s}$ and index $\mathrm{n_s}$

$$\mathrm{P_{prim}(k) \propto A_s\, k^{n_s}} \; . \tag{1.12}$$

Harrison and Zel'dovich predicted a $\mathrm{n_s} = 1$ primordial spectrum assuming that initial fluctuations of the gravitational potential are scale invariant. Today the overall shape of the power spectrum is very different from a power-law because of the scale-dependent evolution of fluctuations.

By structure formation we mean the generation and evolution of the primordial inhomogeneities. Structure formation uses a perturbative approach to describe the evolution of density fluctuations during the various epochs of the Universe.

Scales which are close to horizon size or super-horizon require relativistic perturbation theory, which is based on General Relativity because we can no more ignore the curvature of space-time. We just summarise the results of this approach saying that super-horizon scale fluctuations grow proportionally to the scale factor in the matter dominated era $\delta_\mathbf{k} \propto \mathrm{a}$ and proportionally to the square of the scale factor in the radiation dominated era $\delta_\mathbf{k} \propto \mathrm{a}^2$, with $\delta_\mathbf{k}$ the density contrast $\delta = \delta\rho/\bar{\rho}$ in Fourier space.

Fluctuations at sub-horizon scales can be treated with Newtonian perturbation theory combining the main equations of fluid dynamics (for a perfect fluid) with gravity and adding, to the background solutions, small linear perturbations as

$$\rho_\mathrm{M}(\mathrm{t}, \mathbf{x}) = \bar{\rho}(\mathrm{t}) + \delta\rho(\mathrm{t}, \mathbf{x}) \tag{1.13}$$

for the matter density. Considering adiabatic perturbations in a fluid and Fourier expanding the perturbations in terms of plane waves which evolve independently, one can obtain the final following equation

$$\ddot{\delta}_{\mathbf{k}} + 2\mathrm{H}\dot{\delta}_{\mathbf{k}} + \left[\frac{\mathrm{c}_\mathrm{s}^2\mathrm{k}^2}{\mathrm{a}^2} - 4\pi\mathrm{G}\bar{\rho}\right]\delta_{\mathbf{k}} = 0\ , \tag{1.14}$$

function of the fluctuations $\delta_{\mathbf{k}} = \delta\rho_{\mathbf{k}}/\bar{\rho}$ in Fourier space, with $\mathrm{c}_\mathrm{s} = \partial\mathrm{p}/\partial\rho$ the sound speed and the dot standing for time derivative. The equation is solved by the time evolution of the Fourier amplitudes $\delta_{\mathbf{k}}$ of the perturbation after finding the background solution for the functions $\mathrm{a(t)}$, $\mathrm{H} = \dot{\mathrm{a}}/\mathrm{a}$ and $\bar{\rho}(\mathrm{t})$. The term $2\mathrm{H}\dot{\delta}_{\mathbf{k}}$, which accounts for the expansion of space, is a sort a friction factor and tends to limit the growth of fluctuations. The nature of the solution strongly depends on the sign of the term inside the brackets where the intermediate scale, called the Jeans wave number (Jeans found for a static fluid the same equation but without the term $2\mathrm{H}\dot{\delta}_{\mathbf{k}}$), between the two regimes is

$$\mathrm{k_J} = \mathrm{a}\frac{\sqrt{4\pi\mathrm{G}\bar{\rho}}}{\mathrm{c_s}}\ . \tag{1.15}$$

The first term inside brackets in equation 1.14 is related to the pressure gradients that tend to contrast the compression due to the gravity (second term) generated by density excess of the fluid $\delta_{\mathbf{k}}$. If pressure dominates, for fluctuations with scale $\mathrm{k} > \mathrm{k_J}$, we should expect to recover oscillating solution, which stands for sound waves in the density field, while if gravity dominates the small initial perturbation with $\mathrm{k} < \mathrm{k_J}$ grows leading to more compact structures.

During the evolution of the Universe, perturbations with different scales enters the horizon and the Jeans scale in different epochs when the background fluid density was dominated by radiation or by matter. Following the evolution of each single mode in the density field, we can have an idea of how the amplitude of fluctuations has grown during the evolution of the Universe and predict the shape of the matter power spectrum in the linear regime.

CDM perturbations

Cold dark matter is believed to be the dominant component in the formation of structures. Baryons and hot dark matter (neutrinos), in fact, are too few respect to CDM to have a determinant role, radiation with its high pressure does not form structures and the dark energy, which dominates in the energy density today, is not far from being homogeneous. Since CDM is pressure-less ($\mathrm{c_s} = 0$), the growth equation (1.14) becomes

$$\ddot{\delta}_{\mathbf{k}} + 2\mathrm{H}\dot{\delta}_{\mathbf{k}} + 4\pi\mathrm{G}\bar{\rho}_\mathrm{M}\delta_{\mathbf{k}} = 0\ , \tag{1.16}$$

meaning that CDM fluctuations, which are always super-Jeans scale perturbations, do not oscillate.

The horizon scale λ_H grows with time meaning that larger-scale fluctuations enter the horizon at later time. The time of equality between radiation and matter, $\mathrm{t_{eq}}$, is determinant for the evolution of these density fluctuations fixing a particular scale $\lambda_\mathrm{H}(\mathrm{t_{eq}})$, being the size of the horizon at the time of equality. Fluctuations smaller than $\lambda_{\rm H}(\rm t_{eq})$ enter the horizon during the radiation dominated era at $\rm t < t_{eq}$ when the expansion of the Universe is $\rm H^2 \propto \bar{\rho}_R \propto a^{-4}$, while larger ones enter during the matter dominated $\rm t > t_{eq}$ era when $\rm H^2 \propto \bar{\rho}_M \propto a^{-3}$. Replacing the proper Hubble factor in equation 1.16 and solving for the amplitude of fluctuations with respect to time, we obtain that scale CDM fluctuations smaller than $\lambda_{\rm H}(\rm t_{eq})$, that have entered the horizon in the radiation epoch, evolve with

$$\delta(\rm t) = \delta_{\rm initial}\left(1 + \frac{3}{2}\frac{\rm a}{\rm a_{eq}}\right) , \tag{1.17}$$

while scale fluctuations larger than $\lambda_{\rm H}(\rm t_{eq})$, that enter during the matter era, grow as

$$\delta(\rm a) \propto a \propto t^{2/3} . \tag{1.18}$$

When the Universe becomes dark-energy dominated, $\rm H^2 \propto$ const and CDM perturbations stop growing.

From this simple description, we can explain the overall shape of today power spectrum, composed by two regimes separated by the peak of the power at frequency $\rm k = k_{eq}$, which is the mode associated to horizon at the time of equality, $\rm k_{eq} = \frac{2\pi}{\lambda_H(t_{eq})}$. The largest scale structures observe today in the cosmic web are in fact the consequence of the initial perturbations that, once have entered the horizon in the matter era, have continued to grow as when they were super-horizon fluctuations. The power at these scales has the same shape of the primordial one $\rm P(k) \propto k^{n_s}$. Smaller-scale structures, instead, when entered the horizon in the radiation dominated era, remain more or less frozen since have growth by only a fraction of $\frac{5}{3}$ until the time of equality and start to grow during the matter epoch. Since the growth of super-horizon fluctuations is proportional to $\rm a^2$, the ratio at $\rm t_{eq}$ between the amplitude of a fluctuation that has entered the horizon in the radiation epoch and one still growing on super-horizon scales is $\rm R \propto \left(\frac{a_H}{a_eq}\right)^2$ with $\rm a_H$ the scale factor at the time of entering the horizon. Therefore knowing that $\rm k \propto a_H^{-1}$, we can predict the shape of the power for $\rm k > k_{eq}$

$$\rm P(k) \propto k^{n_s} \times R^2 \propto k^{ns-4} . \tag{1.19}$$

The overall shape of the initial power spectrum, after $\rm t_{eq}$, is shown in figure 1.1 where the largest scales ($\rm k << k_{eq}$) behave like a power-law function with index $\rm n_s$ while small scales, beyond the peak ($\rm k >> k_{eq}$), decrease following a power law with index $\rm n_s - 4$.

Baryonic perturbations

Although CDM is dominant in structure formation, the baryonic component is not negligible and most importantly what we can directly observe is the clustering of baryons within the form of galaxies. In this context, we need to study how perturbations in the baryonic component evolve. The time of decoupling $\rm t_{dec}$ is determinant in the evolution of baryonic fluctuations as the time of equality is for CDM structures. Before $\rm t_{dec}$, baryons and photons were coupled forming a fluid characterised by the radiation pressure. Due to the high pressure of this fluid, baryonic perturbations cannot grow until the time of decoupling. The Jeans length, defined as the inverse of the Jeans mode $\lambda_J \sim k_J^{-1}$, is in fact comparable to the horizon scale. All baryonic fluctuations that enter the horizon before t_{dec} are smaller than λ_J and consequently evolve as an oscillating wave in the baryon-photon fluid. These oscillations, known as Baryonic Acoustic Oscillations (BAO) leave a secondary feature in the matter power spectrum. Perturbations in the matter density before the time of decoupling modify locally the temperature of photons $\langle\frac{\delta\rho}{\rho}\rangle \sim \langle\frac{\delta T}{T}\rangle$: these oscillating waves leave a strong imprint in the power spectrum of CMB anisotropies, which has been observed extremely well by Larson et al. (2011) and Planck Collaboration et al. (2013). After decoupling, the baryon component composed of hydrogen and helium sees just its own pressure, which is the one of a monatomic gas. The baryon Jeans length drops after decoupling letting the baryonic fluctuations to grow as $\delta_B \propto a$. Photons are no more affected by the presence of baryons and the study of CMB anisotropies gives us direct information on fluctuations at the time of decoupling. If we measure the growth of baryonic structure from the time of decoupling up to now, we obtain $a_0/a_{dec} = 1 + z_{dec} \sim 1100$ meaning that baryonic structures we observe today should have a density bigger than a factor of 10^3 respect to anisotropies observed in the CMB, 10^{-5}. The density of galaxies, gas clouds and other baryonic structures in the Universe correspond to a density contrast of the order of $\delta\rho_B/\bar{\rho} \sim 10 - 100 - 1000$, far from the predicted $10^{-5} \times 10^3 = 10^{-2}$. This discrepancy is one of the most accredited reasons for the existence of dark matter. CDM structures begin in fact to grow earlier at the time of equality, $t_{eq} < t_{dec}$. At the time of decoupling, baryons, free to evolve without the support from photon pressure, fall into the CDM gravitational potential wells, reaching in a short time the density contrast of CDM. This explains why the observed density contrast in baryonic structures today is so high.

To account for all the changes in the matter power spectrum shape with respect to the primordial one in the matter dominated era (when perturbations at all scales grow at the same rate, independently of time), an analytical solution, called transfer function T(k) and based on various assumptions, is introduced, such that

$$P(k,z) \; = \; T^2(k)\; D^2(z)\; P_{prim}(k)\;, \tag{1.20}$$

where P(k) is the matter power spectrum measured today, $P_{prim}(k)$ the primordial power of equation 1.12 and D(z) the growth factor which describes the growth of fluctuations at different k-modes. As mentioned before, the asymptotic behaviour of the transfer function is

$$\begin{aligned} T(k) = 1 \quad & \text{for} \quad k << k_{eq} \\ T(k) \propto k^{-2} \quad & \text{for} \quad k >> k_{eq}\;. \end{aligned} \tag{1.21}$$

Detailed evaluations of the transfer function have been carried out by a number of authors. The two most important achievements in evaluating an analytical expression for the transfer function are Bardeen et al. (1986),

$$T(k) = \frac{\ln(1+2.34q)}{2.34q}[1 + 3.89q + (16.1q)^2 + (5.46q)^3 + (6.71q)^4]^{-1/4}\;, \tag{1.22}$$

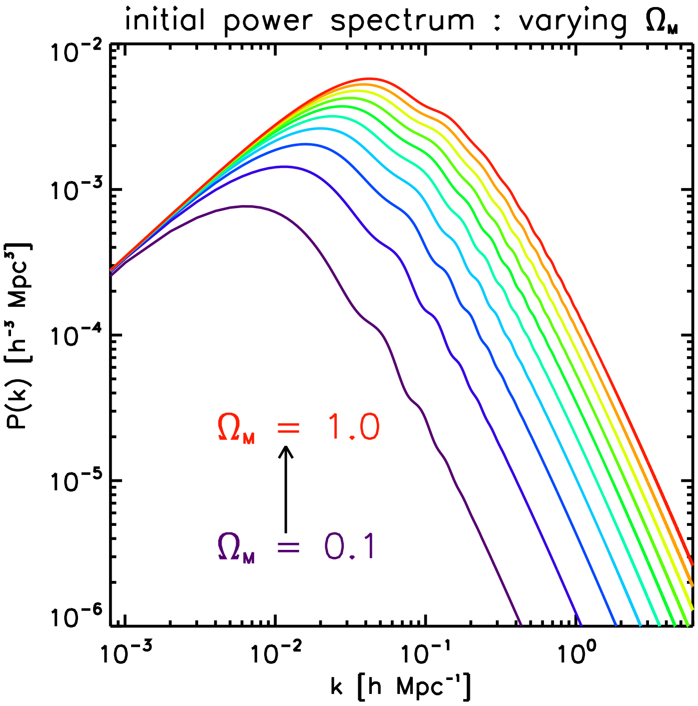

Figure 1.1: theoretical model power spectra obtained using the transfer function of Eisenstein and Hu (1998) and a power-law primordial power spectrum, varying the matter density. The cosmological parameters are fixed to $n_s = 0.95$, $h = 0.70$, $\Omega_k = 0$, $f_B = \Omega_B/\Omega_M = 20\%$ expect for Ω_M which is set as free with values ranging from $\Omega_M = 0.1$ to $\Omega_M = 1.0$ with step $\Delta\Omega_M = 0.1$.

where $q = k/(\Omega_M h^2 Mpc^{-1})$, for the CDM model alone and the work done by Eisenstein and Hu (1998) extended to include baryonic effects.
The initial power spectra $P(k) \propto T^2(k) \times P_{prim}(k)$ plotted in figures 1.1, 1.2, 1.3 are computed with the approximated transfer function calculated by Eisenstein and Hu (1998) for different cosmological parameters. A detailed calculation including all relevant effects could not be done analytically. There are publicly available computer codes (such as the one used in this thesis, Code for Anisotropies in the Microwave Background (CAMB, Lewis and Challinor, 2011)) that evaluate numerically the matter power spectrum setting as input the cosmological parameters. The main approximation of this calculation of the power spectrum (in the analytic solution or in the numerical one) is the linear perturbation theory which assumes that fluctuations in the density are small enough to have $\delta\rho << \bar{\rho}$. This is quite correct for large scale structure which are still in the linear regime $\delta < 1$. For smaller scales, non-linear effects, as the high velocities of galaxies inside clusters and groups $\delta v \sim v$ or the high density, must be taken into account. Numerous efforts have been done to include non-linearities in the model us-

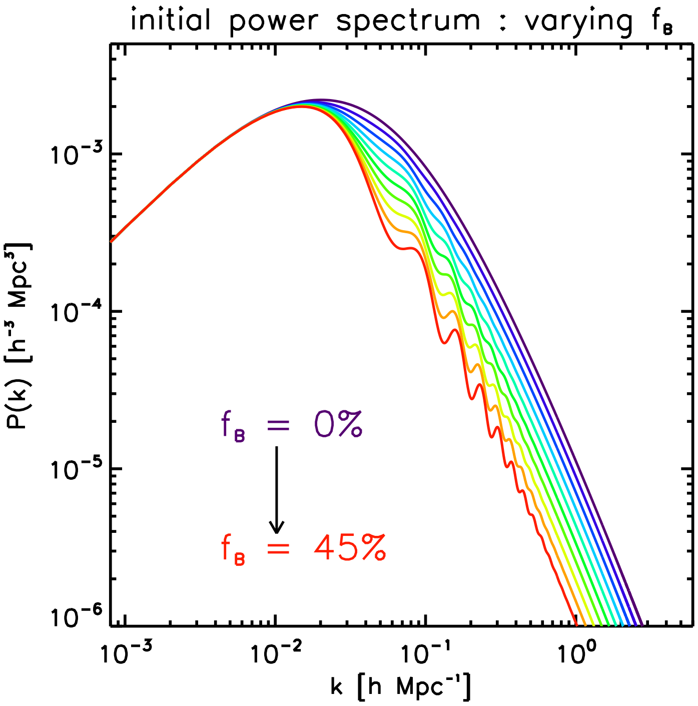

Figure 1.2: theoretical model power spectra obtained using the transfer function of Eisenstein and Hu (1998) and a power-law primordial power spectrum, varying the baryonic fraction. The cosmological parameters are fixed to the same values of figure 1.1 expect for $\Omega_M = 0.30$ which is now fixed and the baryonic fraction which is set as free with values ranging from $f_B = 0\%$ to $f_B = 45\%$ with step $\Delta f_B = 5\%$. Notice the increase of the BAO signal associated to the increase of the baryonic fraction.

ing huge N-body simulations. The code allows the user to select an option to obtain an approximated non-linear power spectrum, based on the HALOFIT prescriptions described in Smith et al. (2003). Figure 1.3 shows the comparison between the linear power spectrum calculated with the Eisenstein and Hu (1998) formulas and the non-linear correction based on the Smith et al. (2003) work at different redshifts. Notice that non-linear and linear power spectra are more consistent with each other at higher redshift due to the lower clustering of structures resulting in a lower effect of non-linearities. The overall amplitude of the power spectrum changes due to the different growth factor D(z) of structures at different epochs.

1.3.2 Shape and amplitude of P(k)

Before comparing the measured power spectrum with a theoretical prediction, we need to introduce the bias factor due to the relation between the clustering of matter predicted

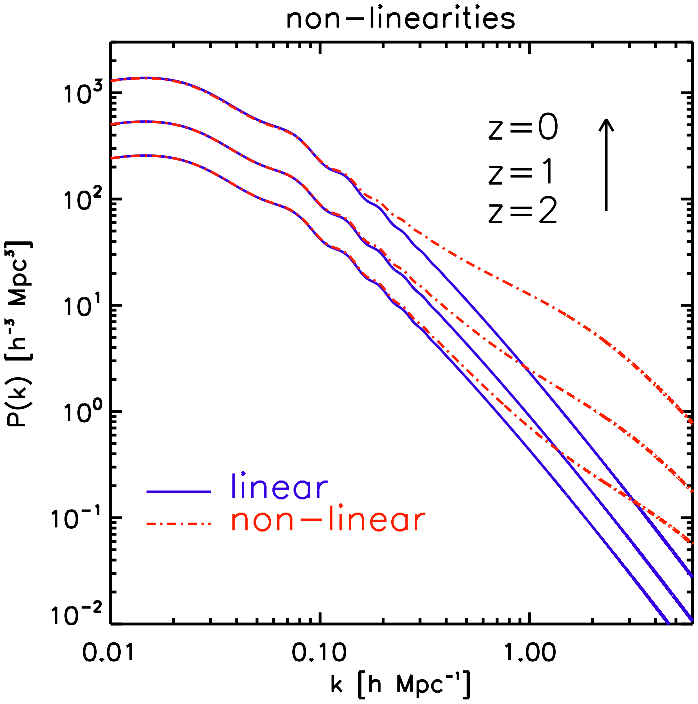

Figure 1.3: theoretical model power spectra at different redshifts ($z = 0$, $z = 1$ and $z = 2$) obtained using the transfer function of Eisenstein and Hu (1998) and a power-law primordial power spectrum (blue solid lines) showing the behaviour of linear theory compared to power spectra obtained in the same way but with the inclusion of the non-linear correction of Smith et al. (2003) (red dashed lines).

by theory and the observed galaxy clustering. To map the underlying dark matter density field, we use visible tracers as galaxies. A general way of interfacing matter and galaxy fluctuations has been introduced by Fry and Gaztanaga (1993) via the bias function F,

$$\delta_G = F(\delta_M) \,, \tag{1.23}$$

which assumes that the bias function is a local mapping between the matter and galaxy density contrast.

Furthermore they consider that the biasing function F can be expanded in Taylor series,

$$\delta_G = \sum_n \frac{b_n \delta_M^n}{n!} \,, \tag{1.24}$$

where the coefficients b_n define the bias function (Fry and Gaztanaga, 1993).

Since the matter density fluctuations are very small at large scales $\delta_M << 1$, we can stop the series to the first b_0 term with the other contributions negligible. Under this assumption, the matter density field is related by a simple normalisation factor to the galaxy one with the consequent power given by

$$\mathrm{P_G(k) \approx b_0^2\, P_M(k)}\,. \tag{1.25}$$

In the following we adopt a linear $\mathrm{b \approx b_0}$ and scale independent $\mathrm{b(k) \approx b}$ bias term.
Given linear theory (an approximation to non-linearities) and a biasing scheme between matter and galaxies, a measurement of the galaxy power spectrum leads to constraints of some cosmological parameters.
Huge future surveys could measure the power at scales $\mathrm{k < k_{eq}}$ constraining the primordial amplitude $\mathrm{n_s}$ since, at those scales, structures have entered the horizon in the matter dominated era maintaining the primordial shape information.
The amplitude integrated over the whole k-space range is proportional to the primordial amplitude $\mathrm{A_s}$. Smoothing the matter density field with a spherical filter of size R gives a further field with spatial variance

$$\sigma_{\mathrm{R}}^2 = \frac{1}{2\pi}\int_0^\infty \mathrm{P(k)W^2(k,R)k^2\,dk} \tag{1.26}$$

where $\mathrm{W(k,R)}$ is the window function for the smoothing, the Fourier Transform of a sphere. Usually $\mathrm{R = 8\,h^{-1}Mpc}$ because at this scale cosmologists predict that non-linearities are negligible and damped out with the smoothing. The value σ_8 is directly proportional to the primordial amplitude $\mathrm{A_s}$. The problem is that we do not have surveys large enough to measure the power at scales $\mathrm{k < k_{eq}}$ to directly infer $\mathrm{n_s}$ and consequently we cannot perform the integral of equation 1.26 to recover $\mathrm{A_s}$.
Another cosmological information could be extracted from the wavenumber of the peak of the power spectrum because is the scale of the horizon at the time of equality, $\mathrm{k_{eq} = (2\,\Omega_M\,H_0^2\,z_{eq})^{1/2}}$, but again this would require sampling very large scales and also the peak is very broad due to equality not being instantaneous.
The cosmological parameters that mainly affect the shape of the power spectrum are the matter density Ω_{M} and the baryonic fraction $\mathrm{f_B} = \frac{\Omega_{\mathrm{B}}}{\Omega_{\mathrm{M}}}$. The evolution of structures is strictly dependent on the amount of matter in the Universe. Varying Ω_{M} and fixing all the other parameters, the shape of the power spectrum changes as shown in figure 1.1. The peak of the power shifts toward higher modes in a Universe with higher mean density, because the time of equality between radiation and matter happens at higher redshift, when the horizon scale was smaller.
The power of large scale fluctuations, instead, is not altered by the variation of Ω_{M} since at $\mathrm{k < k_{eq}}$ perturbations undergo the same growth. The parameter $\mathrm{n_s}$ is fixed to $\mathrm{n_s = 0.95}$ in this figure, as suggested by the latest Planck measurements (Planck Collaboration et al., 2013).
When adding more baryons with respect to CDM (i.e. increasing the baryonic fraction $\mathrm{f_B} = \frac{\Omega_{\mathrm{B}}}{\Omega_{\mathrm{M}}}$), the damping of power on scales $\mathrm{k > k_{eq}}$ is stronger. In fact once the fluctuation enters the horizon, baryons remain frozen until the epoch of decoupling because interact with photons well beyond the equality. After decoupling, they follow the underlying gravitational potential generated by CDM that instead were free to grow already

since the time of equality. Fixing the amount of total matter in the Universe and varying the baryonic fraction, in the early Universe there was less CDM able to cluster between $\mathrm{t_{dec}}$ and $\mathrm{t_{eq}}$ and more frozen baryons. This situation would lead to a suppression of the power for modes that entered the horizon before the time of equality $\mathrm{k > k_{eq}}$ but keeps the peak position $\mathrm{k_{eq}}$ fixed because $\mathrm{k_{eq}}$ depends on the total amount of matter and not on the baryonic fraction (figure 1.2). Comparing the measured power spectrum from a galaxy redshift survey with theoretical models dependent on Ω_{M} and $\mathrm{f_B}$ provides a powerful way of constraining these two cosmological parameters.

An important feature in the power spectrum that has been outlined in the last years are the oscillations visible in figure 1.2 on scales of the order of $\mathrm{k} \sim 0.1\,\mathrm{h\,Mpc^{-1}}$. These are related to the baryonic acoustic oscillations, observed in the CMB spectrum, that leave an imprint on the matter power spectrum and on the clustering. Increasing the density of baryons, the BAO signal is stronger as outlined in the figure. The recent observation of BAO is a proof that baryons are not so negligible in the evolution of structures as it was thought. The most important part is that the amplitude and position of these wiggles are related to the size of the sound horizon at decoupling. Thus BAO provides a measuring stick that measured at different redshift could helping to measure the expansion rate of the Universe, in a completely independent way from the supernova technique.

Recently another issue related to the large-scale structure has been pointed out by Guzzo et al. (2008). Since galaxies are measured in redshift space, the shape of the power spectrum is altered in an anisotropic way enhancing observed structures at large scales due to the falling of galaxies in high density peaks along the line-of-sight and a damping of structures at smaller scales due to the Fingers of God. Instead of looking at redshift-space distortions as an annoying factor that should be taken into account in the modelling of the power spectrum, a correct study of the anisotropic clustering introduced in linear theory (Kaiser, 1987) leads to an estimate of the growth rate parameter. The growth rate of structure is predicted by the gravity theory and a correct estimate of this value could prove if the General Relativity is the correct answer to gravity or not.

1.3.3 Surveys

With the cosmological information enclosed in the power spectrum, many projects have been and are dedicated to the analysis of the distribution of galaxies through the measurements of millions galaxy distances, through their redshifts.
Despite the small number of galaxies and the small volume sampled, the first surveys in the 1980s and 1990s were determinant to discriminate a $\Omega \neq 1$ Universe meaning that the matter density is only a part of the total quantity of energy in the Universe. Due to the advancement of technology, big surveys were started in 1990s, reducing the statistical error of the matter density and constraining other parameters as the baryonic fraction. The two most important of such surveys, which significantly improved our vision of the Universe, are the Two Degree Field Galaxy Redshift Survey (2dFGRS, Colless et al., 2003) and the Sloan Digital Sky Survey (SDSS, Strauss et al., 2002).

The 2dFGRS, fully released in 2003, covers an area of approximately 1500 square degrees observing $\sim 220,000$ reliable redshifts for galaxies in a volume of $\mathrm{V} \sim 10^8\ \mathrm{h^{-3}Mpc^3}$. The 2dFGRS surveys the distribution of galaxies in the local Universe, covering objects with a median redshift of $\mathrm{z} = 0.11$. The power spectrum of the 2dFGRS survey was first measured by Percival et al. (2001) with the survey partially completed, at a 70% stage. The main result of their paper was a constraint on Ω_{M} and $\mathrm{f_B}$, with a strong degeneracy between the two. The study of the power spectrum for the full survey by Cole et al. (2005) was helpful to break the degeneracy. In addition to the close constraint measured with the complete 2dFGRS, one of the major achievement of the survey was the discovery of the BAO signal in the measured power spectrum, providing an important verification of our fundamental model of structure formation.

The SDSS started observing data in 2000 and is currently being further extended (Dawson et al., 2013). The first two projects of the SDSS covered the period between 2000 and 2008 sampling the local Universe with a median redshift of $\mathrm{z} = 0.1$, observing $\sim 800,000$ spectra of galaxies over 7,500 square degrees. A subsample of this distribution of galaxies is represented by a particular class of objects: the Luminous Red Galaxies (LRG). This sample of 46,748 LRG covers $\mathrm{V} \sim 0.72\,\mathrm{h^{-3}Gpc^3}$ over 3816 square degrees and a redshift range $0.16 < \mathrm{z} < 0.47$, making it, at the time, the best sample for the study of large-scale structure. Eisenstein et al. (2005) performed a correlation function analysis detecting the BAO peak at a mean redshift of $\langle \mathrm{z} \rangle = 0.35$ and constraining, for the first time in the large-scale structure approach, the dark energy properties.

Many recent redshift surveys have also been focused on redshift-space distortions to study the anisotropy of the clustering due to the peculiar velocity of galaxies. These projects have discarded many theories of gravity but we are not yet able to discriminate between General Relativity or other theories of modified gravity. The aim of these modern surveys is to extend the study of the clustering of matter at higher redshift, $\mathrm{z} \sim 1$ when the Universe was half of its size. Surveys, able to reach very large distances, are the WiggleZ Dark Energy Survey (Blake et al., 2008) and the current Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al., 2013), extension of the SDSS, i.e. SDSS III.

WiggleZ is an Australian galaxy survey completed in 2011 and composed of $\sim 240,000$ redshifts over an area of 1000 square degrees. The redshift range $0.5 < \mathrm{z} < 1$, where the blue emission-line galaxies of WiggleZ are observed, allows to extend these studies on the clustering of matter up to a mean redshift of $\langle \mathrm{z} \rangle \sim 0.8$. The main goal of the survey has been the study of the anisotropy of the clustering due to RSD to measure the growth rate of structure (Blake et al., 2011b) and the analysis of the properties of dark energy throughout the BAO signal in the power spectrum and correlation function (Blake et al., 2011a; Parkinson et al., 2012).

BOSS is an extension of the SDSS project at higher redshift ($\mathrm{z} < 0.7$) designed to recover the length of BAO at different redshifts. The properties of BOSS are impressive, with 10,000 square degrees sampled and ~ 1.5 million massive galaxy redshifts measured. Anderson et al. (2012) measured through a power spectrum and correlation function analysis the BAO imprint on the clustering of matter at a mean redshift of $\langle \mathrm{z} \rangle \sim 0.57$ using a huge volume of $\mathrm{V} = 2.2\,\mathrm{h^{-3}Gpc^3}$.

The VIMOS Public Extragalactic Redshift Survey (VIPERS, Guzzo et al., 2014; Garilli et al., 2014) has been developed in this context to extend the study of the distribution of galaxies to z $\sim$ 1. Despite the small covering area of $\sim$ 24 square degrees, the high redshift extension $0.4 < z < 1.2$ restrict in a fairly large sampled volume, makes VIPERS comparable with the properties of the 2dFRGS and SDSS at z $\sim$ 0.

The main property of VIPERS is, in fact, its redshift range that allows us to measure the clustering of matter up to redshift z $\sim$ 1. The uniqueness of VIPERS is also due to the highest sampling rate ever tried in a survey at these redshifts. A general introduction of VIPERS will be given in the next chapter.

Chapter 2

The VIPERS sample

VIPERS is an ongoing ESO Large Program which aims at measuring redshift for $\sim 100,000$ galaxies using the spectrograph VIsible Multi-Object Spectrograph (VIMOS) at VLT. At the time of writing (november 2014), observations are $\sim 90\%$ complete. The first Public Data Release (PDR-1), available at http://vipers.inaf.it, occurred in September 2013 with the survey complete at $\sim 70\%$.

The main goal of VIPERS is to map the three dimensional distribution of galaxies in order to push our understanding of galaxy clustering, growth of structure and galaxy properties at an epoch when the Universe was half its current age, $\mathrm{z} \sim 1$. To achieve its goals, VIPERS is an extended redshift survey in which galaxies are observed in a redshift range between $0.4 < \mathrm{z} < 1.2$. Such depths inevitably limit the angular size and thus the volume explored in past redshift surveys such as VIMOS VLT Deep Survey (VVDS, Le Fèvre et al., 2005), the DEEP2 survey (Coil et al., 2008) and the zCOSMOS survey (Lilly et al., 2009). However, despite being centred at a mean redshift of $\langle \mathrm{z} \rangle \sim 0.7$, the most remarkable innovation of VIPERS is its similarity to local surveys like 2dFGRS and SDSS in terms of volume and number density. All these surveys are characterised by a high sampling density, compared to the sparser samples of the recent generation of BAO-oriented surveys (WiggleZ, BOSS).

VIPERS covers an area of ~ 24 square degrees with a mosaic of 288 VIMOS pointings corresponding, given the redshift interval, to a volume of $5 \times 10^7\ \mathrm{h}^{-3}\mathrm{Mpc}^3$.

Such a combination of sampling and volume is unique among redshift surveys at $\mathrm{z} > 0.5$. In this chapter, a simple overview of the VIPERS survey design and strategy is provided. For a more detailed description, we refer to Guzzo et al. (2014) and Garilli et al. (2014).

2.1 General properties

VIPERS has been designed to cover part of the W1 and W4 fields from the T0005 release of the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS, Cuillandre et al., 2012) Wide photometric catalogue.

CFHTLS Wide is a 5-band photometric survey composed of four fields, W1, W2, W3 and W4, covering a total area of 133 square degrees.

The first attempt to construct a survey at high redshift similar to VVDS, DEEP2 and zCOSMOS but with a sufficient large volume to measure meaningful cosmological information was an extension of VVDS, VVDS-Wide (Garilli et al., 2008), which covered ~ 8 square degrees down to a magnitude $i_{AB} = 22.5$, but observing all objects (stars and galaxies). VIPERS was designed to maximise the number of galaxies observed in the range of interest while at the same time attempting to select against stars.

The star/galaxy classification scheme developed to construct the VIPERS target sample is particularly important as in a purely magnitude limited sample of objects at $i_{AB} = 22.5$ the fraction of stars can be larger than 30%. The method adopted to classify stars and galaxies for VIPERS uses knowledge of the object size, provided by the half-light radius r_h (i.e. the radius containing half of the object's flux) and has been tested for a significant number of objects in both W1 and W4 provided by the VVDS Deep and Wide surveys. The stellar contamination is found to be only 3.2% meaning that the methods used to pre-select galaxies are very efficient.

Given the luminosity function of galaxies and results from previous VIMOS surveys such as VVDS and zCOSMOS, we knew that a magnitude-limited sample with $i_{AB} < 22.5 - 23.0$ would cover the redshift range out to $z \sim 1.2$ and could be assembled with fairly short VIMOS exposure times (< 1 hour). The redshift range was further isolated through a simple and robust colour-colour selection on the (r-i) vs (u-g) plane, that allowed to discard $z < 0.5$ galaxies. We could be confident in advance that this selection method would efficiently remove galaxies at $z < 0.5$, while yielding $> 98\%$ completeness for $z > 0.6$. A precise calibration of this separation method was made possible by the location of the VVDS-Wide ($i_{AB} < 22.5$) and VVDS-Deep ($i_{AB} < 24$) samples within the W4 and W1 fields of CFHTLS, respectively.

Once objects have been targeted to be observed, redshifts are measured with VIMOS, a 4-channel imaging spectrograph where each channel (a "quadrant") covers $\sim 7 \times 8$ arcmin^2 for a total field of view (a "pointing") of ~ 218 arcmin^2. Each channel is a complete spectrograph with the possibility to insert 30×30 cm^2 slit masks at the entrance focal plane, as well as broad-band filters and grisms. The standard lay-out of the four quadrants on the sky is reproduced in figure 2.1. The electromagnetic spectrum could not be obtained for all the objects in the photometric catalogue pre-selected with the colour-colour selection and star/galaxy classification. The Slit Positioning Optimisation Code (SPOC, Bottini et al., 2005) is an algorithm used to select spectroscopic targets in each quadrant and designed to maximise the number of objects at each pointing. Objects selected by SPOC are then observed with the VIMOS spectrograph. The verification and validation of the measured redshift still require human intervention and has been centralised in the data reduction and management centre at INAF - IASF Milano. A quality flag is assigned to each measured redshift based on the quality of the corresponding spectrum. Here we use only galaxies with flags between 2 and 9, which correspond to a global redshift confidence level of 98%.

The errors on redshift measurement have been estimated for PDR-1 using 1215 objects with reliable redshift, observed two times because deliberately re-observed or because lying at the border of the quadrants where two quadrants overlap. The distribution of the

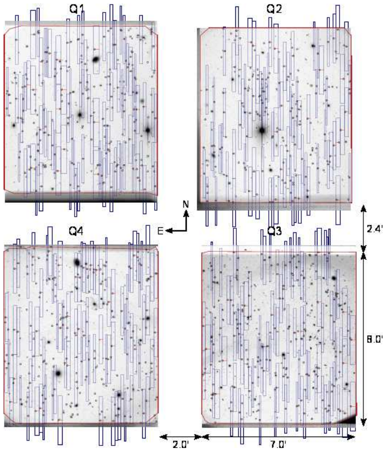

Figure 2.1: example of the detailed footprint and disposition of the four quadrants in a full VIMOS pointing (W1P082 in this case).

differences between the two redshifts of the same object, normalised to the corresponding redshift expansion factor $1 + \mathrm{z}$, is well described by a Gaussian with a dispersion of $\sigma = 200\,\mathrm{km\,s^{-1}}$ which is related to a mean 1σ error of $\sigma_{\mathrm{v}} = \sigma/\sqrt{2} = 141\,\mathrm{km\,s^{-1}}$.
The distributions of galaxies mapped with VIPERS in the two fields, W1 and W4, are shown in figure 2.2 and 2.3 for PDR-1. PDR-1 consists on $\sim 55,000$ redshifts composed

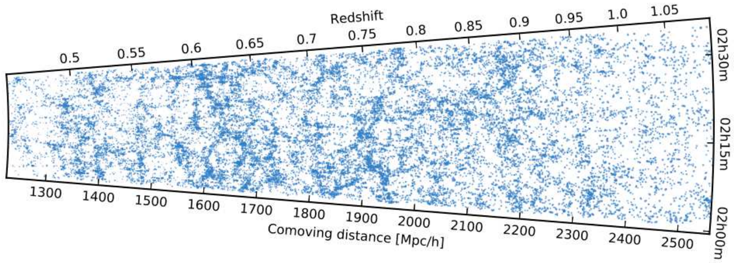

Figure 2.2: distribution of galaxies for the W1 field in the PDR-1 catalogue.

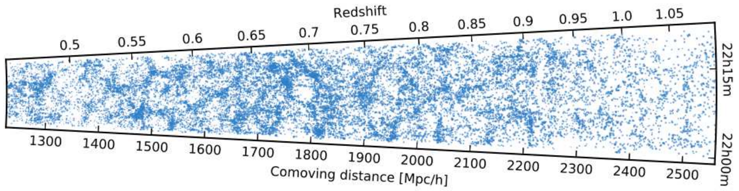

Figure 2.3: distribution of galaxies for the W4 field in the PDR-1 catalogue.

mainly of galaxies between $0.4 < z < 1.2$. Due to the statistical nature of the colour-cut selection, the catalogue is contaminated by stars and by galaxies extending outside the considered redshift range. Additionally, 20% of the redshifts are considered not sufficiently reliable for statistical analysis (flag$= 1$, $\sim 50\%$ confidence level of being wrong), and thus will be discarded from the analysis presented here. The remaining catalogue is then composed of $\sim 45,000$ reliable galaxies, which are suitable for accurate studies of galaxy clustering.
Current data (november, 2014) reaches $\sim 65,000$ high quality spectra for a total of $\sim 55,000$ galaxies.
Figure 2.4 is showing the distribution of PDR-1 VIPERS galaxies compared to the main and LRG SDSS samples. The big leap toward higher redshifts taken with VIPERS is evident.

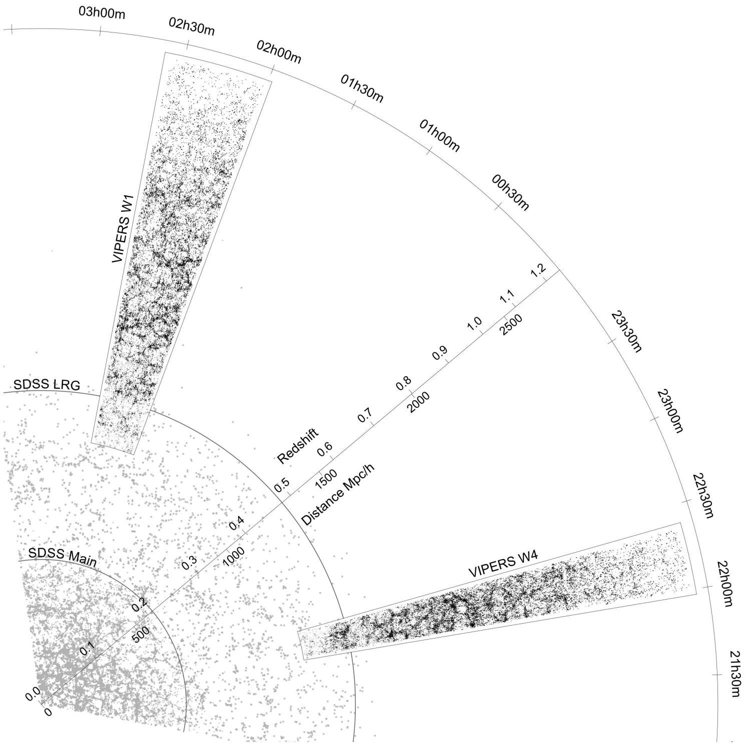

Figure 2.4: distribution of galaxies for the VIPERS W1 and W4 fields compared to the spatial distribution of galaxies in the main SDSS and in the LRG SDSS catalogues.

2.2 The VIPERS selection function

The VIPERS angular selection function is the result of the combination of several different angular completeness functions. Two of the VIPERS angular selection functions

are binary masks: one related to defects in the parent photometric sample (mostly areas masked by bright stars), while the other is a spectroscopic angular mask due to the VIMOS footprint, generated by the composition of several pointings on the sky.
The photometric mask, which is known a priori from the CFHTLS catalogue, consist primarily of patches around bright stars owing to the broad diffraction pattern and internal reflections in the telescope optics and has a small effect in terms of missing area on the VIPERS observing fields, $\sim 2.5\%$ for each field.
The spectroscopic mask, called the "Nagoya" angular mask in the internal convention, generates a particular grid pattern in the sky due to the gaps between quadrants. Moreover, it accounts for secondary features such as the quadrant that may be partly vignetted by the guide probe arm. The spectroscopic angular mask is computed after the observations because has to consider, not only the sky pattern due to the four VIMOS quadrants (which is known a priori), but also the presence of failed pointings.
As shown in figure 2.5 where the distribution of galaxies on the sky is plotted, holes between quadrants and failed pointings reduce strongly the total area of the parent observing field. The angular mask corresponding to the PDR-1 catalogue is the version 3.0 with a missing sky area of $\sim 47\%$ for the field W1 and $\sim 36\%$ for W4. With current data, one more row in the W1 field has been completed while, for the field W4, more pointings corresponding to old failures have been covered. The most recent spectroscopic mask related to the current state of observations is the "Nagoya" 4.0 angular mask and the missing area on the sky is of the order of 30% for W1 and 34% for W4.

Moreover, due to the selection of targets done by the SPOC algorithm, within each of the four VIMOS quadrants an average of 40% of the observed objects in the photometric catalogue satisfying the selection criteria are actually placed behind a slit and observed. The ratio between the number of SPOC selected objects respect to the initial available targets (pre-selected to reduce star contamination and avoid $z < 0.5$ galaxies), is defined as the Target Sampling Rate (TSR), a quantity (known before the observations) which varies with location on the sky due to fluctuations in the surface density of objects. Since SPOC is designed to maximise the number of spectroscopic targets, TSR is inversely proportional to the surface density of objects. In high sky density region, due to the limited size of a slit, objects too close cannot be observed, i.e. we miss a lot of galaxies (low TSR). Vice-versa, almost all galaxies in less populated region on the sky are observed because slits do not overlap. In figure 2.6, we show the TSR on a per-quadrant basis, using the ratio of assigned targets to potential targets.
Another quantity that has to be accounted for in the VIPERS selection function is the Spectroscopic Success Rate (SSR), defined as the number of reliable redshift measured with respect to the number of targeted galaxies due to the varying observing conditions and technical issues. Even in this case, the SSR can be evaluated as a function of VIMOS quadrant for the PDR-1 catalogue. It could be noticed, in figure 2.7, the high SSR level in each quadrant with the majority of regions with SSR $> 80\%$. A few observations under problematic conditions (either technical or atmospheric) are clearly marked out by the brown and purple rectangles.
Detailed knowledge of all these contributions to the VIPERS selection function is a crucial ingredient for any quantitative measurement of galaxy clustering.

2.3 Clustering investigations with VIPERS

Numerous works on the clustering of galaxies have already been performed using the PDR-1 VIPERS sample.
The main goal of VIPERS is an accurate measurement of the growth rate of large-scale structure at $z \sim 1$, in order to improve the estimation made by Guzzo et al. (2008) for the VVDS-Wide sample observing a larger volume to reduce statistical error bars. The study of the anisotropy of clustering has been carried out by de la Torre et al. (2013) analysing the dependence of the correlation function on scales parallel and perpendicular to the line-of-sight, including in the estimation of the correlation function all the selection function effects of VIPERS. Applying the multipoles analysis to the correlation function of the PDR-1 catalogue, de la Torre et al. (2013) have been able to provide a measurement of the growth rate $f\sigma_8 = 0.47 \pm 0.08$ at a mean redshift of $\langle z \rangle = 0.8$, allowing to add a further constraint on gravity, which remains consistent with Einstein's General Relativity but does not discriminate between different gravity models.
The PDR-1 catalogue has also been used to measure the clustering ratio η_R, a new observable introduced in Bel et al. (2014), and defined as the ratio between the correlation function $\xi_R(r)$ of a smoothed field on sphere with radius R and its variance σ^2_R. What is interesting about this statistic is that, assuming the galaxy and matter densities to be related by a simple linear biasing scheme on scales R, the clustering ratio is independent on the bias factor and insensitive to linear redshift distortions. Verifying the accuracy and efficiency of this approach using mock catalogues that match the survey selection process, Bel et al. (2014) applied the η_R method to galaxies in a redshift range $0.65 < z < 1.2$ obtaining a best fit for the matter density parameter of $\Omega_M = 0.270^{+0.029}_{-0.025}$, fixing all the cosmology (expect for the matter density) to the best known prior.
Due to the wide class of galaxies observed by VIPERS, Marulli et al. (2013) used the correlation function approach applied again to PDR-1 to measure the clustering of galaxies as a function of absolute magnitudes and stellar masses, inferring the galaxy bias as a function of luminosity, stellar mass and redshift, assuming a flat ΛCDM model to derive the dark matter two-point correlation function.
The first attempt to measure the clustering of galaxies in Fourier space has been performed in Granett et al. (2012). The 2D angular VIPERS power spectrum measured from the photometric CFHTLS catalogue has been deprojected, using the redshift distribution of a preliminary VIPERS sample composed of $\sim 14,000$ spectroscopic redshifts, in order to obtain the 3D power spectrum. The consequent estimation of cosmological results leads to a measurement of the bias factor $b = 1.38 \pm 0.05$ ($\sigma_8 = 0.8$) and the matter density $\Omega_M = 0.30 \pm 0.06$ within a redshift range $0.5 < z < 1.2$, while keeping other ΛCDM parameters fixed.

Given the redshift range observed, the high sampling rate and the volume sampled, VIPERS enables broader goals involving galaxy evolution, similarly to the achievements of 2dFGRS and SDSS, but at $z \sim 0.1$. Recent works in that direction are the study of the non-linear galaxy bias (Di Porto et al., 2014), the analysis of the galaxy stellar mass functions at intermediate redshift (Davidzon et al., 2013) and the study on the formation of massive red-sequence galaxies (Fritz et al., 2014).

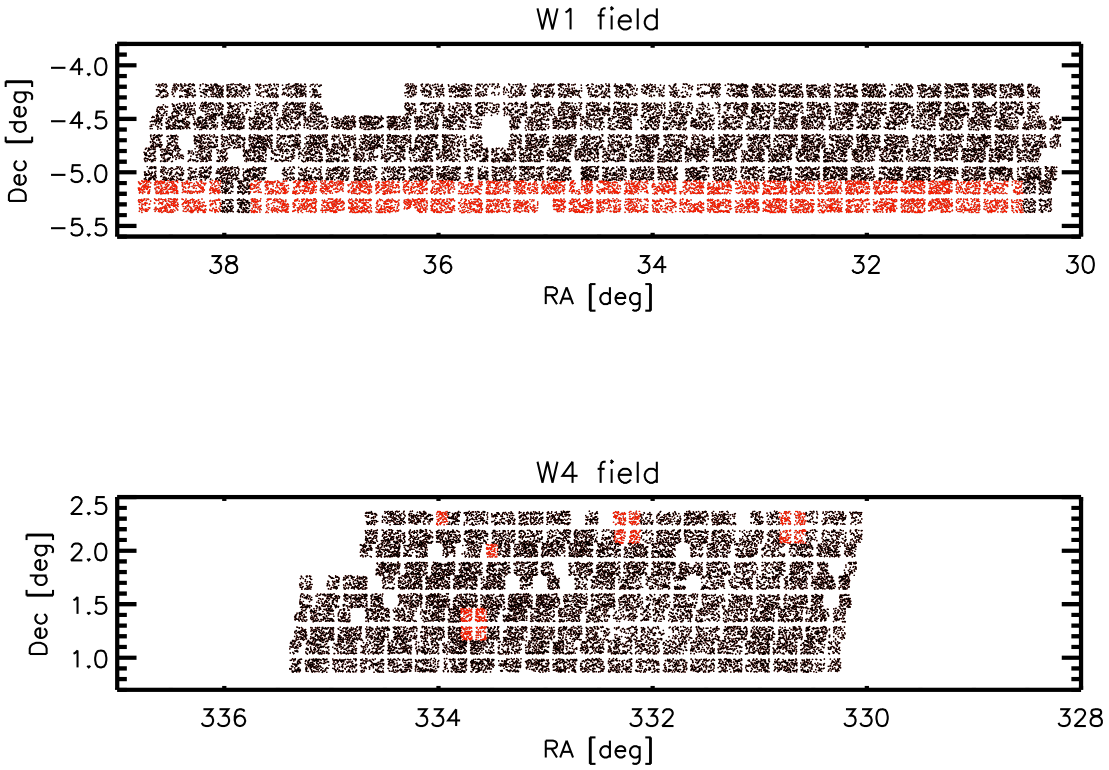

Figure 2.5: *Top panel*: distribution of VIPERS galaxies on the sky as a function of right ascension and declination for the field W1. The black dots correspond to the PDR-1 catalogue with the "Nagoya" 3.0 spectroscopic mask while the current stage of the observations ("Nagoya" 4.0) are represented by black and red points. Grid holes are due to the observing strategy of VIMOS, missing quadrants represent failed pointings and random holes (even if not easily visible in figure) are missing part of the photometric catalogue due mainly to the presence of stars. The big missing part at RA ∼ 37° corresponds to the VVDS region. *Bottom panel*: same as top panel but for the W4 field. The big missing part on the left, at RA ∼ 335°, corresponds to a border in the CFHTLS Wide catalogue.

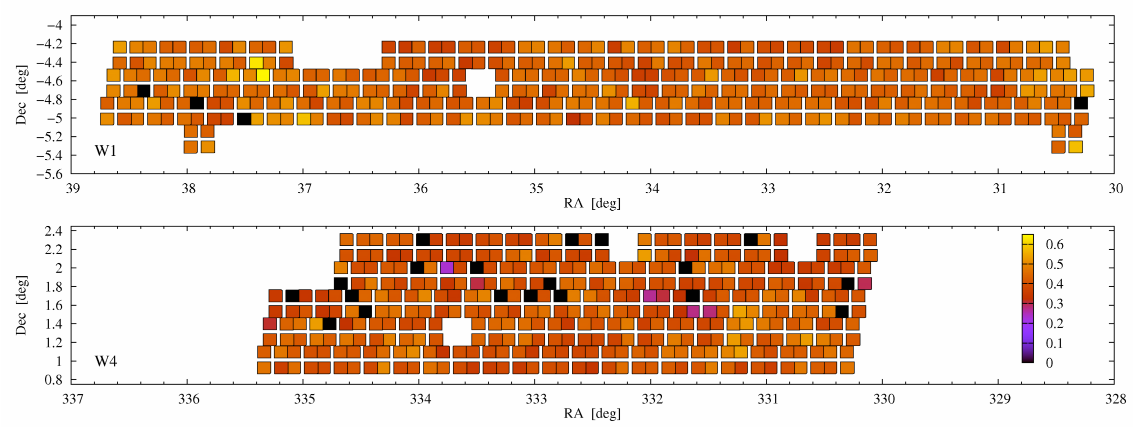

Figure 2.6: behaviour of the Target Sampling Rate (TSR) as a function of quadrant for the PDR-1 in both fields. As shown, the average TSR is around 40% while black quadrants correspond to a failure in the observation with the consequent loss of data.

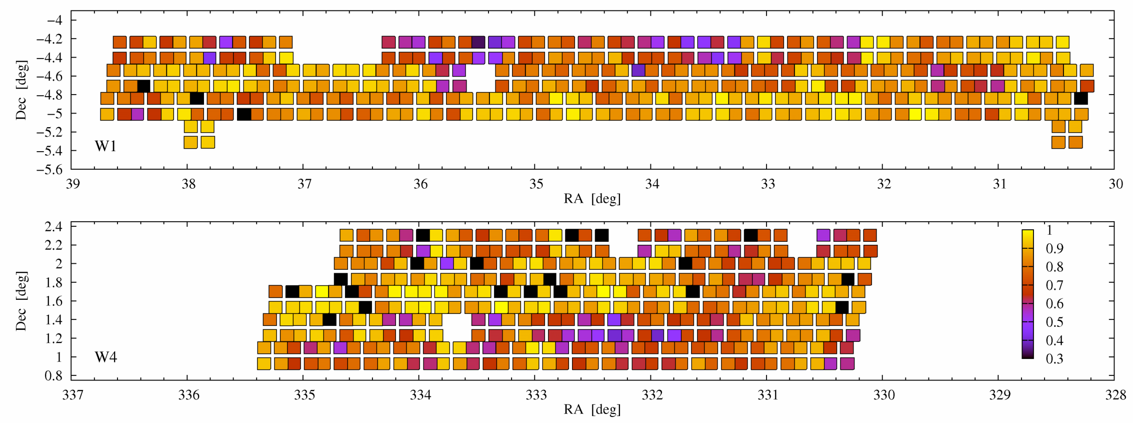

Figure 2.7: behaviour of the Spectroscopic Sampling Rate (SSR) as a function of quadrant. As can be seen, for the majority of quadrants the success rate is larger than 80%.

Chapter 3

Power Spectrum Estimator

In this chapter, we provide a practical guide to measure the power spectrum from a discrete distribution of galaxies. We construct simple configurations of points, drawn from Γ simulations of the density field (appendix B) generated from a known input power spectrum. The advantage of this procedure is that we do not have to deal with unknown bias factors or non-linear effects. An unbiased estimator should recover the input power spectrum without any systematic effects. We start with the simplest setting of particles distributed in a cubical sample with no density gradient. Step by step we add to the sample more realistic features. First we include the cone-like geometry to the sample while, in a second moment, a density gradient dependent on redshift (typical of magnitude-limited surveys) and we discuss how the power spectrum estimator accounts and corrects these factors.

3.1 Cubical box

3.1.1 Fourier Transform

The Fourier Transform (FT) method, used in several past surveys, starts from the definition of the power spectrum which is the covariance of the density field in Fourier space. Let $\rho(\mathbf{x})$ be the matter density field at the position $\mathbf{x}$ and $\bar{\rho}$ is the mean density. Then the density contrast is defined as

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} , \tag{3.1}$$

and characterises the fluctuations of $\rho(\mathbf{x})$ around its mean $\bar{\rho}$. The Fourier Transform of $\delta(\mathbf{x})$ can be defined as

$$\delta(\mathbf{k}) = \int_{\mathrm{V}} \delta(\mathbf{x})\, \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}}\, \mathrm{d}^3\mathrm{x} , \tag{3.2}$$

where V is in principle an infinite volume. In practise, it corresponds to the size of the sample. We assume that V is large enough to represent the overall density field of this toy model.

From the definition of the power spectrum in equation 1.7, the power spectrum estimator is simply related to $\delta(\mathbf{k})$ with the formula

$$\mathrm{P(k)} = \frac{\langle|\delta(\mathbf{k})|^2\rangle}{\mathrm{V}} \ , \tag{3.3}$$

where $\langle ... \rangle$ denotes the ensemble average and the volume of the survey V has been introduced in order to express the power spectrum in units of volume since, from equation 1.11, $\mathrm{P(k)} \equiv \int \xi(\mathrm{r})\, \mathrm{e}^{\mathbf{ik\cdot x}}\, \mathrm{d^3x}$. However, in any point process we do not observe directly the continuous density field. It is sampled by a finite number of particles which are used to trace the underlying fluctuations. The continuous field $\rho(\mathbf{x})$ could be replaced with a discrete density

$$\mathrm{n}(\mathbf{x}) = \sum_{\mathrm{G}=1}^{\mathrm{N_G}} \delta^{\mathrm{D}}(\mathbf{x} - \mathbf{x}_{\mathrm{G}}) \ , \tag{3.4}$$

where each particle at position $\mathbf{x}_{\mathrm{G}}$ has been replaced by a Dirac-δ function and the sum is performed over the total number $\mathrm{N_G}$ of objects. This ensures that $\bar{\mathrm{n}} = \bar{\rho}$. Substituting equation 3.4 in 3.2, the density contrast in Fourier space becomes

$$\delta(\mathbf{k}) = -\delta^{\mathrm{K}}(\mathbf{k} - \mathbf{0}) + \frac{1}{\bar{\mathrm{n}}} \sum_{\mathrm{G}=1}^{\mathrm{N_G}} \mathrm{e}^{-\mathrm{i}\,\mathbf{x}_{\mathbf{G}}\cdot\mathbf{k}} \ . \tag{3.5}$$

The δ^{K} is the Kronecker-delta which is equal to 0 if $\mathbf{k} \neq \mathbf{0}$ and 1 if $\mathbf{k} = \mathbf{0}$.
Since no gradient is imposed to the density, $\bar{\mathrm{n}}$ does not depend on position and is obtained by $\bar{\mathrm{n}} = \mathrm{N_G/V}$. The mean density is evaluated directly from the distribution of particles used to measure the power spectrum. In principle $\bar{\mathrm{n}}$ could be different respect to the true underlying value. Our inability of knowing a priori the mean density of the sample leads to systematic effects in the power spectrum estimation, called "integral constrain", which we will account for modifying the theoretical prediction of the power spectrum when we will compare measurements with theory (chapter 4).
The FT procedure evaluates directly the real and imaginary part of the density field in Fourier space as

$$\begin{aligned} \delta_{\mathrm{REAL}}(\mathbf{k}) &= -\delta^{\mathrm{K}}(\mathbf{k} - \mathbf{0}) + \frac{1}{\bar{\mathrm{n}}} \sum_{\mathrm{G}=1}^{\mathrm{N_G}} \cos(\mathbf{x}_{\mathrm{G}} \cdot \mathbf{k}) \\ \delta_{\mathrm{IM}}(\mathbf{k}) &= -\delta^{\mathrm{K}}(\mathbf{k} - \mathbf{0}) + \frac{1}{\bar{\mathrm{n}}} \sum_{\mathrm{G}=1}^{\mathrm{N_G}} \sin(\mathbf{x}_{\mathrm{G}} \cdot \mathbf{k}) \ , \end{aligned} \tag{3.6}$$

with the three dimensional power spectrum estimator defined as

$$\mathrm{P}(\mathbf{k}) = \frac{\delta^2_{\mathrm{REAL}}(\mathbf{k}) + \delta^2_{\mathrm{IM}}(\mathbf{k})}{\mathrm{V}} - \frac{1}{\bar{\mathrm{n}}} \ . \tag{3.7}$$

The Poisson shot-noise term $\mathrm{P_{SN}} = 1/\bar{\mathrm{n}}$ must be subtracted from the power spectrum estimator. The shot-noise term is due to the self-particle contribution to the statistics and it comes from the discreteness of the sample. If particles are randomly distributed, it could be demonstrated with definition 3.3 and through equation 3.6 that the power is not 0 but equal to the inverse of the mean density.

In the first test to introduce the power spectrum estimator, we consider particles distributed in a cubical box with size L and volume $V = L^3$. We generate a k-space grid with the same step in all three directions (k_x, k_y, k_z) given by $\Delta k = 2\pi/L$, with a minimum frequency $k_{min} = \Delta k$, and a value for k_{max} dependent on the smallest scale we want to reach. k_{min} is limited by fluctuations with wavelength comparable to the size of the survey $\lambda_{max} \sim L$ or $k_{min} \sim 2\pi/L$ because we cannot analyse modes larger than the box. k_{max} instead can be extended up to very small scales. The problem here is more practical because the increase of k_{max} is directly proportional to the enhancement of the computation-time. Since we have to calculate the contribution of each particle to the density field in each single mode, the number of iterations in the FT method is proportional to the number of modes in the k-space grid times the number of particles: $N_{it} \propto \frac{k_{max}}{k_{min}}^3 \times N_G$. The monopole power spectrum is calculated spherical averaging the 3D power of expression 3.7,

$$P(|k|) = \frac{1}{N_k} \sum_{k<|\mathbf{k}|<k+\delta k} P(\mathbf{k}) \ , \tag{3.8}$$

where N_k is the number of modes in the shell.

We apply the FT estimator to Γ realisations with particles distributed in a cubical box with size $L = 486.4\,h^{-1}$Mpc.

Statistical errors in the power spectrum, under the assumption of Gaussian fluctuations, are proportional to the volume of the sample V with the variance of the power, called "cosmic variance", such that $\sigma_P^2(k)/P^2(k) \propto V/V_k$ where V_k is the volume defined in Fourier space (Feldman et al., 1994). A number of 30 independent realisations are used in this analysis in order to reduce cosmic variance and have an idea of statistical errors. We evaluate the power spectrum measuring the Fourier Transform of the density contrast as in equation 3.7. In order to reduce the computation-time, we choose a very sparse k-space grid with minimum frequency $k_{min} = \frac{2\pi}{L} \sim 0.013\,h\,Mpc^{-1}$, step $\Delta k = k_{min}$ and maximum frequency given by $k_{max} = 32 \times \Delta k \sim 0.41\,h\,Mpc^{-1}$. The number 32 is chosen to reduce the number of points in the k-space grid to a reasonable value of $(2 \times 32)^3$ where the factor of 2 accounts for negative frequencies. Since the number of particles in our Γ simulation is very big ($N_G \sim 2.2 \times 10^6$) and the number of FT iterations is $N_{it} \propto N_G$, we also randomly under-sample at 5% level the number of points in each realisation. The shot-noise factor will increase due to the decreasing of the mean density but our calculation will speed up by a factor of 20.

The averaged power spectrum over 30 mock catalogues after applying the spherical average (equation 3.8) is represented in figure 3.1. Statistical errors correspond to the 1σ dispersion of the power spectrum P(k) for all the realisations with respect to the mean $\bar{P}(k)$,

$$\sigma_i^2 = \frac{\sum_N \left[P(k_i) - \bar{P}(k_i)\right]^2}{(N-1)N} \ , \tag{3.9}$$

with N = 30 the number of catalogues. Errors are bigger at large scales due to cosmic variance; modes comparable to the size of the cube are less sampled than those at small wavelength. Due to the limited number of particles used for the computation of the power, the amplitude of smaller scales is shot-noise limited.
The comparison between the expected input power spectrum and the observed one gives a perfect match within statistical uncertainties on the whole scale range considered: the FT approach is a good estimator of the power spectrum. We refer to figure 3.1 as the "monopole" power spectrum because, from equation 3.8, it depends only on the modulus of k. Despite the high value for the shot-noise, the P_{SN} correction is working well even at small scales (high k-modes) where the shot-noise contribution dominates. The problem related to this technique is the long computation-time required to perform the Fourier Transform of the density field.

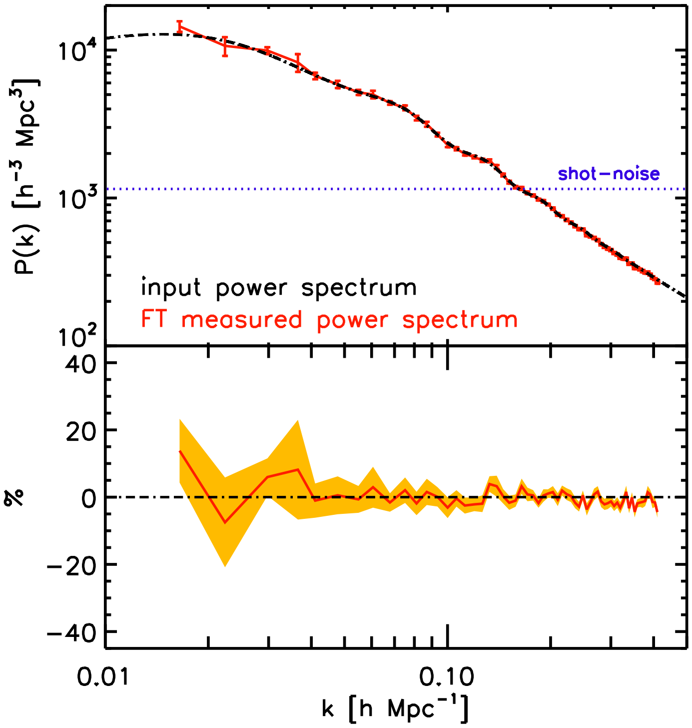

Figure 3.1: *Top panel*: comparison between the recovered power spectrum from 30 cubical box (red line) and the input power spectrum of the simulation (black line). Errors correspond to the 1σ dispersion for all the samples respect to the averaged power. The dotted blue line correspond to the shot-noise level which is quite dominant at smaller scales. *Bottom panel*: ratio between the measured power spectrum and the input expected one with related statistical errors. The black dot-dashed line shows the level of the input power spectrum.

In modern surveys, the number of observed galaxies is of the order of millions leading to the development of faster estimators. The simple three dimensional Fourier Transform has been replaced by a Fast Fourier Transform method based on a clever and fast algorithm.

3.1.2 Fast Fourier Transform

In the last 15 years due the the increasing amount of simulated data, power spectra are typically computed using a Fast Fourier Transform (FFT) algorithm in order to speed up the computation of the Fourier Transform of the density field $\delta(\mathbf{x})$. The high performance of the FFT is due to the reduced number of iterations. For the one dimensional case, computing the discrete FT of N points takes $\mathrm{O(N^2)}$ arithmetical operations, while an FFT can perform the same computation in only $\mathrm{O(N\log N)}$ operations.

As an input of FFT, one must supply the density field regularly sampled on a grid on which the FFT algorithm will apply periodic boundary conditions. The parameters of the grid characterise the k-space domain where the power spectrum will be calculated. The total range L spanned by the grid is related to both the minimum frequency of the sampled power, $\mathrm{k_{min}} = 2\pi/\mathrm{L}$, and its resolution, $\Delta\mathrm{k} = \mathrm{k_{min}}$. The resolution of the grid H (distance between two grid points) fixes $\mathrm{k_{max}}$, the maximum wavenumber. It is recommended (even if not necessary) to use a cubical geometry for the grid in configuration space: the Fourier counterpart will lead in fact to a regular sampling of the power with the same steps in all the three directions, $\Delta\mathrm{k_x} = \Delta\mathrm{k_y} = \Delta\mathrm{k_z}$. The choice of H is determinant because limits the maximum frequency at which the power spectrum can be measured. The Nyquist theorem states that a signal can be reconstructed if the sampling frequency $\mathrm{k_s}$ is at least double the highest frequency $\mathrm{k_{max}}$ of the signal itself, $\mathrm{k_s} \geq 2\mathrm{k_{max}}$. This means that if we sample the signal $\delta(\mathbf{x})$ with periodicity $\lambda_\mathrm{s} = \mathrm{H}$ ($\mathrm{k_s} = \frac{2\pi}{\mathrm{H}}$), the highest frequency of the signal we can reconstruct is $\mathrm{k_{max}} = \frac{2\pi}{\mathrm{H}} \times \frac{1}{2} = \pi/\mathrm{H}$, which is called the Nyquist frequency $\mathrm{k_N}$.

As said before, to get the Fourier Transform of the density field at each point of the grid, the number density must be first smoothed in cubical cells with size H,

$$\mathrm{n}(\mathbf{x}_\mathrm{P}) = \frac{\mathrm{N}(\mathbf{x}_\mathrm{P})}{\mathrm{H}^3} \ , \tag{3.10}$$

where $\mathrm{N}(\mathbf{x}_\mathrm{P})$ is the number of particles assigned at each point of the grid at the grid-position $\mathbf{x}_\mathrm{P}$. This is what we call the Mass Assignment Scheme (MAS). The density contrast becomes

$$\delta(\mathbf{x}_\mathrm{P}) = \frac{\mathrm{n}(\mathbf{x}_\mathrm{P})}{\bar{\mathrm{n}}} - 1 \ , \tag{3.11}$$

with $\bar{\mathrm{n}}$ defined, under the assumption of no gradient in number density, as the total number of points divided by the volume of the box, $\bar{\mathrm{n}} = \mathrm{N_G/V}$.

We then perform the fast Fourier Transform (FFT) using the publicly available software fftw3 (Fastest Fourier Transform in the West, Frigo and Johnson, 2012). Applying the FFT algorithm to $\delta(\mathbf{x}_\mathrm{P})$, we obtain the density field in Fourier space $\delta(\mathbf{k})$ sampled on a grid of values. The monopole power spectrum is then simply obtained by the spherical average (equation 3.8) of the 3D power estimator after subtracting the shot-noise term

$$\hat{\mathrm{P}}(\mathbf{k}) = \frac{|\mathrm{FFT}\,\delta(\mathbf{x})|^2}{\mathrm{V}} - \frac{1}{\bar{\mathrm{n}}} \,. \tag{3.12}$$

As shown by equation 3.10, the measured power spectrum will depend on the chosen MAS. The contribution of a particle to the density field of a grid point is bigger proportionally to the distance between the two. Mathematically the MAS is a convolution between the number density of objects $\mathrm{n}(\mathbf{x})$ and the window function $\mathrm{W}(\mathbf{x})$ of the cell which quantifies how much a particle contributes to the surrounding grid points

$$\mathrm{n}(\mathbf{x}_\mathrm{P}) = \int \mathrm{n}(\mathbf{x}')\,\mathrm{W}(\mathbf{x}_\mathrm{P} - \mathbf{x}')\,\mathrm{d}^3\mathrm{x}' \,. \tag{3.13}$$

In our work, we implement the two most popular MAS which are characterised by different shapes for $\mathrm{W}(\mathbf{x})$: the Nearest Grid Point (NGP) and Cloud-in-Cell (CIC).
The NGP scheme assigns particles to their nearest grid point. Each particle will give a contribution only to the nearest grid point and 0 to all the others. In one dimension, the window function would be a top-hat

$$\mathrm{W}(\mathrm{x}) = \begin{cases} 1, & \text{if } |\mathrm{d}| \leq \mathrm{H}/2 \\ 0, & \text{if } |\mathrm{d}| > \mathrm{H}/2 \,. \end{cases} \tag{3.14}$$

Since a convolution in configuration space becomes a multiplication in Fourier space, this effect can be easily taken into account simply dividing the density contrast in Fourier space $\delta(\mathrm{k})$ by the Fourier Transform of $\mathrm{W}(\mathrm{x})$

$$\delta_\mathrm{NGP}(\mathrm{x}) = [\delta \star \mathrm{W}](\mathrm{x}) \qquad \Rightarrow \qquad \delta_\mathrm{NGP}(\mathrm{k}) = \delta(\mathrm{k}) \times \mathrm{W}(\mathrm{k}) \,, \tag{3.15}$$

with $\mathrm{W}(\mathrm{k})$ being

$$\mathrm{W}(\mathrm{k}) = \mathrm{sinc}\left(\frac{\pi \mathrm{k}}{2\mathrm{k}_\mathrm{N}}\right) \,. \tag{3.16}$$

The CIC scheme is more sophisticated than NGP: each galaxy increases the number density of nearby cells with a weight proportional to the distance between the particle and the grid point. The window function in configuration space is therefore

$$\mathrm{W}(\mathrm{x}) = \begin{cases} 1 - |\mathrm{d}|/\mathrm{H}, & \text{if } |\mathrm{d}| \leq \mathrm{H} \\ 0, & \text{otherwise} \,, \end{cases} \tag{3.17}$$

and its Fourier Transform

$$\mathrm{W}(\mathrm{k}) = \mathrm{sinc}^2\left(\frac{\pi \mathrm{k}}{2\mathrm{k}_\mathrm{N}}\right) \,. \tag{3.18}$$

Again to correct for the MAS effect, the measured $\delta(\mathrm{k})$ must be divided by the corresponding $\mathrm{W}(\mathrm{k})$.
For CIC assignment, when the grid in configuration space is small enough that barely

enclose the survey, periodic condition must be applied. For instance in a cubical survey with size L enclosed in a grid with the same size, a galaxy at position (x,x,L) will contribute to the density field of cells at position (x,x,L) and (x,x,0).
The window function, associated to cubical cells, is simply given by the multiplication of the window in all the three directions

$$\mathrm{W}(\mathbf{x}) = \mathrm{W}(\mathrm{x})\,\mathrm{W}(\mathrm{y})\,\mathrm{W}(\mathrm{z})\ . \tag{3.19}$$

The corresponding Fourier Transform is given by

$$\mathrm{W}(\mathbf{k}) = \left[\mathrm{sinc}^2\left(\frac{\pi \mathrm{k_x}}{2\mathrm{k_N}}\right)\mathrm{sinc}^2\left(\frac{\pi \mathrm{k_y}}{2\mathrm{k_N}}\right)\mathrm{sinc}^2\left(\frac{\pi \mathrm{k_z}}{2\mathrm{k_N}}\right)\right]^{\mathrm{p}}\ , \tag{3.20}$$

where $\mathrm{p} = 1, 2$ for NGP and CIC, respectively.
A well-known effect in Fourier theory is the "aliasing" problem: the Fourier Transform of a function, which has been sampled at a certain sampling rate, is equal to the sum of infinitely many copies of the Fourier Transform ("alias") of the effective signal. When we perform an FFT analysis, we sample the density contrast at regular grid point assigning particles with a MAS to cells. The estimated power spectrum $\hat{\mathrm{P}}(\mathbf{k})$ of equation 3.12 is then related to the true one $\mathrm{P}(\mathbf{k})$ by equation

$$\hat{\mathrm{P}}(\mathbf{k}) = \sum_{\mathbf{n}} |\mathrm{W}(\mathbf{k} + 2\mathrm{k_N}\mathbf{n})|^2 \mathrm{P}(\mathbf{k} + 2\mathrm{k_N}\mathbf{n}) + \frac{1}{\bar{\mathrm{n}}}\sum_{\mathbf{n}} |\mathrm{W}(\mathbf{k} + 2\mathrm{k_N}\mathbf{n})|^2 \tag{3.21}$$

(demonstration in Jing, 2005) with $\mathbf{n}$ an integer vector and the factor W^2 introduced by the convolution of the density field with the smoothing window for the MAS. The "aliasing" term comes from the contribution to the power from higher order $\mathbf{n} \neq \mathbf{0}$.
Jing (2005) proposed an iterative method to extract from equation 3.21 the true power spectrum. Both effects of the convolution and the alias are significant near the Nyquist wavenumber $\mathrm{k_N}$. We do not care too much on what is happening around $\mathrm{k_N}$ because, given the speed of the FFT, we can push $\mathrm{k_N}$ to be very high (at uninteresting modes), decreasing the value of H, where we do not care about any "aliasing" problem. This is why we choose to correct the estimated 3D power spectrum only for the first term $\mathbf{n} = \mathbf{0}$ in equation 3.21 simply dividing the estimated power of equation 3.12 by 3.20

$$\mathrm{P}(\mathbf{k}) = \hat{\mathrm{P}}(\mathrm{k_x}, \mathrm{k_y}, \mathrm{k_z}) \times \left[\mathrm{sinc}^2\left(\frac{\pi \mathrm{k_x}}{2\mathrm{k_N}}\right)\mathrm{sinc}^2\left(\frac{\pi \mathrm{k_y}}{2\mathrm{k_N}}\right)\mathrm{sinc}^2\left(\frac{\pi \mathrm{k_z}}{2\mathrm{k_N}}\right)\right]^{-2\mathrm{p}}\ . \tag{3.22}$$

After correcting for the MAS, the monopole power spectrum is obtained applying the usual spherical average of 3.8.
The FFT approach is tested considering again 30 realisations constructed with Γ distribution of the density field. Given the efficiency and speed of the FFT algorithm, we do not need to randomly select points as we did for the FT approach in order to reduce the high number of objects.
As a first test, we smooth all the $\mathrm{N_{TOT}} \sim 2.2 \times 10^6$ particles using NGP as MAS on an FFT grid with the same size of the sample $\mathrm{L} = 486.4\,\mathrm{h}^{-1}\mathrm{Mpc}$ and a cell-size

$H = 4 \times 0.95\,h^{-1}$Mpc. The maximum mode sampled is given by the Nyquist frequency and fixed by the cell-size, $k_N = \frac{\pi}{H} \sim 0.83\,h\,\mathrm{Mpc}^{-1}$, while the minimum sampled frequency is given by the largest scale $k_{min} = \frac{2\pi}{L} \sim 0.013\,h\,\mathrm{Mpc}^{-1}$. The number of cells in the FFT grid is $N = \left[\frac{L}{H}\right]^3 = 128^3$. Figure 3.2 outlines the difference between the recovered power spectrum with and without the correction for the MAS applying the wrong estimator 3.12 in one case and the corrected one 3.22 in the other. Given the large number of particles, the shot-noise term compared to the FT case is negligible, even at the smallest considered scales.

The corrected power for the NGP assignment is better recovering the expected power spectrum respect to the uncorrected one, with a good agreement up to mode $k \sim 50\,\%\,k_N$. Figure 3.3 shows the comparison between the behaviour of the measured power spec-

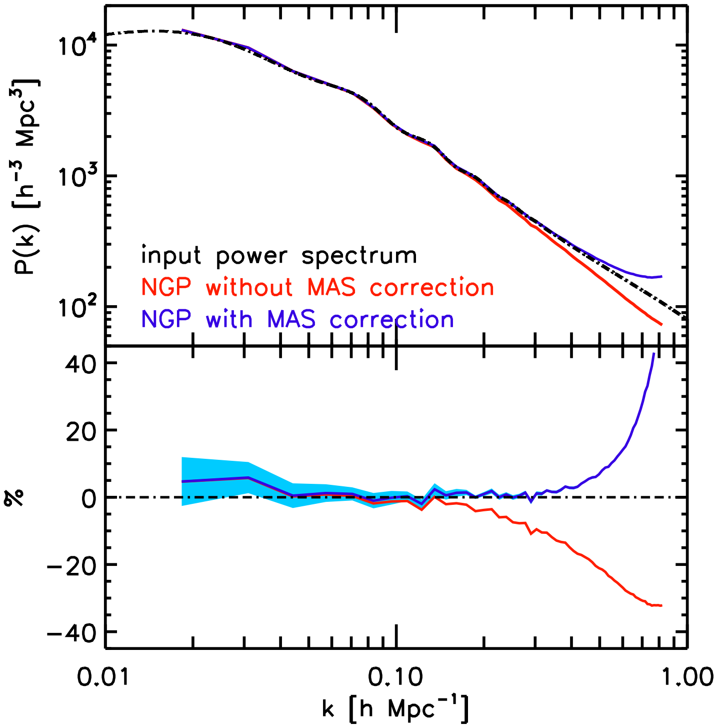

Figure 3.2: *Top panel* : test on the robustness of the MAS correction for the NGP assignment scheme. The black dot-dashed line represents the input power spectrum of the Γ simulation. The averaged power spectrum over 30 cubical samples is measured with the FFT technique smoothing the distribution of particles with NGP scheme. The red line shows the recovered power spectrum without the MAS correction (equation 3.12) while the blue line corresponds to the power corrected for the effect of NGP (equation 3.22). Errors are plotted only for the corrected case and are calculated as the dispersion of the power over all the 30 realisations. *Bottom panel* : ratio between the observed power and the input one plotted in order to outline the goodness of the correction.

trum for the two different MAS: NGP and CIC. The parameters of the FFT grid is the same as for the previous test and both recovered power spectra have been corrected for the MAS. Since the smoothing of particle density is affecting scales comparable to the size of the cell H, the effect of choosing one method or the other alters the recovered power spectrum at modes with $\mathrm{k} \sim \mathrm{k_N}$. The CIC scheme is more accurate than the NGP giving a good match between the observed and expected input power up to frequencies $\mathrm{k} \sim 75\%\,\mathrm{k_N}$ respect to the $\mathrm{k} \sim 50\%\,\mathrm{k_N}$ accuracy of the NGP scheme. This is due to the fact that the contribution of each particle to the surrounded grid points encodes information of the relative distance between the grid points and the particle considered.

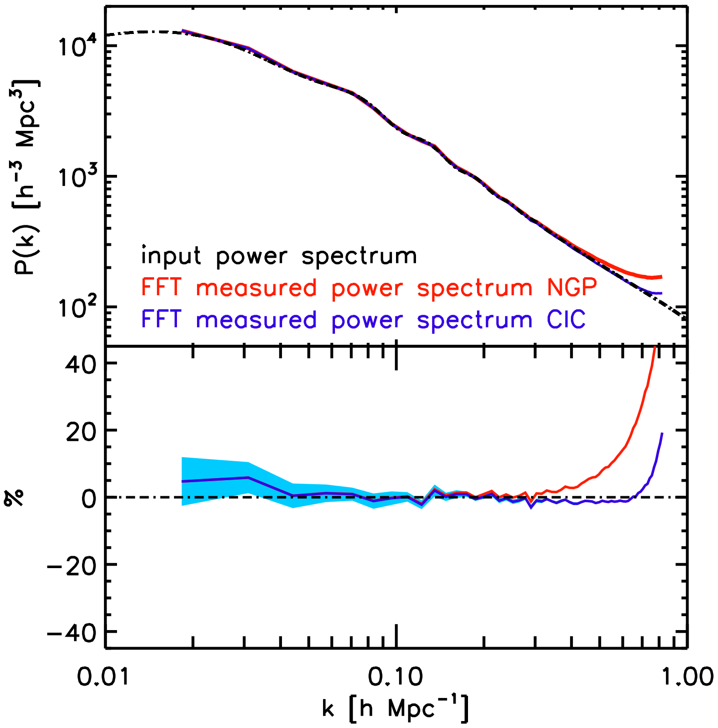

Figure 3.3: *Top panel*: comparison between the measured power spectrum with the number density of particles smoothed with NGP (red line) and CIC (blue line) and both corrected, as in equation 3.22 for the MAS effect. The input power spectrum, black dashed line, is over-plotted in order to discriminate between one method or the other. *Bottom panel*: ratio between the two measured power spectra and the expected input one. The CIC procedure works better respect to the NGP recovering the right amplitude and shape of the power up to higher frequencies.

3.1.3 FT vs FFT

In order to have a fair comparison between the FT and FFT technique, we measure the power spectrum from the same distribution of particles for the two techniques and compare the results with the input power.

Again we consider 30 realisations constructed from a Γ simulation with no gradient applied to the density and objects distributed into a cubical geometry. We fix for both methods the same parameters in order to have a fair comparison. The k-space grid for the FT method is the same as the one used in the previous test with $\mathrm{N} = (2 \times 32)^3 = 64^3$ grid points, $\mathrm{k_{max}} = \mathrm{k_N} = 0.41\,\mathrm{h\,Mpc^{-1}}$, $\mathrm{k_{min}} = 0.013\,\mathrm{h\,Mpc^{-1}}$ and $\Delta\mathrm{k} = \mathrm{k_{min}}$. To obtain the same grid in Fourier space for the FFT method, we fix the parameters of the FFT grid in configuration space to $\mathrm{L} = 486.4\,\mathrm{h^{-1}Mpc}$ and cell-size $\mathrm{H} = 8 \times 0.95\,\mathrm{h^{-1}Mpc}$, obtaining the same $\mathrm{k_{min}}$ and $\mathrm{k_{max}}(= \mathrm{k_N})$ as for the FT one. In order to reduce the number of iterations for the FT method, we consider a fraction of 5% of all particles even for the FFT approach smoothing them with CIC as MAS.

Figure 3.4 plots the two measured power spectra after averaging them over all the 30 realisations. The FT method is able to recover the input power spectrum up to very high frequencies. The FFT technique with the implementation of the CIC scheme, properly corrected with 3.22, starts to become inefficient at $\mathrm{k} \sim 75\%\,\mathrm{k_N} \sim 0.3\,\mathrm{h\,Mpc^{-1}}$, as we have just tested in the previous subsection. The efficiency of the FT method could be reached by the FFT choosing a thiner grid and maintaining, at the same time, the computation-time much lower than the FT method. For instance, halving the cell-size $\mathrm{H} = 4 \times 0.95\,\mathrm{h^{-1}Mpc}$ leads to an increase of the maximum frequency, where the FFT estimator is still working quite well, at $\mathrm{k} \sim 75\%\,\mathrm{k_N} \sim 0.6\,\mathrm{h\,Mpc^{-1}}$ which is higher than the maximum frequency of $\mathrm{k} = 0.41\,\mathrm{h\,Mpc^{-1}}$ (figure 3.5).

In the following of the thesis, we are going to adopt the FFT estimator because of the fast computation and use the CIC assignment scheme, keeping in mind that the efficiency of this approach is limited to modes smaller than $\mathrm{k} \sim 75\%\,\mathrm{k_N}$ once we correct for the effect of the MAS.

3.2 Cone-like geometry

Surveys such as the 2dFGRS and SDSS have a cone-like geometry with the observer positioned at the vertex and a big angular extension. The shape of modern surveys as BOSS, WIGGLEz and VIPERS is a truncated cone more extended along the redshift direction. To test the impact of the peculiar survey geometry on the recovered power spectrum and show how the power spectrum estimator should be modified, we consider cone-like distributions of points.

Our choice of cone-like geometry is motivated by the redshift survey VIPERS. We introduce the VIPERS geometry analysing mock catalogues with the same VIPERS dimension and volume drawn from the cubical Γ simulations described in appendix B.

In this section we limit our study on a W1 field geometry with an angular extension in right ascension between $\mathrm{RA_{min}} = 30.1°$ and $\mathrm{RA_{max}} = 38.8°$, $\mathrm{DEC_{min}} = -5.4°$ and $\mathrm{DEC_{max}} = -4.15°$ in declination and a redshift range $0.93 < \mathrm{z} < 1.1$. Fixed the bound-

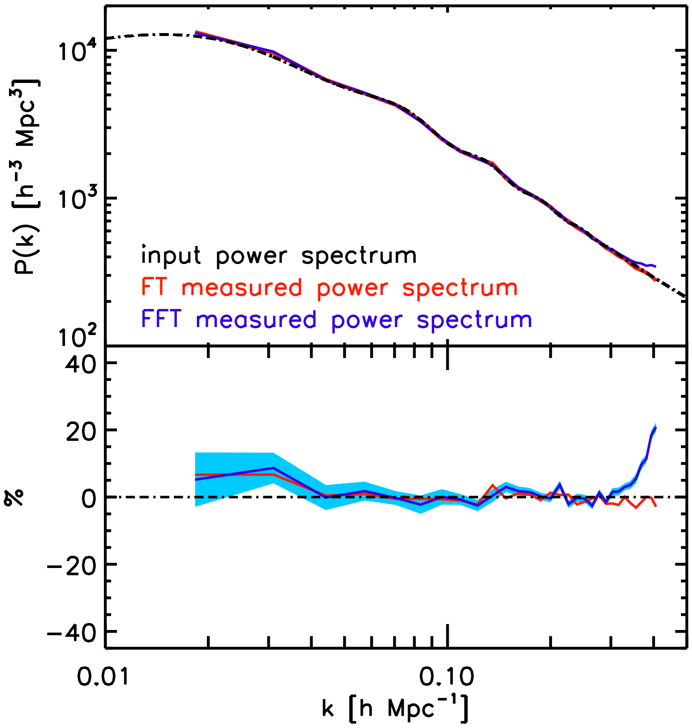

Figure 3.4: *Top panel*: measured power spectrum from 30 cubical simulations recovered with the FT approach (red line) and with the FFT one (blue line) and compared with the input power for Γ simulation (black dotted line). *Bottom panel*: ratio between the two measured power spectra and the expected one. Note that the difference between the two spectra is important only at frequencies $\mathrm{k} \sim \mathrm{k_N}$.

aries of our samples, we generate 240 independent mock catalogues cut from the 30 boxes used in the previous section. The large number of mock samples is important to reduce cosmic variance and outline systematic errors without being affected by statistical ones. Since we decide to adopt the FFT approach, the discrete distribution of particles has been firstly smoothed onto a grid. We choose a cubical grid of size $\mathrm{L} = 500\,\mathrm{h}^{-1}\mathrm{Mpc}$ fixing the minimum frequency to $\mathrm{k_{min}} = 2\pi/800 \sim 0.008\,\mathrm{h\,Mpc}^{-1}$. The value of the power measured at such large scales will not give us any information because the grid used is bigger than the size of the mock. The cell size is $\mathrm{H} = 4 \times 0.95 = 3.8\,\mathrm{h}^{-1}\mathrm{Mpc}$ fixing the Nyquist frequency to a value of $\mathrm{k_N} = \pi/3.8 \sim 0.8\,\mathrm{h\,Mpc}^{-1}$. Objects inside the volume of the survey are assigned to cells with the CIC scheme. The density contrast is calculated with equation 3.11 comparing the weighted number of galaxies in each cell with the mean number density $\bar{\mathrm{n}}$ which, for the moment, is constant. Again $\bar{\mathrm{n}}$ is measured as the

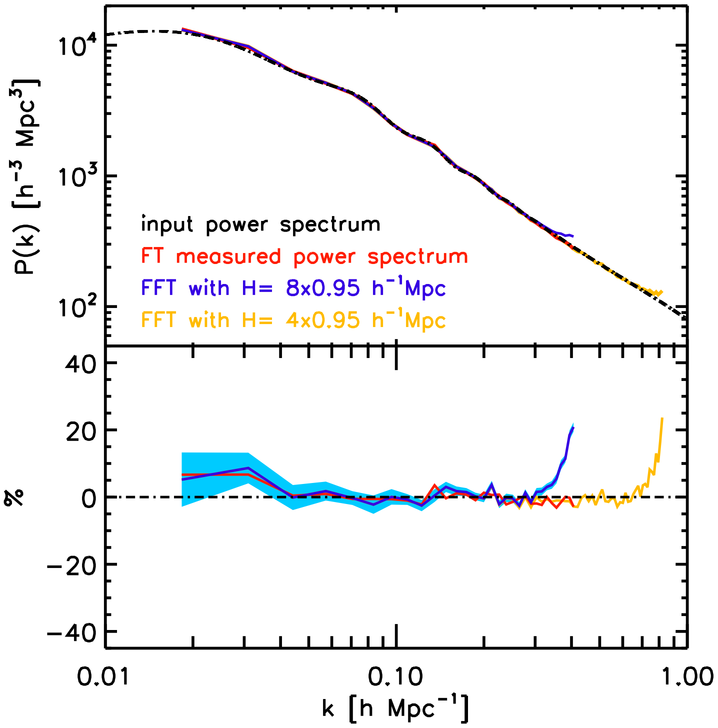

Figure 3.5: *Top panel*: same figure as 3.4 with the over-plotting of another averaged power spectrum (orange line) recovered with the FFT approach, CIC assignment with related MAS correction and a finer grid with size-cell $\mathrm{H} = 4 \times 0.95\,\mathrm{h}^{-1}\mathrm{Mpc}$. *Bottom panel*: ratio between the measured and input power spectrum. The power calculated with the finer grid in configuration space recovers the expected power up to $\mathrm{k} \sim 0.6\,\mathrm{h\,Mpc}^{-1}$.

ratio between the total number of galaxies with respect to the volume of the survey $\mathrm{V_S}$,

$$\mathrm{V_S} = \int_\mathrm{r}\int_\psi\int_\theta \mathrm{r}^2 \sin\theta\, \mathrm{d}\psi\, \mathrm{d}\theta\, \mathrm{dr} = \Delta\psi \left[\frac{\mathrm{r}^3_\mathrm{max} - \mathrm{r}^3_\mathrm{min}}{3}\right]\left[\sin\theta_\mathrm{max} - \sin\theta_\mathrm{min}\right], \qquad (3.23)$$

where $\mathrm{r_{min}}(\mathrm{r_{max}})$ is the minimum (maximum) comoving distance associated to redshift z=0.93 (z=1.1), $\theta = \pi/2 - \mathrm{DEC}$ and $\psi = \mathrm{RA}$.

The volume of these samples is $\mathrm{V_S} \sim 5.6 \times 10^6\,\mathrm{h}^{-3}\mathrm{Mpc}^3$ and the mean constant density averaged over 240 mock catalogues is of the order of $\bar{\mathrm{n}} \sim 0.003\,\mathrm{h}^3\,\mathrm{Mpc}^{-3}$.

Since the cubical grid is set up on top of the light cone, the assignment scheme is characterised by three regimes: the cell is completely outside the cone, the cell is completely inside and the cell crosses the cone. In the first case, we assign the value of 0 to the over-density field δ while, in the second case, we simply calculate the density field as for the cubical geometry. Cells at the edges of our sample that are partially outside and partially inside represent a problem. In average, the number of galaxies in those cells is lower with respect to cells completely inside the sample modifying the observed

clustering. This will introduce spurious correlation on scales comparable to the distance between these cells, mainly at scales comparable to the size of the cone.
A quick solution could be to avoid those cells with the consequence of losing information. Another is to calculate the fraction of the cell-volume inside the survey and weigh each cell for that value. We decide to follow the literature generating a random unclustered catalogue with the same geometry of the sample in order to mimic the mean density of the sample in each cell. The number of galaxies in each cell is compared to the number of random points in that cell: the number of particles in cells at the boundaries of the sample are therefore affected in the same way. The number of unclustered objects should be big enough to cover the whole volume of the survey with high precision: the mean particle distance λ should be much lower than the size of the cell $\lambda << \mathrm{H}$.
The estimator of the density contrast sampled at each point of the grid becomes

$$\delta(\mathbf{x}_\mathrm{P}) = \frac{\mathrm{n_G}(\mathbf{x}_\mathrm{P}) - \alpha \mathrm{n_R}(\mathbf{x}_\mathrm{P})}{\bar{\mathrm{n}}} \,, \tag{3.24}$$

where the subscripts G and R stand respectively for galaxy and random points while α is the ratio between the total number of galaxies and the total number of random points, $\alpha = \mathrm{N_G}/\mathrm{N_R}$. The number density $\mathrm{n_G}$ and $\mathrm{n_R}$ are measured by smoothing the distribution of mock galaxies and random points with the chosen assignment scheme (NGP, CIC) into a grid. Once we compute the density contrast with expression 3.24, the estimator of the power spectrum does not change, following the procedure described before with the correction for the MAS and the subtraction of the shot-noise term. Due to the introduction of the random catalogue, the shot-noise term is increased by

$$\mathrm{P_{SN}} = \frac{1+\alpha}{\bar{\mathrm{n}}} \,. \tag{3.25}$$

Another reason to generate a large number of random objects with $\mathrm{N_R} >> \mathrm{N_G}$ is to minimise the contribution of the random catalogue discreteness to the shot-noise term. In our case, we construct a random catalogue with 20 times more objects than the corresponding mock sample enhancing the shot-noise term by only a 5% factor, $\mathrm{P_{SN}} \sim \frac{1+5\%}{0.003} \sim 350\,\mathrm{h^{-3}Mpc^3}$.

The effect of the W1 cone-like geometry on the averaged power spectrum over 240 mock catalogues, plotted in figure 3.6, is strong at scales comparable with the size of the survey $0.01 < \mathrm{k} < 0.2\,[\mathrm{h\,Mpc^{-1}}]$ and negligible at smaller scales $\mathrm{k} > 0.2\,\mathrm{h\,Mpc^{-1}}$. From a theoretical point of view, this suppression of the power at large scales is well explained by the convolution between the input power spectrum and the window function that accounts for the limited size of the sample

$$\mathrm{P_{obs}}(\mathbf{k}) = \int \mathrm{P_{input}}(\mathbf{k}')|\mathrm{W}(\mathbf{k}-\mathbf{k}')|^2 \frac{\mathrm{d^3k'}}{(2\pi)^3} \,, \tag{3.26}$$

with W the window function of the survey (cone) in Fourier space. The decreasing of the power depends on the peculiar shape of the analysed sample and on the anisotropy of the distribution of galaxies.

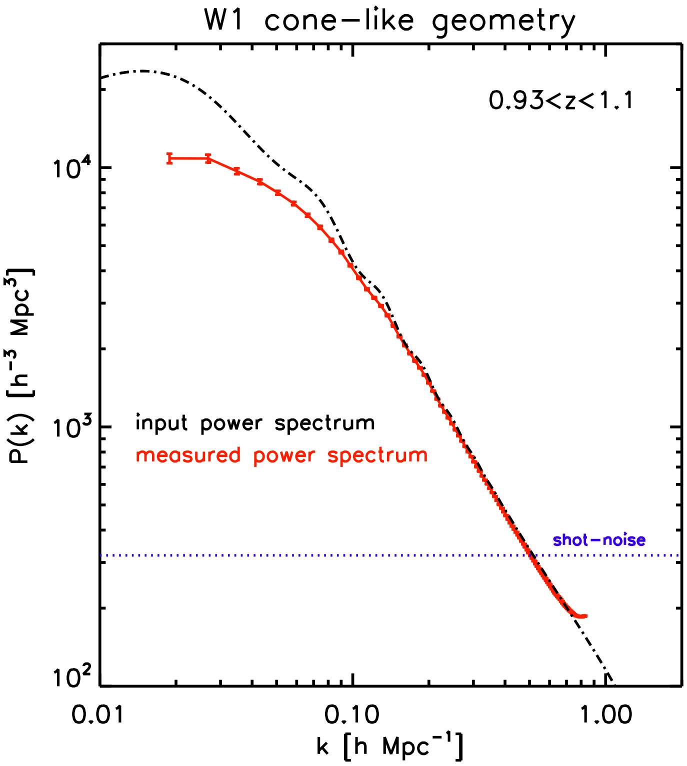

Figure 3.6: measured power spectrum from 240 mocks with the same VIPERS W1 cone-like geometry and redshift range $0.93 < z < 1.1$ (red line). The black dot-dashed line represents the input power spectrum of the simulation at a mean redshift of $\langle z \rangle = 1.02$. The effect of the VIPERS peculiar geometry is clearly visible at large scales $k < 0.2\,h\,\mathrm{Mpc}^{-1}$.

Compact surveys with a geometry well extended in all the three directions as the 2dF-GRS and the SDSS (or simply the cubical box studied before) undergo a negligible window function effect even at large scales. The VIPERS-like anisotropic geometry however, very narrow in one direction, leads to an important modification of the recovered shape of the power spectrum. We will discuss more about the window function problem and how to account for that in chapter 4.

3.3 Radial selection

Magnitude-limited surveys are affected by radial selection effects, which result from the inability of the survey to detect fainter galaxies at large redshifts. This selection effect introduces a gradient in the density distribution of surveyed objects. In the following we are going to explain how this impacts the power spectrum measurements.
Usually the redshift distribution of galaxies is computed dividing the sample in redshift (or comoving distance) bins and counting the number of galaxies in each bin. The his-

togram is then smoothed with a Gaussian filtering in order to mix over- and underdense regions and recover the true mean density. The scale at which fluctuations should be mixed up to obtain a good estimator of the mean density at a particular redshift depends on survey properties. The choice of the variance σ in the Gaussian filtering is not trivial and the power spectrum estimator is very sensitive to this scale. In section 6.2.1, we will analyse the impact of these free parameter in the power spectrum estimator.
For the moment, our purpose is to improve the power spectrum estimator with the introduction of a redshift-dependent mean density that we know a priori. We apply the known radial selection function $\bar{n}(r)$ of figure 3.7 to our set of 240 mock catalogues removing more galaxies at higher redshift following the input distribution. FKP implement a power spectrum estimator, based on the FT method, to include the non-constant mean density typical of magnitude-limited sample. We follow the FKP procedure adapting their method with the FFT algorithm.

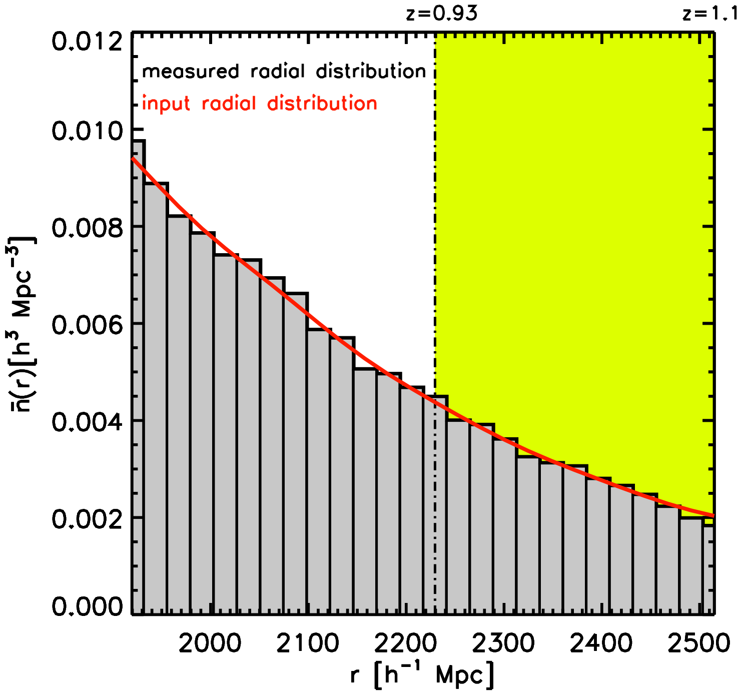

Figure 3.7: input radial distribution of galaxies applied to the 240 mock catalogues constructed with Γ simulation (red line). The grey-filled histogram is the measured radial density averaged over all the 240 samples with a bin size of $\Delta r \sim 25\,h^{-1}$Mpc.

Since the unclustered catalogue is used to mimic the mean density and must reproduce the features of the data, the first step is to generate a random catalogue with the same geometry of the sample including the mean decreasing density which, in our particular case, is known a priori. The second step is to introduce a normalisation factor N in the denominator of the density contrast (3.24) which accounts for the redshift-dependence of $\bar{\mathrm{n}}$.

The normalised FKP density field is therefore defined as

$$\delta(\mathbf{x}) = \frac{\mathrm{n_G}(\mathbf{x}) - \alpha \mathrm{n_R}(\mathbf{x})}{\mathrm{N}} , \tag{3.27}$$

where the normalisation N is given by

$$\mathrm{N} = \int \bar{\mathrm{n}}^2(\mathbf{x})\, \mathrm{d}^3\mathrm{x}' = \alpha \sum_{\mathrm{R}} \bar{\mathrm{n}}(\mathbf{x}_\mathrm{R}) . \tag{3.28}$$

Again the number density of galaxies $\mathrm{n_G}(\mathbf{x})$ and random points $\mathrm{n_R}(\mathbf{x})$ have to be smoothed onto a grid in order to perform the Fourier Transform of the field with an FFT.

With the new definition of the density field, the 3D power spectrum estimator sampled on a k-space grid becomes, as before,

$$\mathrm{P}(\mathbf{k}) = |\mathrm{FFT}\, \delta(\mathbf{x})|^2 - \mathrm{P_{SN}} , \tag{3.29}$$

where the shot-noise term is now given by

$$\mathrm{P_{SN}} = \frac{1+\alpha}{\mathrm{N}^2} \int \bar{\mathrm{n}}(\mathbf{x})\, \mathrm{d}^3\mathrm{x} = \frac{(1+\alpha)\, \mathrm{N_G}}{\mathrm{N}^2} . \tag{3.30}$$

The second equality of equations 3.28 and 3.30 is due to the fact that the unclustered catalogue mimics the mean density distribution of galaxies, so we use a convenient way to perform the spatial integrals: $\int \mathrm{d}^3\mathrm{x}\, \bar{\mathrm{n}}(\mathbf{x}) ... \rightarrow \alpha \sum_\mathrm{R} ...$ where the sum is over the random catalogue which we assume are sufficiently numerous to define $\bar{\mathrm{n}}$.

We measure the power spectrum from the 240 Γ mock catalogues with the VIPERS W1 cone-like geometry, in a redshift range $0.93 < \mathrm{z} < 1.1$ and the decreasing radial selection included. The power is estimated with the procedure just described using the FFT-grid parameters $\mathrm{L} = 800\, \mathrm{h}^{-1}\mathrm{Mpc}$ and $\mathrm{H} = 4 \times 0.95\, \mathrm{h}^{-1}\mathrm{Mpc}$ and smoothing particles with the CIC scheme.

Figure 3.8 shows the comparison between the recovered power spectrum before and after applying the decreasing mean density to the same set of mocks. The power estimated with the FKP method is in good agreement with the power estimated before the inclusion of the radial selection function, meaning that the FKP estimator perfectly treat the radial-dependent density of galaxies.

In the next section, we generalise the estimation by taking into account a possible weighting of the density contrast.

3.4 Weighting galaxies

Suppose each galaxy contributes to the density field with a weight $\mathrm{w}(\mathbf{x})$. The power spectrum estimator should be slightly modified including the weighting scheme. The

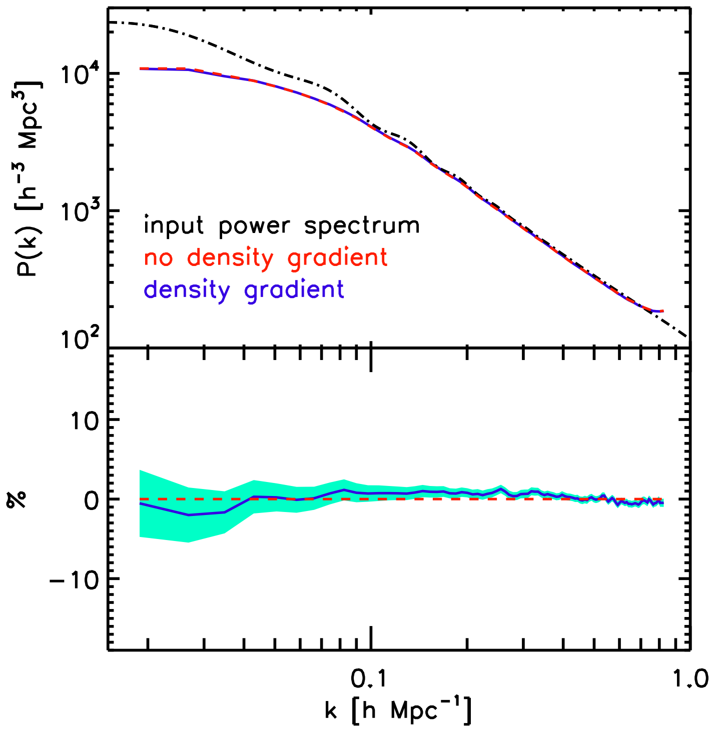

Figure 3.8: *Top panel*: comparison between the measured power spectrum averaged over 240 mock catalogues with a mean constant density (red line), the one recovered from the same set of mocks once a decreasing radial distribution is applied (blue line) and the input power of the Γ simulation (black dot-dashed line). *Bottom panel*: ratio between the power spectrum measured before and after the inclusion of a non-constant radial distribution.

density field is improved to account the galaxy weight

$$\delta(\mathbf{x}) = \mathrm{w}(\mathbf{x})\frac{\mathrm{n_G}(\mathbf{x}) - \alpha \mathrm{n_R}(\mathbf{x})}{\mathrm{N}} \; , \tag{3.31}$$

where $\mathrm{n_G}(\mathbf{x})$ and $\mathrm{n_G}(\mathbf{x})$ are still the smoothed number density of the distribution of particles and random points respectively. The assignment scheme (MAS) to grid points, necessary for the FFT algorithm, should count each galaxy with $\mathrm{w}(\mathbf{x})$ rather than 1. The normalisation factor N in the denominator of the density field is modified as follows

$$\mathrm{N} = \int \bar{\mathrm{n}}^2(\mathbf{x}) \; \mathrm{w}^2(\mathbf{x}) \, \mathrm{d}^3\mathrm{x} = \alpha \sum_{\mathrm{R}} \bar{\mathrm{n}}(\mathbf{x}_\mathrm{R})\mathrm{w}^2(\mathbf{x}_\mathrm{R}) \; , \tag{3.32}$$

with the factor α, that accounts for the different number between mock galaxies and random points, now computed as

$$\alpha = \frac{\sum_{\mathrm{G}} \mathrm{w}(\mathbf{x}_{\mathrm{G}})}{\sum_{\mathrm{R}} \mathrm{w}(\mathbf{x}_{\mathrm{R}})} \,. \tag{3.33}$$

The power spectrum estimator is, as in equation 3.29, the square of the Fourier Transform of the density field minus the shot-noise term redefined as:

$$\mathrm{P}_{\mathrm{SN}} = \frac{\sum_{\mathrm{G}} \mathrm{w}^2(\mathbf{x}_{\mathrm{G}}) + \alpha^2 \sum_{\mathrm{R}} \mathrm{w}^2(\mathbf{x}_{\mathrm{R}})}{\mathrm{N}^2} \tag{3.34}$$

The power spectrum estimator that accounts for different weights for galaxies will be used to correct the effect on the clustering due to the particular sampling strategy of VIPERS.

In the next chapter we will also analyse the FKP weighting scheme introduced to minimise the variance of the power spectrum estimator.

Chapter 4

Window function

In this chapter we focus on the effect of the non trivial geometry (due to the observational strategy) on the measurement of the power spectrum. In general, the goal is to measure the galaxy power spectrum in the infinite Universe and to compare it with the predictions for the matter power spectrum computed from theory. However we are limited by observation with a distribution of galaxies in a finite volume $\mathrm{V_S}$.
The observed density contrast $\delta_{\mathrm{obs}}(\mathbf{x})$ is

$$\delta_{\mathrm{obs}}(\mathbf{x}) = \delta(\mathbf{x}) \times \mathrm{W}(\mathbf{x}) \ , \tag{4.1}$$

where $\delta(\mathbf{x})$ is the true galaxy density contrast of the whole Universe and $\mathrm{W}(\mathbf{x})$ is a three dimensional top-hat function which is equal to 1 if the region we are analysing is inside the survey ($\mathbf{x} \in \mathrm{V_S}$) and 0 if outside. The window $\mathrm{W}(\mathbf{x})$ models the fact that we are trying to infer cosmological quantities observing only a fraction of the entire Universe. The problem is enhanced in Fourier space where a simple multiplication becomes a convolution (appendix A). Therefore the measured power spectrum $\mathrm{P_{obs}}(\mathbf{k})$ differs from the true $\mathrm{P_T}(\mathbf{k})$. Formally, equation 4.1 implies that the underlying power and the window function are related via a convolution (see FKP equation 2.1.6)

$$\mathrm{P_{obs}}(\mathbf{k}) = \int \mathrm{P_T}(\mathbf{k}')|\mathrm{W}(\mathbf{k}-\mathbf{k}')|^2 \frac{\mathrm{d^3k'}}{(2\pi)^3} \ , \tag{4.2}$$

where the window function $\mathrm{W}(\mathbf{k})$ in Fourier space is simply the Fourier Transform of the window $\mathrm{W}(\mathbf{x})$

$$\mathrm{W}(\mathbf{k}) = \int_{\mathrm{V_S}} \mathrm{W}(\mathbf{x}) \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \, \mathrm{d^3x} \ . \tag{4.3}$$

Mathematically the convolution is an operation that mixes the amplitude of the power at different modes. Suppose we want to measure the power at a frequency $\mathbf{k}$: due to the convolution effect, this value does not only depend on the power amplitude at that particular mode but also on power at close frequencies with a weight given by the window function. If the window function is very narrow, the mixing of modes is negligible. In

the ideal case of $\mathrm{W}(\mathbf{k}) = \delta^{\mathrm{D}}(\mathbf{k})$, the observed power spectrum is the same as the true one: $\mathrm{P}_{\mathrm{obs}}(\mathbf{k}) = \mathrm{P}_{\mathrm{T}}(\mathbf{k})$. Vice-versa if the window is quite extended, the power at that particular $\mathbf{k}$ is affected by nearby frequencies leading to a strong convolution effects.
The extension of the window function is inversely proportional to the size of the survey. Surveys extended in all three directions have small convolution effects while smaller surveys, more elongated in one direction, will bring an anisotropic window function with a relatively strong change in power shape. Theoretically the independence of modes expressed in equation 4.4 makes the power spectrum analysis a better choice with respect to the correlation function where bins, at different scales, are correlated,

$$\mathrm{P}(|\mathbf{k}_1|)\,(2\pi)^3\,\delta^{\mathrm{D}}(\mathbf{k}_1 - \mathbf{k}_2) = \langle \delta(\mathbf{k}_1)\delta^*(\mathbf{k}_2)\rangle \; . \tag{4.4}$$

This is no longer true with the mixing of modes introduced by the convolution of the power with the window function.
The term $\frac{1}{V}$ introduced in the 3.3 estimator to express the power spectrum with the unity of a volume, could be explained with equation 4.4. The factor $\int \mathrm{W}^2 ...\ \mathrm{d}^3\mathrm{k}$ introduces in fact a normalisation that must be corrected. This term is actually equal to

$$\int |\mathrm{W}(\mathbf{k})|^2 \,\frac{\mathrm{d}^3\mathrm{k}}{(2\pi)^3} = \int_{\mathrm{V}_{\mathrm{survey}}} \mathrm{w}^2(\mathbf{x})\,\mathrm{d}^3\mathrm{x} = \mathrm{V}_{\mathrm{survey}} \; , \tag{4.5}$$

and the inverse of this should be included in the power spectrum estimator to correct the normalisation introduced by the window function.
As a result, the window function introduces on one hand a normalisation and on the other can introduce spurious mixing of modes. The latter effect modifies the shape of the power spectrum and in the next sections we will explain how to handle it.

4.1 Computing the window function

To have a fair comparison between the measured power spectrum and a theoretical model, to extract cosmological quantities, we have to compute the three dimensional convolution of equation 4.2. Before performing the convolution, we firstly have to focus on the computation of the window function using the estimator introduced by FKP. We consider different geometries in order to study how the window depends on the particular shape of the survey. The window function is the Fourier Transform of a simple top-hat function with domain extended on the volume of the survey. For simple geometries, the window function could be calculated analytically as the Fourier Transform of a three dimensional function equals to 1 inside the box and 0 outside. For complicated geometries, the analytical computation of the window is no more possible and it is substituted with a Monte Carlo technique. The first step is the construction of an unclustered catalogue with the same geometry of the survey and the same selection function. As in the estimation of the power spectrum, since the Fourier Transform is performed with the FFT algorithm, we need to enclose the information onto a grid smoothing random objects with the MAS used for the power spectrum estimation. The window function in configuration space at the grid position $\mathbf{x}_\mathrm{P}$ is defined as

$$\mathrm{W}(\mathbf{x}) = \frac{\mathrm{N}(\mathbf{x})/\mathrm{H}^3}{\mathrm{N}} \ , \tag{4.6}$$

where N the normalisation factor of equation 3.32 and $\mathrm{N}(\mathbf{x})$ the weighted number of random points in each cell after applying the MAS smoothing. The window function of equation 4.2 is the Fourier Transform of $\mathrm{W}(\mathbf{x})$ which we compute with the FFT technique.

$$|\mathrm{W}(\mathbf{k})|^2 = |\mathrm{FFT}\, \mathrm{W}(\mathbf{x})|^2 - \mathrm{W}_\mathrm{SN} \ , \tag{4.7}$$

where $\mathrm{W}_\mathrm{SN} = \mathrm{N}_\mathrm{R}/\mathrm{N}^2$ stands for the shot-noise term introduced by the discreteness of the random catalogue. Finally, since we apply the MAS, we correct for the window introduced by cells (equation 3.20) obtaining

$$\mathrm{W}_\mathrm{obs}(\mathbf{k}) = \frac{\mathrm{W}(\mathbf{k})}{\left[\mathrm{sinc}\left(\frac{\pi \mathrm{k_x}}{2\mathrm{k_N}}\right)\mathrm{sinc}\left(\frac{\pi \mathrm{k_y}}{2\mathrm{k_N}}\right)\mathrm{sinc}\left(\frac{\pi \mathrm{k_z}}{2\mathrm{k_N}}\right)\right]^\mathrm{p}} \ , \tag{4.8}$$

where again p=1, 2 respectively for NGP and CIC and the obs subscript stands for "observed".

4.1.1 Cubical geometry

To test if our described method of computing the window function is correct, we measure the window function of a cubical box with size $\mathrm{L} = 500\,\mathrm{h}^{-1}\mathrm{Mpc}$ and compare it with the expected analytic solution

$$\mathrm{W}(\mathbf{k}) \propto \sin\left(\frac{\mathrm{k_x L}}{2}\right) \sin\left(\frac{\mathrm{k_y L}}{2}\right) \sin\left(\frac{\mathrm{k_z L}}{2}\right) \ , \tag{4.9}$$

with L the size of the cube. We generate an unclustered catalogue with a quite large number density of points $\sim 0.08\,\mathrm{h}^3\,\mathrm{Mpc}^{-3}$ in order to reduce the shot-noise term, $\mathrm{W}_\mathrm{SN} \propto 1/\bar{\mathrm{n}}$. The number of unclustered objects should also be big enough to fill all cells in order to limit stochasticity in the window. Since the number of points in each cell follows a Poisson distribution with mean $\lambda = \mathrm{N_R}/\mathrm{L}^3 \times \mathrm{H}^3$, the number $\mathrm{N_R} = 10^7$ is large enough to have a very low probability of finding 0 occurrence cells.

Random points are then assigned using CIC scheme into a large cubical grid, $\mathrm{L}_\mathrm{grid} = 4000\,\mathrm{h}^{-1}\mathrm{Mpc}$, in order to sample the window with high precision (remember that the resolution of the k-space grid depends on L, $\Delta\mathrm{k} = \frac{2\pi}{\mathrm{L}}$). The cell size $\mathrm{H} = 10\,\mathrm{h}^{-1}\mathrm{Mpc}$ is quite big to limit the FFT iterations to a reasonable value of $\mathrm{N} = [\mathrm{L}_\mathrm{grid}/\mathrm{H}]^3 = 400^3$. The window is calculated following the estimator of equation 4.7, subtracting the shot-noise term and correcting for the CIC assignment scheme.

The square of the projection of the window over the three k-directions $(\mathrm{k_x}, \mathrm{k_y}, \mathrm{k_z})$ is plotted in figure 4.1. Due to the symmetry of the cubical box, the projected window function shows the same behaviour along the three directions. In the top panel, we also outline the perfect match between the measured window function and the analytical solution of equation 4.9 concluding that our numerical approach is correct. The lower panel shows the window function in logarithmic scale to emphasise the odd and even harmonics related to the size of the survey. Even for the cubical box the window function is different from a Dirac-δ generating some effects in the recovered power spectrum. The peaks in the window should mix up the amplitude of the power at different modes. However, the ratio between the first bump at k=0 and the other peaks is so big that the contribution of higher harmonics is very small.

In the case of cubical geometry, the main effect of the window function is a normalisation factor to rescale for the limited size of the sample. The mixing of modes is negligible, as one can notice in figure 3.3 where, even at frequencies comparable to the size of the survey $\mathrm{k} \sim \frac{2\pi}{\mathrm{L}}$, the estimated power spectrum averaged over 30 realisations is consistent to the input theoretical model of the simulation within statistical errors.

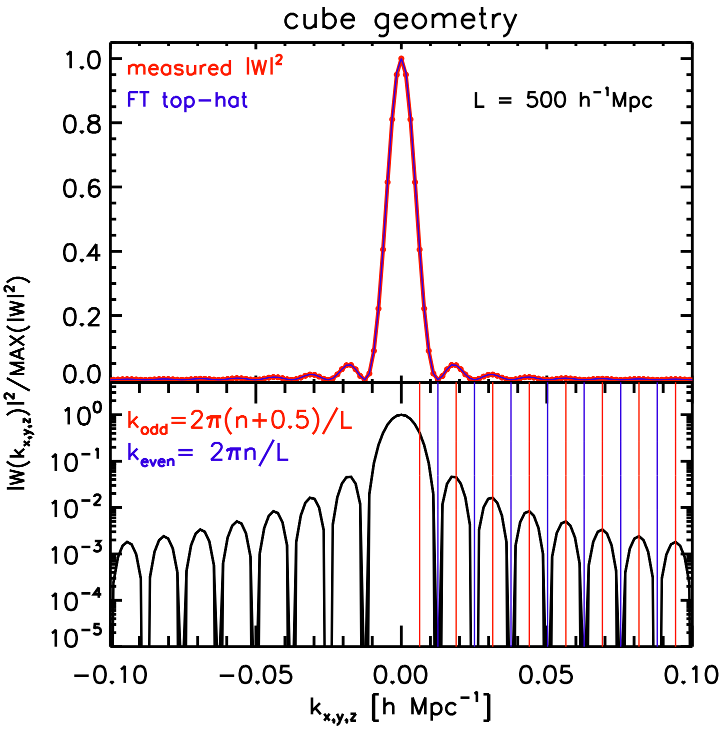

Figure 4.1: *Top panel*: projected window function along the $\mathrm{k}_{\mathrm{x,y,z}}$ directions for a cubical geometry with size $\mathrm{L} = 500\,\mathrm{h}^{-1}\mathrm{Mpc}$ (red line). Due to the symmetry of a cubical box the three window functions are completely superposed. The blue line represents the analytic solution for a cubical geometry window as in equation 4.9. *Bottom panel*: same as the top panel in logarithmic scale to outline the harmonics related to the size L.

4.1.2 Box geometry

Another simple geometry to deal with is a box with different size in each direction. As for the cubical box case, the window could be calculated analytically. However we perform again the window function numerically filling with random points a volume of $\mathrm{V} = \mathrm{L_x}\,\mathrm{L_y}\,\mathrm{L_z} = 150 \times 250 \times 350\ \mathrm{h^{-3}Mpc^3}$ and smoothing them with the CIC assignment scheme into an FFT grid with values $\mathrm{L_{grid}} = 4000\,\mathrm{h^{-1}Mpc}$ and $\mathrm{H} = 10\,\mathrm{h^{-1}Mpc}$. Computed the window as in equation 4.7, we plot the projections along the three main axes ($\mathrm{k_{x,y,z}}$) in figure 4.2. Due to the asymmetry of the box, the three window functions related to the different directions are not the same as in the case of the cubical geometry. As said before, narrower is the size of the survey in one direction and more extended is its window function. The extension of the box along the x-axes is very small compared to the other sizes causing, as expected, the broadest window. Vice-versa the size of the box in the z-direction leads to the narrowest window. To observe the effect of this box window function on the recovered power spectrum, we properly cut the 30 Γ realisations, introduced in section 3.1.1, in order to reproduce the box geometry here described. The measured power spectrum with CIC assignment and FFT-grid parameters $\mathrm{L} = 1000\,\mathrm{h^{-1}Mpc}$ and $\mathrm{H} = 5\,\mathrm{h^{-1}Mpc}$, averaged over 30 realisations, is plotted in figure 4.3 and compared to the input power of Γ simulations.

The convolution effects due to the mixing of modes are no more negligible even if the box geometry is very simple. The suppression is stronger at scales $\mathrm{k} \sim \frac{2\pi}{\mathrm{L_z}}$ because these modes are sampled only by the z-direction.

4.1.3 Cone-like geometry

The last step, before studying the sample with the VIPERS W1 cone-like shape, is to consider the behaviour of a window function for a truncated cone-like geometry. We suppose to deal with a survey quite extended along the redshift (radial) direction $\mathrm{r} = [100 - 800]\,\mathrm{h^{-1}Mpc}$ and with an angular aperture of $30^\circ \times 10^\circ$ (figure 4.4).

Our goal is to measure the window function of such a fake survey generating a cone-like random catalogue with height mimicking the redshift direction. We firstly rotate the random sample in order to align the cone with the main coordinates, aligning the radial direction with the z-axes, the angular aperture of 10° with the y-direction and the 30° angle with x, and then we compress the information on an FFT grid with the same parameters used for the box. The sizes of the sample, at the distance $\bar{\mathrm{r}}$ that halves the volume of the sample, are: $\Delta\mathrm{x} \sim 400\,\mathrm{h^{-1}Mpc}$, $\Delta\mathrm{y} \sim 100\,\mathrm{h^{-1}Mpc}$ and $\Delta\mathrm{z} \sim 700\,\mathrm{h^{-1}Mpc}$. The projected windows as a function of the three k-coordinates are plotted in figure 4.5 for the positive domain and in logarithmic scale. Again we notice that the widest window is the one corresponding to the narrowest y-direction. For this configuration, we cannot predict the position of higher harmonics because, in each direction, we do not have a fixed size due to the cone-shape geometry. Oscillations are related to a superimposition of harmonics where the dominant part is given by frequencies at $\mathrm{k_n} = 2\pi(\mathrm{n} + 0.5)/\Delta\mathrm{x}, \mathrm{y}, \mathrm{z}$.

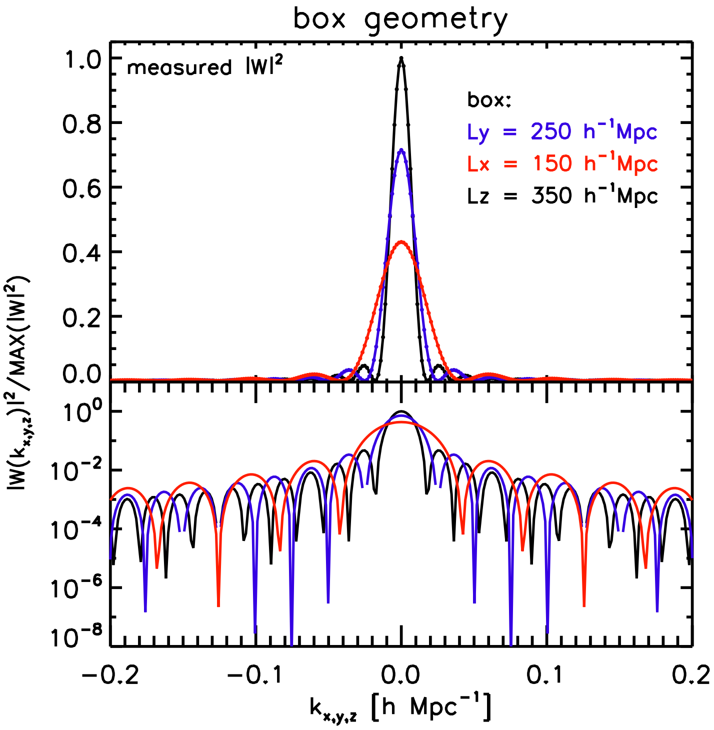

Figure 4.2: *Top panel*: projections along the directions $k_{x,y,z}$ of the window function related to a box geometry with sizes $L_x = 150$, $L_y = 250$, $L_z = 350$ [h^{-1}Mpc]. The narrowest window is related to the largest z-direction (black line), vice-versa for the x-direction (red line) and the intermediate case is represented by the y-direction (blue line). *Bottom panel*: same windows as in the top panel but in logarithmic scale. Notice the positions of the harmonic peaks (associated to different scales, L_x, L_y, L_z) which are now different for the three directions respect to the cubical geometry.

4.1.4 W1 VIPERS geometry

Finally we present the window function for the sample constructed from Γ simulation with the VIPERS W1 geometry in a high-redshift bin $0.93 < z < 1.1$. In this section, we anticipate what will be the behaviour of the VIPERS window function. Later we will explore the VIPERS window function for both the W1 and W4 fields in different redshift bins, including other features as the angular mask. We rotate the 240 samples in order to align the distribution of particles with the cartesian axes and guess the features of the window from the 3D survey extension. Figure 4.6 shows the projected distribution of galaxies in the y-z plane once we apply the rotation aligning the z-coordinate with the redshift direction of the sample.

The survey sizes at the mean redshift of the sample are $\Delta x \sim 60\,h^{-1}$Mpc, $\Delta y \sim 400\,h^{-1}$Mpc and $\Delta z \sim 300\,h^{-1}$Mpc. The main component of the superimposed

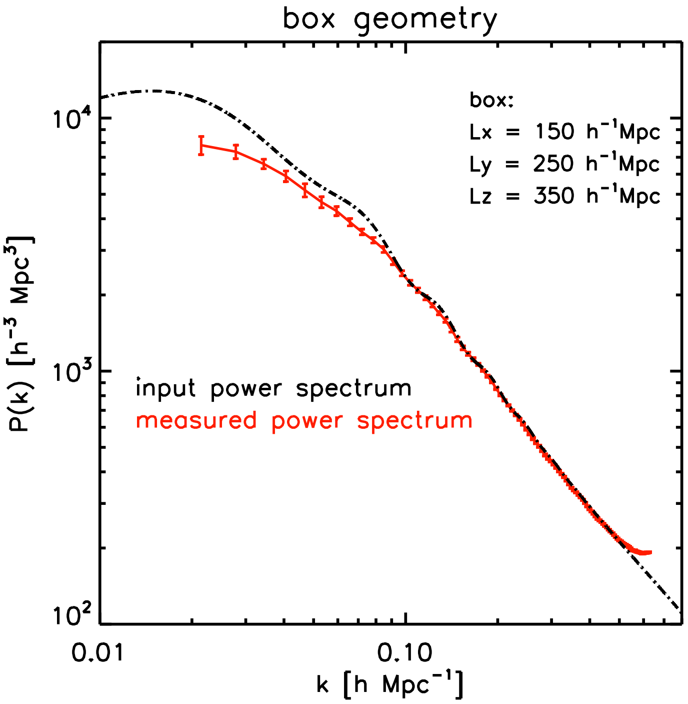

Figure 4.3: comparison between the input power spectrum of Γ simulation (black dot-dashed line) and the measured one (red line), averaged over 30 realisations with the box geometry described in the text.

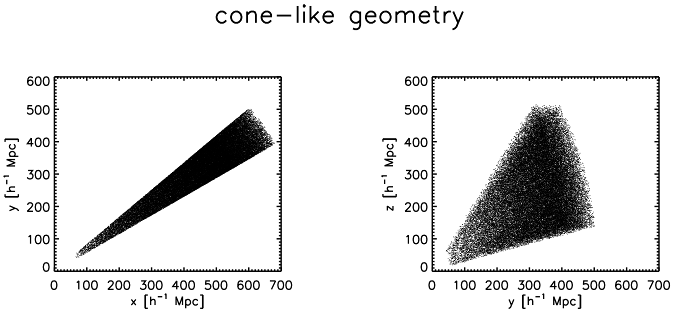

Figure 4.4: position of particles for the tested cone-like geometry in cartesian coordinates.

harmonic oscillations in the window should be related to these values. Because of the anisotropy of the survey, the convolution effects are quite strong as result from the

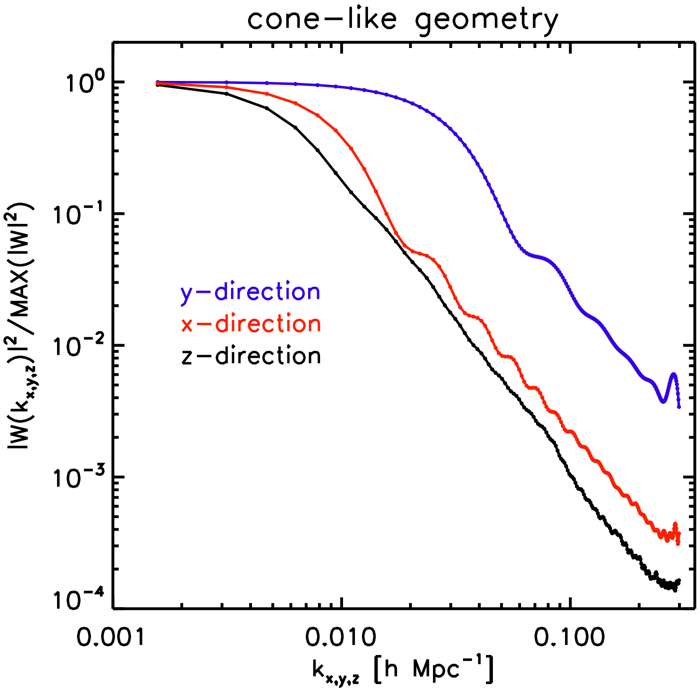

Figure 4.5: projected window function along the directions $k_{x,y,z}$ for a cone-like geometry quite extended in redshift and with a small angular aperture. The windows related to the angles of the sample (blue and red line) are broader respect to the window resulting from the redshift direction (black line).

suppression of the evaluated power in figure 3.6. The window function in Fourier space should characterise this anisotropy. We expect a behaviour similar for the y- and z-direction, where the survey has similar dimensions, and a broader window in the x-direction. We construct again an unclustered catalogue with the same geometry and with the same distribution in redshift of the mock catalogues. We align the random sample with the coordinate axes and measure the number density of random points assigning particles with CIC as MAS into an FFT grid with $\mathrm{H} = 5\,\mathrm{h}^{-1}\mathrm{Mpc}$ and $\mathrm{L} = 2000\,\mathrm{h}^{-1}\mathrm{Mpc}$.

The projected window function along the three main axes, correct for the shot-noise term and for the MAS effect, is plotted in figure 4.7. We recover exactly what we have predicted: the narrowest x-direction leads to the widest projected window while the other two directions generate similar windows. Peaks are consistent with higher order harmonics $\mathrm{k_n} = 2\pi(\mathrm{n} + 0.5)/\Delta\mathrm{x}, \mathrm{y}, \mathrm{z}$ even if there are contributions due to different $\Delta\mathrm{x}, \mathrm{y}, \mathrm{z}$ because of the cone-like geometry.

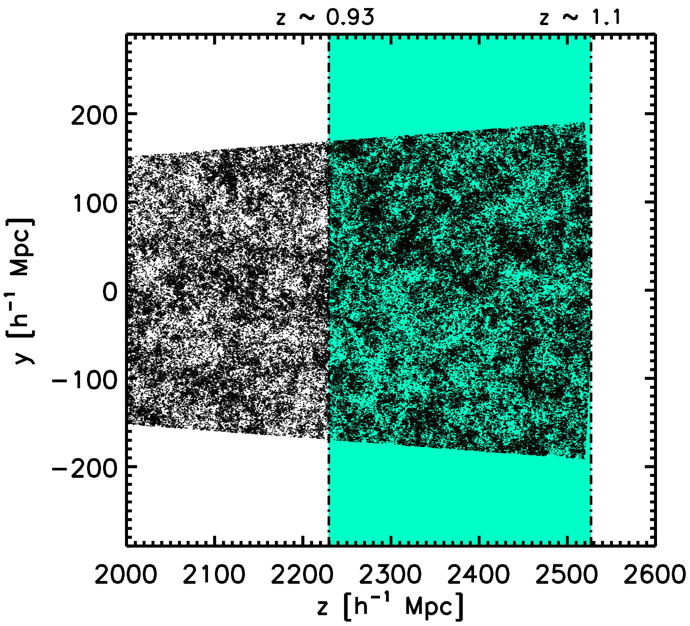

Figure 4.6: distribution of mock galaxies in one of the 240 W1 VIPERS-like samples projected along the y-z plane.

4.2 Convolving the model

A way to account for the convolution effect is to deconvolve the measured power spectrum with the window function of the survey. There are some iterative procedures such as the "Lucy deconvolution" to address this problem but they are computationally heavy processes. Even the convolution of the model is computationally time-consuming, in particular when it has to be performed repeatedly on different models. In real space, we do not have to perform the 3D integral of equation 4.2 because the operation of the convolution and the spherical average commute: we can move the operation of spherical averaging inside the integral obtaining

$$\mathrm{P}(|\mathbf{k}|) = \int \mathrm{P}(|\mathbf{k}'|)\,\mathrm{W}^2(|\mathbf{k}-\mathbf{k}'|)\,\frac{\mathrm{d}^3\mathrm{k}'}{(2\pi)^3}\ , \tag{4.10}$$

which is simpler to compute. In redshift space, this trick is not possible. We have to perform the whole integral and then spherical average the result. Since we have to deal with data in redshift space, we choose not to apply the shortcut of equation 4.10 even for real space simulated data.

Instead of calculating the whole direct convolution, we use the approach of Sato et al.

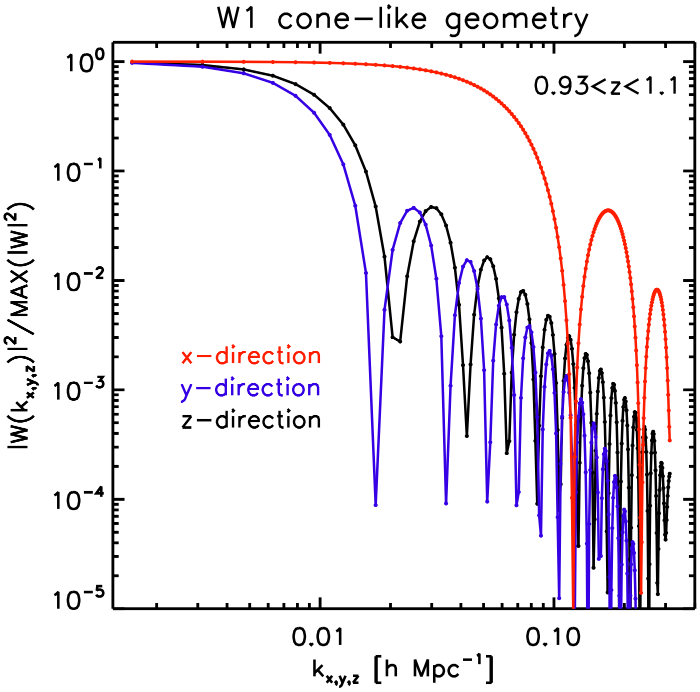

Figure 4.7: projected window function along the directions $k_{x,y,z}$ for a W1 VIPERS-like geometry in a redshift bin $0.93 < z < 1.1$ for the positive k-space domain and in logarithmic scale.

(2011), which takes advantage of the fact that a convolution in Fourier space becomes a multiplication in configuration space and vice-versa. Fourier transforming equation 4.2 and Fourier transforming back, one can obtain the following relation which is even faster with the FFT algorithm

$$P_{conv}(\mathbf{k}) = FFT^{-1}\, FFT\big[W^2 * P_M\big] - IC = FFT^{-1}\left[FFT(W^2) \times FFT(P_M)\right] - IC\,, \tag{4.11}$$

where P_M is the model power spectrum and P_{conv} the convolved power that should be compared with the measured one. To perform equation 4.11, we need to create a k-space grid domain where, at each point, the monopole theoretical power spectrum is interpolated and the window is calculated.

The subtracting term IC, added in the equation 4.11, stands for "integral constraint". It appears when we are observing a limited portion of the Universe where the mean of the density contrast could not be zero. The zero-mode amplitude of the power must have a value of zero $P(k = 0) = 0$ as a result of the definition $\langle\delta(\mathbf{x})\rangle = 0$ (because

$\delta(\mathbf{k}-\mathbf{0}) = \int \delta(\mathbf{x})\, \mathrm{d}^3\mathrm{x})$. The factor IC is (Percival et al., 2007)

$$\mathrm{IC} = \frac{|\mathrm{W}(\mathbf{k})|^2}{|\mathrm{W}(0)|^2} \int \mathrm{P}(\mathbf{k})|\mathrm{W}(\mathbf{k})|^2\, \mathrm{d}^3\mathrm{k}\ . \tag{4.12}$$

With this correction, the power is forced to be zero at frequency k=0.

4.2.1 Testing the convolution

Box geometry

To test if the convolving method just described is correct, we compare the measured power spectrum, averaged over 30 catalogues with the box geometry $\mathrm{L_x} = 150$, $\mathrm{L_y} = 250$, $\mathrm{L_z} = 350\,[\mathrm{h^{-1}Mpc}]$ of section 4.1.2, with the input power of the simulation convolved with the proper window function. The model power spectrum is corrected including the window function effect with expression 4.11 and subtracting the integral constraint. The window function W^2 for the box geometry, already computed in section 4.1.2, has been recalculated changing the FFT-grid parameters to $\mathrm{L} = 1000\,\mathrm{h^{-1}Mpc}$ and cell size $\mathrm{H} = 5\,\mathrm{h^{-1}Mpc}$ ($\mathrm{N} = 200^3$ grid points). We adopt the CIC assignment scheme to smooth the distribution of particles onto the FFT grid used for the computation of the window. The input power spectrum on the Γ simulation is the model $\mathrm{P_M}$ of equation 4.11 interpolated on a k-space grid with domain $\mathrm{k_{min}} = 2\pi/\mathrm{L}$, step $\Delta\mathrm{k} = \mathrm{k_{min}}$ and $\mathrm{k_{max}} = \pi/\mathrm{H}$. After the convolution, a spherical average is applied in order to obtain the monopole convolved power spectrum. The measured power spectrum is the same as plotted in figure 4.3. Figure 4.8 shows a good match between the corrected model and the averaged power: our approach of measuring the window function and convolving it with the model is correct up to high modes. The smallest scales with modes comparable to the Nyquist frequency are again affected by MAS problems.

W1 VIPERS-like geometry

Finally we compare the measured power spectrum of mock in a W1 VIPERS-like geometry with the convolved input model. The parameters of the FFT grid used to measure the density contrast and to compute the window function in configuration space are $\mathrm{L} = 500\,\mathrm{h^{-1}Mpc}$ and $\mathrm{H} = 2\,\mathrm{h^{-1}Mpc}$ with $\mathrm{N} = 250^3$ grid points. The Fourier-space grid, where we sample the power spectrum, the window function and interpolate the monopole input power, has parameters $\mathrm{k_{min}} \sim 0.012\,\mathrm{h\,Mpc^{-1}}$, $\Delta\mathrm{k} = \mathrm{k_{min}}$ and $\mathrm{k_{max}} \sim 1.57\,\mathrm{h\,Mpc^{-1}}$. In figure 4.9, we plot the comparison between the measured power, averaged over 240 realisations, and the input power spectrum of Γ simulation convolved with the recovered window function related to the W1 VIPERS-like geometry in a redshift bin $0.93 < \mathrm{z} < 1.1$. Again the ratio between the measurements and the corrected model at scales where the suppression of the window is strong is good within statistical uncertainties. This test makes us confident that we are able to correctly model the effect of the survey geometry even for a complicated shape as VIPERS.

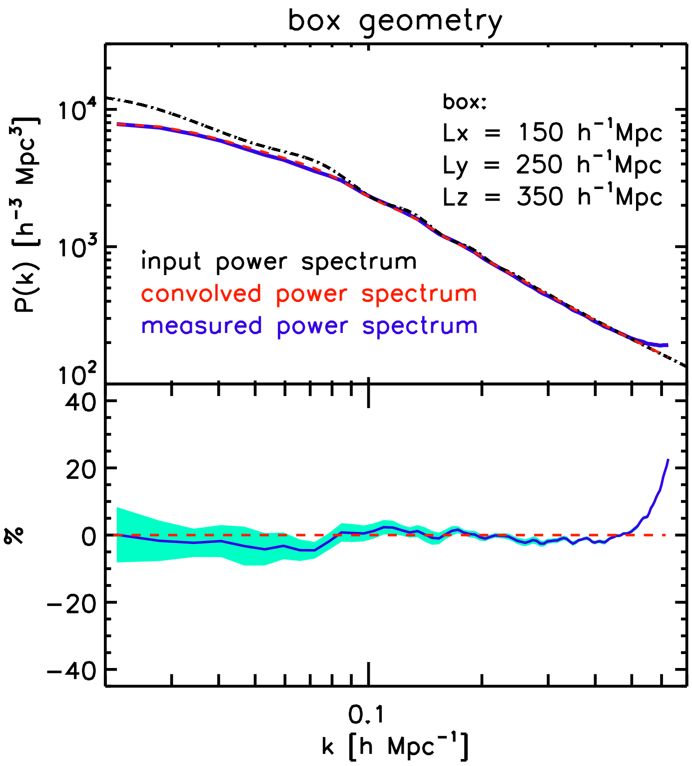

Figure 4.8: *Top panel*: same as figure 4.3 with the inclusion of the convolved input power spectrum (red line). *Bottom panel*: ratio between the measured power spectrum from the box geometry analysed in the text and the input power convolved with the proper window function. The good agreement means that the window function is correctly taken into account in our modelling.

4.2.2 FKP weighting scheme

In addition to the implementation of a new power spectrum estimator able to include non-constant mean density, FKP calculated theoretically an optimal weight to minimise the variance of the power estimator under the assumption that fluctuations are Gaussian

$$\mathrm{w}(\mathbf{x}) = \frac{1}{1+\bar{\mathrm{n}}(\mathbf{x})\mathrm{P}(\mathrm{k})} \ . \tag{4.13}$$

This weight is the analogue of the "$1 + 4\pi\bar{\mathrm{n}}\mathrm{J}_3$" weighting scheme often used in correlation analysis (Hamilton, 1993). In order to set the weighting scheme, we need to know the power P(k) which is what we want to measure. However the particular choice of P(k) should not affect too much the recovered power. We let change this parameter, keeping it constant, in order to see its impact on the recovered power spectrum. We fix the FFT-grid parameters as in the previous subsection, smooth with CIC scheme and follow the procedure for the estimation of the power of section 3.4. The galaxy power spectrum

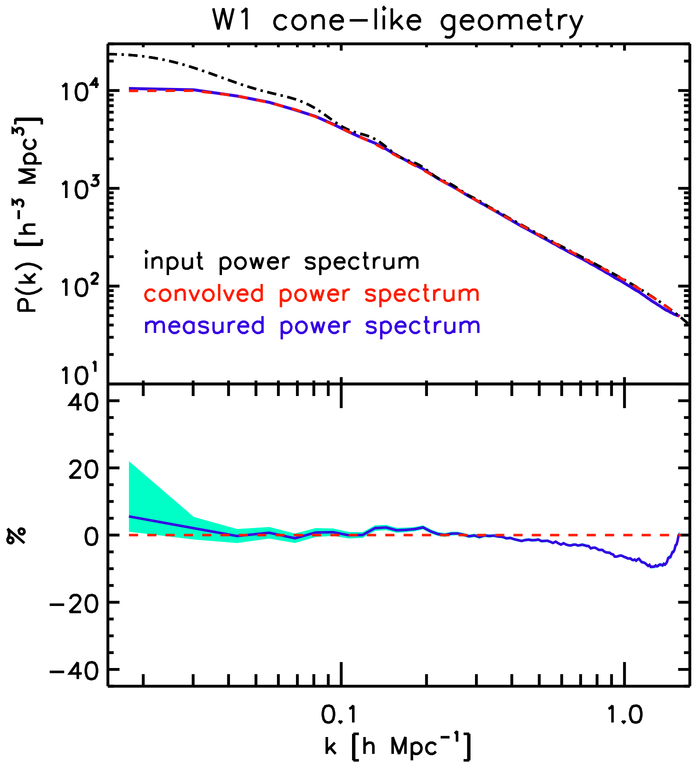

Figure 4.9: *Top panel*: comparison between the measured power spectrum (blue solid line) averaged over 240 mock catalogues with the W1 cone-like geometry and the input power (black dashed line) convolved with the window function computed for the W1 cone-like shape (red solid line). *Bottom panel*: relative error between the measured power and the model convolved with the window function to show how the window function correction is working well up to high modes.

averaged over 240 mock catalogues with W1 VIPERS-like geometry, in a redshift range between $0.93 < z < 1.1$ and with different FKP weights parameterised by the constant value $P(k) \equiv P_{FKP}$ is plotted in the top panel of figure 4.10. The best-fit value for the weight P_{FKP} in $w_{FKP}(\mathbf{x})$ should be found comparing these different observed power spectra with the convolution between the input power and the VIPERS window function. Even the computation of the window should account for the FKP weights: the window function estimator should slightly change weighting each random point for $w_{FKP}(\mathbf{x})$. The weighted number density of random objects in each cell becomes

$$W(\mathbf{x}) = w(\mathbf{x}) \frac{N(\mathbf{x})/H^3}{N} , \tag{4.14}$$

with the normalisation N given by equation 3.32.

The input model convolved with different windows, function of $w_{FKP}(\mathbf{x})$, generates different models, each of them should be compared with the proper observed power. Results are shown in the bottom panel of figure 4.10 where the ratio between the measured power with different FKP weights with the proper input convolved model is plotted. Different values for the FKP weight slightly change the observed power but the inclusion of this weighting scheme does not seem to improve the estimation of the true galaxy power spectrum. Even the variance of the power spectrum estimator does not change

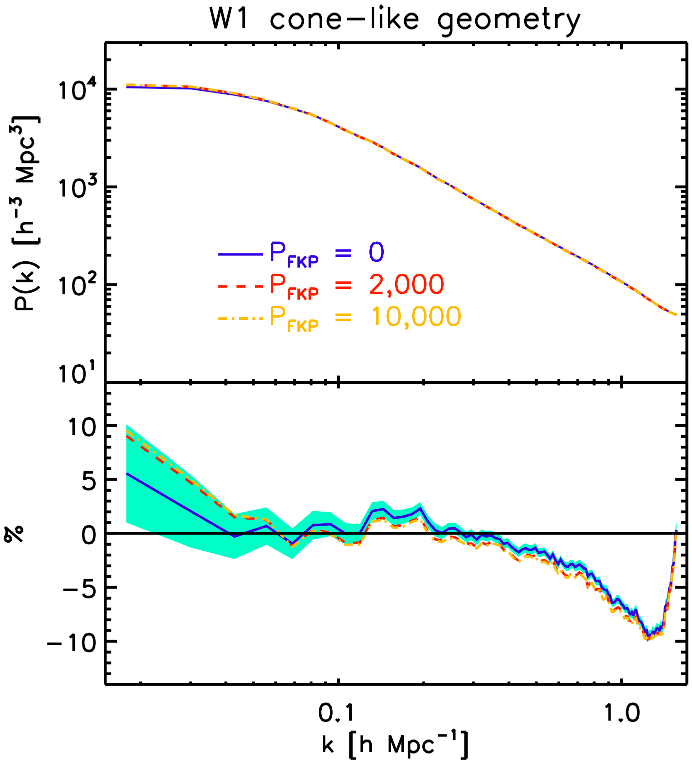

Figure 4.10: *Top panel*: measured power spectra with different FKP weights. The difference between the spectra seems to be negligible. *Bottom panel*: ratio between the power spectra with different FKP weights and the input model convolved with the proper window function.

significantly as a function of the particular FKP weight considered.

Chapter 5

Clustering estimation on VIPERS mocks

In this chapter we analyse the effects of VIPERS features on the recovered power spectrum using VIPERS-like mocks for W1 and W4 fields (de la Torre et al., 2013). We start analysing a distribution of mock galaxies with the same overall geometry of VIPERS adding at each step more realistic features to our mocks. In this way we control each possible source of systematics that could alter the shape and amplitude of the measurements.

5.1 VIPERS mock catalogues

To test the robustness of our clustering estimation we use simulated data which are designed to be a realistic match to the VIPERS sample. de la Torre et al. (2013) created two set of mock samples based on the Halo Occupation Distribution (HOD) technique. These two sets only differ by the input halo catalogue that has been used. In the first set of mock catalogues, he used the haloes extracted from the MultiDark dark matter N-body simulation (Prada et al., 2012). This simulation assumes a flat ΛCDM cosmology with parameters $(\Omega_{\rm M}, \Omega_\Lambda, \Omega_{\rm B}, {\rm h}, {\rm n_s}, \sigma_8) = (0.27, 0.73, 0.0469, 0.7, 0.95, 0.82)$.
The large volume of the simulation $V = 1\,{\rm h^{-3}Gpc^3}$ covered by $N = 2048^3$ particles is important to select independent mock samples. The haloes have been identified using a friends-of-friends algorithm with a relative linking length of $b = 0.17$ times the inter-particle separation $0.083\,{\rm h^{-1}Mpc}$. De la Torre then applied to the complete halo catalogues the algorithm presented in Carlson and White (2010) to remap halo positions and velocities in the initial simulation cube onto a cuboid of the same volume but different geometry. This is done to accommodate a maximum number of disjoint VIPERS W1 and W4 fields within the volume of the simulation. This process generated 26 and 31 independent light-cones for the W1 and W4 field respectively in a redshift range $0.4 < z < 1.2$. The light-cones are built by considering haloes from the different snapshots, disposing them according to their distance from the coordinate origin of the light-cone.
The samples are then populated with galaxies using the HOD technique that has been

calibrated on the first internal release. In this process, each halo has been populated with galaxies according to its mass, the mean number of galaxies in a halo of a given mass being given by the HOD. Finally, velocities are added to the galaxies measuring their redshift-space positions. While to central galaxies are assigned the velocity of their host halo, satellite galaxies have an additional random component for which each Cartesian velocity component is drawn from a Gaussian distribution with a standard deviation that depends on the mass of the host halo. Details about the galaxy mock catalogues construction are given in de la Torre et al. (2013).

The second set of mock catalogues, described again in de la Torre et al. (2013), is based on halo catalogues created with the Pinocchio code (Monaco et al., 2002). This code follows the evolution of a set of particles on a regular grid using an ellipsoidal model to compute collapse times and identify dark matter halos, and the Zel'dovich approximation to displace the haloes from their initial position. Due to the Zel'dovich approximation, the positions and velocities of galaxies suffer of accuracy on scales below $10\,\mathrm{h}^{-1}$Mpc. However this approach has the advantage of being very fast and can be used to generate a large number of independent halo catalogue realisations.

De la Torre created 200 independent halo mock realisations assuming the same cosmology as the MultiDark N-body simulation for both fields. The remaining steps in generating galaxy mock samples are similar to those used for the mocks based on the MultiDark simulation.

In our work, due to the better accuracy of the MultiDark mock catalogues, we make use of them to test the method and analyse the clustering of galaxies in VIPERS-like samples. However the Pinocchio mock catalogues, given the large number, are still necessary to estimate the correlation matrix in order to infer statistical errors on cosmological parameters.

The final step in obtaining fully realistic VIPERS mock catalogues is to add the detailed survey selection function which we are going to include step by step in the next sections.

5.2 The "parent" catalogue

5.2.1 Power spectrum measurements

The first step is to analyse mock catalogues with the VIPERS cone-like geometry for both fields and study the window function effect on the recovered power spectrum due to the overall geometry of the survey.

We consider the simplest configuration for the distribution of galaxies with particles inside the angular boundaries of VIPERS, a magnitude cut of $\mathrm{i}_{\mathrm{AB}} < 22.5$ applied and the effect of the colour selection on the radial distribution.

For the moment we focus on the real-space distribution of particles studying the redshift-space distortions in a second moment. The conversion between redshifts and distances has been made considering the true MultiDark cosmology, a flat ΛCDM cosmology with matter density $\Omega_{\mathrm{M}} = 0.27$, and letting the Hubble parameter h equal to 1.

In these first tests we decide to limit the redshift range between $0.6 < \mathrm{z} < 0.9$ where most of the galaxies are concentrated. The volume of the survey is of the order of

V[$0.6 < z < 0.9$] $\sim 7.1 \times 10^6\,h^{-3}$Mpc^3 for the W1 field and V[$0.6 < z < 0.9$] $\sim$ $5.8 \times 10^6\,h^{-3}$Mpc^3 for W4. Due to the different geometry between W1 and W4 that leads to different window functions, we must treat the two fields separately.
The mean density as a function of the radial comoving distance is measured dividing the whole redshift range between $0.4 < z < 1.2$ into $N = 100$ bins, counting the number of particles for each mock that fall in each bin and dividing by the volume of the bin to obtain a number density. We repeat this procedure for each mock catalogue and average over all the distribution of mocks in order to reduce the noise in the radial selection: an error in the estimation of the mean density leads to systematics in the recovered power spectrum. To further smooth the histogram, we apply a Gaussian filtering with $\sigma = 50\,h^{-1}$Mpc. We end up with two redshift distributions (figure 5.1), one for W1 and the other for W4, which are very similar in shape since the selection of galaxies is done in the same way for the two observing fields. The difference is in the lower number of galaxies in the W4 field due to the more limited angular aperture. For even further

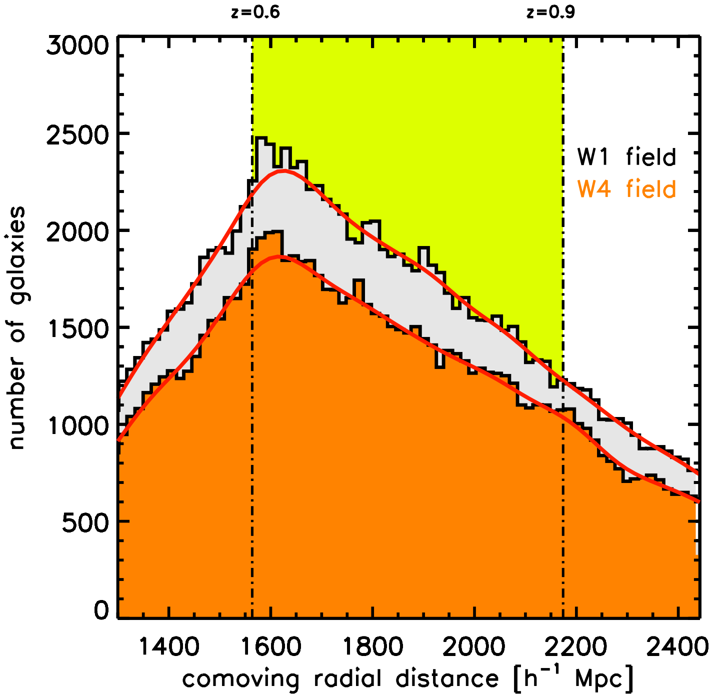

Figure 5.1: average radial distribution over 26 mock W1 catalogues (grey histogram) and 31 W4 catalogues (orange histogram) measured with a radial bin size of $\Delta r \sim 15\,h^{-1}$Mpc. The solid red lines correspond to the Gaussian filtering of the histograms with a smoothing scale $\sigma = 50\,h^{-1}$Mpc. The yellow region correspond to the radial extension used to do the analysis.

smoothing as we will see for data, we can combine the two redshift distributions even if

it is not necessary for mocks.
Based on these smoothed radial selections, we construct two unclustered catalogues, one with W1 geometry and the other with W4, to mimic the mean density of our mock samples. The number of random points is very large compared to the number of galaxies $\mathrm{N_R} \sim 75\,\mathrm{N_G}$ to fill all the volume with high precision and reduce the shot-noise contribution due to the discreteness of the unclustered catalogue.

Following the estimator described in section 3.3, we estimate for each W1 and W4 mock the galaxy power spectrum using an FFT grid with size $\mathrm{L} = 800\,\mathrm{h^{-1}Mpc}$ and cell size $\mathrm{H} = 2\,\mathrm{h^{-1}Mpc}$. Fixing these parameters, the maximum frequency we can sample is the Nyquist frequency at $\mathrm{k_{max}} = \mathrm{k_N} = \pi/\mathrm{H} = 1.57\,\mathrm{h\,Mpc^{-1}}$ and the minimum corresponds to $\mathrm{k_{min}} = 2\pi/\mathrm{L} = 0.008\,\mathrm{h\,Mpc^{-1}}$. Since the size of the grid is much larger than the size of the survey, modes with frequencies $\mathrm{k} \sim \mathrm{k_{min}}$ are not reliable. The minimum mode we trust is the one corresponding to a wavelength comparable with the largest size of the survey $\lambda_{\mathrm{max}} \sim \Delta\mathrm{z} \sim 600\,\mathrm{h^{-1}Mpc} \rightarrow \mathrm{k_{min}} \sim 2\pi/\Delta\mathrm{z} \sim 0.01\,\mathrm{h\,Mpc^{-1}}$ which could be at least sampled in the radial direction. We assign particles to the FFT grid using the CIC scheme, because from our tests results to be the most accurate method, and setting the FKP weight of equation 4.13 equals to 1 for all particles ($\mathrm{P_{FKP}} = 0$).
The 26 power spectra for the W1 field and 31 for W4 are then averaged over all the realisations ending with two independent power spectra plotted in figure 5.2. Figure 5.2 also shows a theoretical model plotted in order to outline how much the geometry of the survey affects the recovered power spectrum for both fields. The theoretical model is constructed using the CAMB (Lewis and Challinor, 2011) algorithm with HALOFIT (Smith et al., 2003) in order to include non-linearities which are dominant at smaller scales. The cosmology used to set the input parameters for CAMB is the MultiDark cosmology. The growth rate is computed to the mean redshift of $\langle\mathrm{z}\rangle \sim 0.73$ determined by averaging the position of galaxies for all mocks. The normalisation of the power is then corrected accounting for the MultiDark $\sigma_{8,\mathrm{MD}} = 0.82$: a parameter that unfortunately cannot be set with CAMB. To account for this, we simply run CAMB two times, one to generate the model at a certain redshift $\langle\mathrm{z}\rangle$ with HALOFIT, the other without the HALOFIT option at redshift 0 to recover the linear power spectrum and consequently measure the variance $\sigma_{8,\mathrm{CAMB}}$. We therefore renormalise the theoretical power spectrum as

$$\mathrm{P_{model}(k)} = \mathrm{P_{CAMB}(k)} \times \left(\frac{\sigma_{8,\mathrm{MD}}}{\sigma_{8,\mathrm{CAMB}}}\right)^2 . \tag{5.1}$$

which is reasonable on large scales, but does change the HALOFIT model. Since we are dealing with mock galaxies and not dark-matter particles, we fit by eye a linear and scale-independent bias that should give a good match at smaller scales. A simple fit by eye at large scales would lead to a mismatch due to the convolution with the window function. We obtain a value for the bias of the order of $\mathrm{b}^2 \sim 2$ for both the observing fields because, as expected, the two set of mocks are constructed from the same underlying density field.

The two recovered power spectra are slightly different at large scales because of the different window functions. The statistical errors plotted in figure 5.2 are measured with

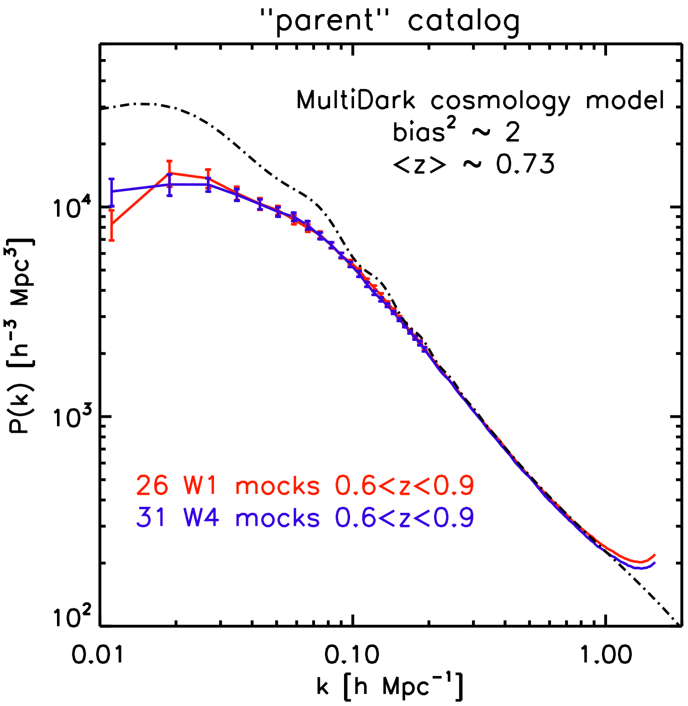

Figure 5.2: power spectrum measured from the MultiDark VIPERS mock catalogues in a redshift bin $0.6 < z < 0.9$ with only the cone-like geometry and averaged over the 26 and 31 mocks respectively for the W1 (solid red line) and W4 (solid blue line) field. Errors correspond the the 1σ dispersion over all the mock catalogues. The black dashed line corresponds to a MultiDark cosmology model generated with the CAMB algorithm at the mean redshift of the samples, $\langle z \rangle \sim 0.73$ normalised by the best fitting bias factor $b \sim \sqrt{2}$.

equation 3.9 as the dispersion of the observed power over all the mock catalogues. From this first test we notice, as outlined in section 3.3, that the overall shape and geometry of VIPERS survey strongly suppress the power at large scales affecting the shape and the amplitude. To account for this, we repeat the procedure of section 4.1.4 analysing the anisotropy of the window and adding the convolution effect to the model.

5.2.2 “Parent” window function

With this simple cone-like shape related to the overall geometry of VIPERS, the window function could be calculated analytically as the Fourier Transform of a function which is 1 inside the “parent” volume and 0 outside.

Since the true VIPERS geometry is more complicated than a simple cone-like geometry, due to the angular selection of galaxies, and we will not be able to analytically Fourier

Transform the window function, we prefer to analyse the 3D window function throughout a Monte-Carlo approach.
Working with an unclustered catalogue with the same geometry of the just analysed mock catalogues, the same radial selection and composed by a large number of points to neglect the shot-noise, we evaluate the window function for both the W1 and W4 fields in a redshift range $0.6 < z < 0.9$. Before assigning random particles to grid points using the usual CIC scheme, we rotate the random sample in order to align the main axis of the cone-like shape parallel to the cartesian coordinates: the x-coordinate is associated to right ascension, y with declination and z with redshift. The grid is the same used for the measure of the power spectrum with $\mathrm{H} = 2\,\mathrm{h}^{-1}\mathrm{Mpc}$ and $\mathrm{L} = 800\,\mathrm{h}^{-1}\mathrm{Mpc}$.

We project the 3D window function over the three directions $\mathrm{k}_{\mathrm{x,y,z}}$ to have a qualitative idea of the impact of the convolution on the estimated power spectrum. Results are plotted in figure 5.3 showing similar behaviour for both fields and a strong anisotropy over the three directions. The most problematic window is the one associated with the narrowest direction (declination). The window associated with the redshift direction instead is very narrow and will lead to a negligible convolution effect on the power. The convolution between the theoretical power spectrum generated with CAMB and the 3D window function is then performed using equation 4.11 and subtracting the integral constraint of relation 4.12. The good match in the top panel of figure 5.4 between the convolved model power spectrum and the recovered one from the mock catalogues tells us how well we can treat the geometry effects for both fields. The excellent agreement with the measured power is outlined in the bottom panel. Again the highest considered

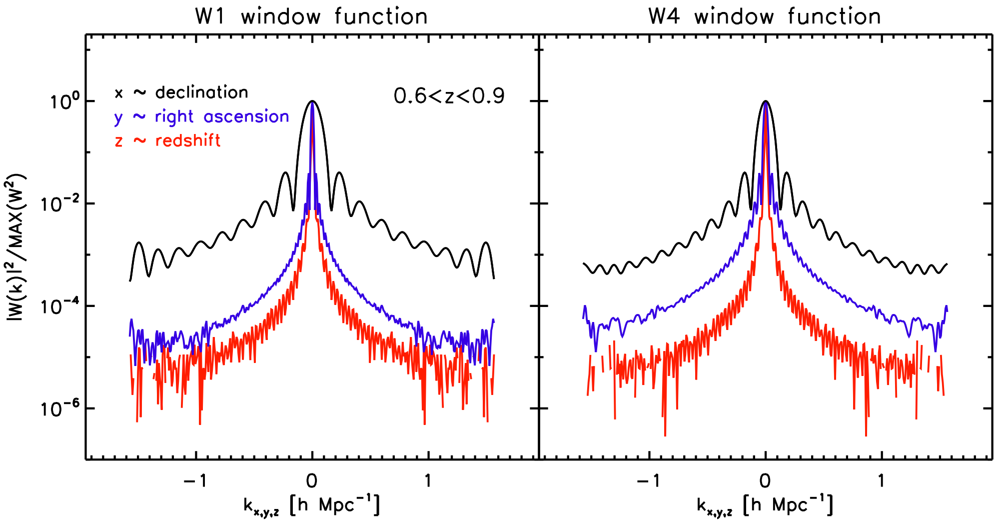

Figure 5.3: *Left panel*: projected window functions for the field W1 along the Fourier coordinates $\mathrm{k}_{\mathrm{x,y,z}}$, counterpart of the main directions of the VIPERS cone in configuration space. *Right panel*: same as the left panel but for the W4 field.

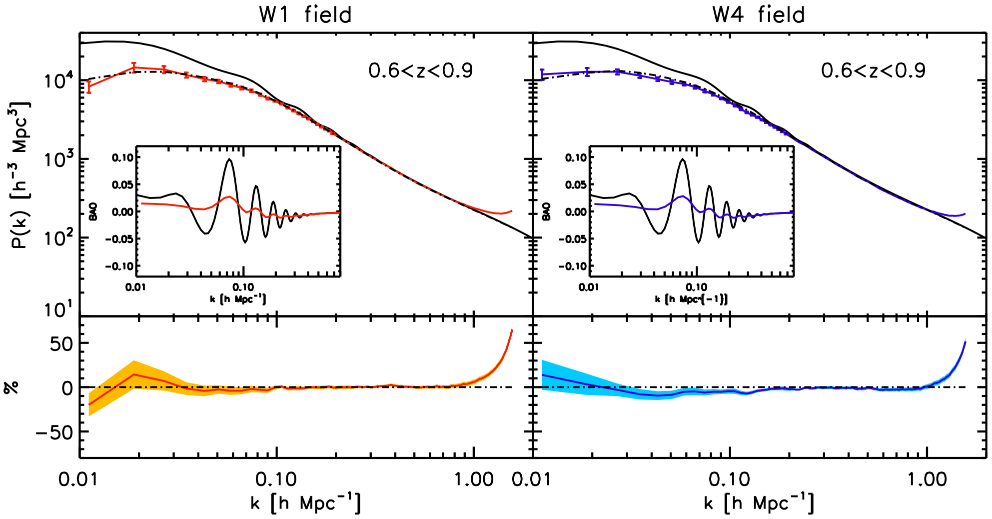

Figure 5.4: *Top left panel* : the solid line represents the theoretical model power spectrum generated with CAMB with the MultiDark cosmology, at the mean redshift of the sample $\langle z \rangle \sim 0.73$ and increased by a b^2 factor. The dot-dashed line is the model power spectrum convolved with the VIPERS-W1 window function with only the cone-like shape. The solid red line is the measured power spectrum averaged over 26 mock catalogues with statistical errors estimated from the dispersion over all the samples. The inset shows the BAO signal for the considered model power spectrum before (black line) and after (red line) applying the window. *Bottom left panel*: ratio between the measurement and the convolved model. *Right panel*: same as the left panel but for the W4 field.

modes are affected by "aliasing" and MAS problems which we choose to not correct because at those scales we do not trust the modelling of the power spectrum and, correcting the measurement, should not give us more information. Note how the amplitude of the BAO signal is severely damped, reducing to a few percent (inset).
From this first test, we conclude that, due to the VIPERS geometry, we cannot use particular features of the power (as BAO) to infer cosmological parameters but the overall shape could be determinant to set cosmological constraint in particular on the matter and baryon density.

5.3 The "mask" catalogue

5.3.1 Power spectrum measurements

The VIMOS spectrograph at VLT can observe galaxies in four quadrants. The combination of different single pointings generates a particular grid pattern on the sky with gaps between quadrants (figure 2.1). The effect of this angular selection is a loss of volume (almost one third in both fields for current data) and consequently an alteration of the window function.

In order to deal with more realistic mock catalogues, we apply the angular "Nagoya" 4.0 mask to the "parent" catalogues selecting only galaxies that fall inside the quadrants. We also apply the photometric mask to the angular distribution of mock galaxies in order to avoid, as in the VIPERS data, galaxies near very bright stars. The photometric mask does not change the distribution of galaxies cutting only a few percent of all the galaxies. To test the impact of the new window function introduced by the mask, we measure again the monopole power spectrum corresponding to these catalogues in a redshift range $0.6 < z < 0.9$. The procedure is similar to the measurements of the power spectrum for "parent" catalogues. We measure the mean density for each field averaging the number of galaxies in each bin over all the samples and Gaussian filter the histogram to obtain a smooth function. The shape and amplitude of the mean density of galaxies should not change respect to the "parent" catalogue because the number of particles which are not inside the angular mask at each redshift bin is proportional to the loss of volume. The particular geometry with gaps and quadrants is taken into account considering an unclustered catalogue with the same feature: we simply take the synthetic catalogue constructed for the "parent" mocks and apply to it the spectroscopic and photometric masks. The density contrast of equation 3.27 is calculated on a grid with the same parameters as for the "parent" catalogue assigning particles to grid points using the CIC scheme. The Fourier Transform of the density contrast, performed using the FFT algorithm, squared, corrected for the shot-noise and for the MAS effect, gives the 3D power spectrum on a k-space grid. The spherical average power spectrum in bins with size $\Delta k = \frac{2\pi}{L}$ is plotted in figure 5.5 and compared with the power measured from the "parent" mock catalogues.

The window function introduced by the loss of volume caused by the angular mask is strongly suppressing the power at all considered scales. At first sight, the effect is similar to a normalisation factor that could depend on a wrong correction of the amplitude of the power for the new volume. However, the decreasing of the power is stronger at large scales and softer at smaller scales where, increasing the sampled frequency, the "mask" power approaches the "parent" one. The behaviour of the "mask" power is compatible with a convolution between the underlying power and a window function which is broader than the one related to the "parent" geometry.

5.3.2 "Mask" window function

In this case, the window function could not be calculated analytically given the peculiar geometry introduced by the angular mask. As for the previous "parent" configuration, we adopt a Monte-Carlo approach generating a random catalogue with the same radial particle distribution as the "mask" mock catalogues and the same geometry: a cone-like shape with applied the VIPERS spectroscopic and photometric mask. After rotating the synthetic sample as usual, the 3D window function is evaluated fixing the same parameters for the grid as before and with the same procedure. Projections over the three

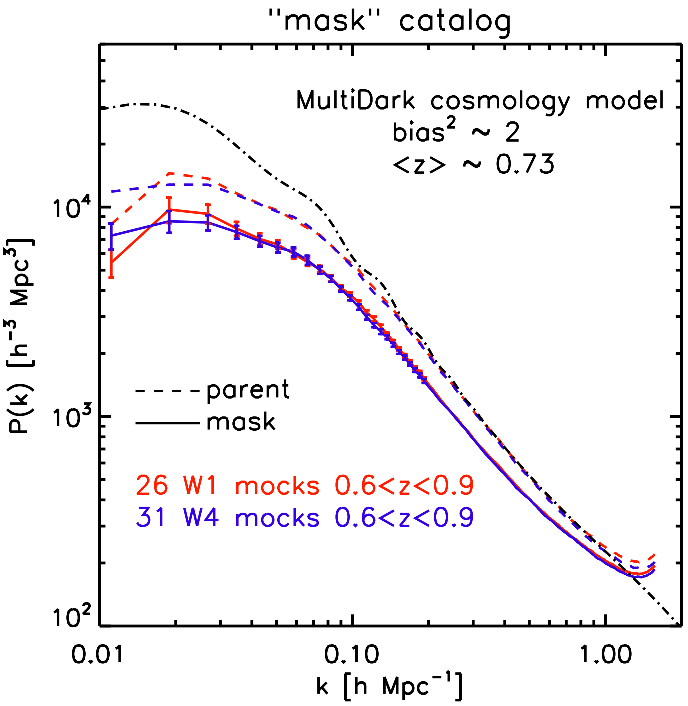

Figure 5.5: the dot-dashed black line corresponds to the MultiDark cosmology model power spectrum constructed with CAMB, while the red and blue dashed lines are the measured power for respectively the W1 and W4 field with the "parent" geometry. Solid lines shows the behaviour of the power once we apply to the "parent" samples the angular selection of galaxies for both field, W1 (red line) and W4 (blue line).

main k-coordinates (associated in configuration space with the main axes of the cone-like shape) for the W1 and W4 fields are plotted in figure 5.6. Again the most problematic direction is related to the declination leading to a broader window and consequently to a strong mixing of modes with the relative suppression of the power.
A comparison in figure 5.7 between the window function related to the W1 cone-like geometry and the window once we apply the angular mask shows that the angular selection of galaxies affects the window creating strange bumps features at different frequencies. The peaks in the window can be explained with the periodic pattern introduced by the angular mask both in declination and in right ascension. These peaks are quite broad because the particular angular scale translates into different scales as a function of redshift in configuration space. Each scale generates harmonics at different modes broadening the bumps in the window.
Again to include in the model the suppression of the power seen in figure 5.5, we perform a 3D convolution between the theoretical model and the 3D window function just

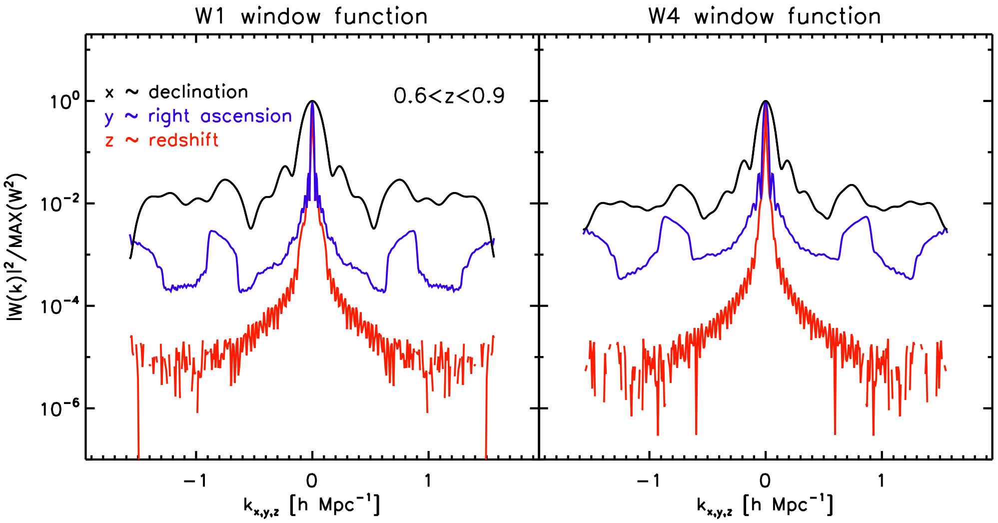

Figure 5.6: *Left panel*: behaviour of the window function projected along the coordinates $k_{x,y,z}$ for the VIPERS W1 sample with cone-like geometry and angular mask applied. *Right panel*: same as the left panel but for the W4 field.

calculated. Figure 5.8 shows our ability to model the effect of the window function even for this complicated configuration. The deviation from the model at the largest scales is due to "aliasing" and MAS problems.

5.4 Redshift space

The observed redshift of a galaxy z_{obs}, measured from its light spectrum, composed of the emission of millions of stars, is an estimate of its recession velocity caused by the expansion of space. Assuming a fiducial cosmology, this quantity could be related to the comoving distance between us and the corresponding galaxy, $z_{obs} = z_{cosm}$ with z_{cosm} the redshift uniquely determined by the distance of the object. A complication arises from the peculiar velocity of galaxies that adds, a priori, an unknown term z_{pec} in the observed redshift,

$$z_{obs} = z_{cosm} + z_{pec} \ . \tag{5.2}$$

At the cosmological scales we are interested, z_{cosm} is always much bigger than z_{pec}: z_{obs} is still a good estimate for the distance. The problem is that statistically the peculiar velocity distorts the clustering at all scales modifying the shape of the power spectrum. The coherent infall of galaxies towards over-dense regions in fact enhances the observed clustering in the large scale regime causing an increase of the estimated power. At scales comparable to clusters and superclusters, the presence of virialized structures with the consequent random velocities of galaxies suppresses the clustering damping the observed

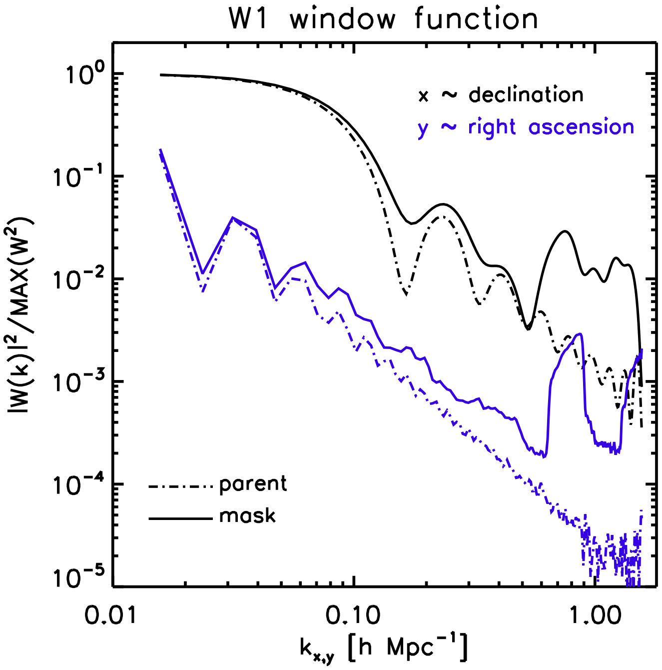

Figure 5.7: comparison of the projected window function along the DEC-direction (black lines) and the RA-direction (blue lines) between a "parent" geometry (dot-dashed lines) and the "mask" one (solid lines) for the field W1.

power spectrum. The combination of these two effects can be seen in figure 5.9 where the recovered power spectrum from VIPERS mock catalogues in real space is compared to the one measured in redshift space. Figure 5.9 shows clearly the predictions in the two regimes: at large scales the recovered power is increased while at smaller scales the expected damping is visible.

To analyse the redshift space distortions in the recovered power spectrum, we use VIPERS mock catalogues with cone-like geometry and the angular (spectroscopic and photometric) mask applied in a redshift bin $0.6 < z < 0.9$. To convert the velocity of particles in redshift, we start from the definition of the physical distance r in the comoving frame, $r = a(t)\chi$, with χ the comoving distance. Taking the derivative with respect to time, we obtain

$$\dot{r} = \dot{a}\chi + a\dot{\chi} = Hr + a\dot{\chi} = v_H + a\dot{\chi} \ , \tag{5.3}$$

where $v_H = Hr$ is the Hubble flow velocity. Since the peculiar velocity of a galaxy is defined as $v_p = \dot{r} - v_H$, the velocity in comoving space will be $\dot{\chi} = \frac{v_p}{a}$ and the variation

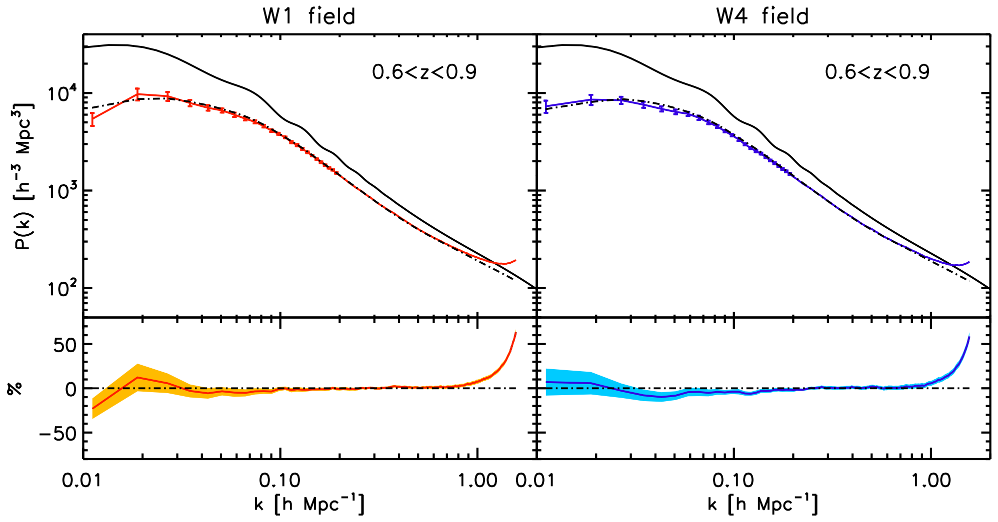

Figure 5.8: *Top left panel*: the solid line shows the theoretical power spectrum drawn with the MultiDark cosmology. A comparison between the convolved theoretical power and the measured one from W1 mock samples with the angular mask applied is also plotted. *Bottom left panel*: ratio between the convolved model and the measured one to outline the good match between measurements and theory. *Right panel*: same as the left panel but for the W4 field.

in redshift associated to the peculiar velocity $\Delta z = z_{pec} = v/c$,

$$z_{pec} = \frac{v_{\parallel}}{c}(1 + z_{cosm}) \;, \tag{5.4}$$

where $v_{\parallel}$ the line-of-sight component of the velocity.

In order to infer cosmological parameters from the comparison of the power spectrum of galaxies measured in redshift space with the corresponding theoretical prediction, we need to include redshift-space distortions in our modelling.

5.4.1 Redshift-space distortions

Since galaxies are measured in redshift space, the theoretical model should account for redshift-space distortions which alters the isotropy of the power spectrum. The real-space power spectrum P_r generated with CAMB is corrected using a simple RSD model considering the linear Kaiser term (Kaiser, 1987) which is dominant in the large scale regime and a damping term (Peacock and Dodds, 1994, dispersion model) which strongly affects the shape at smaller scales

$$P_s(k, \mu) = b^2\, P_r(k)(1 + \beta\mu^2)^2 \exp^{-(\sigma k \mu/2)^2} \;, \tag{5.5}$$

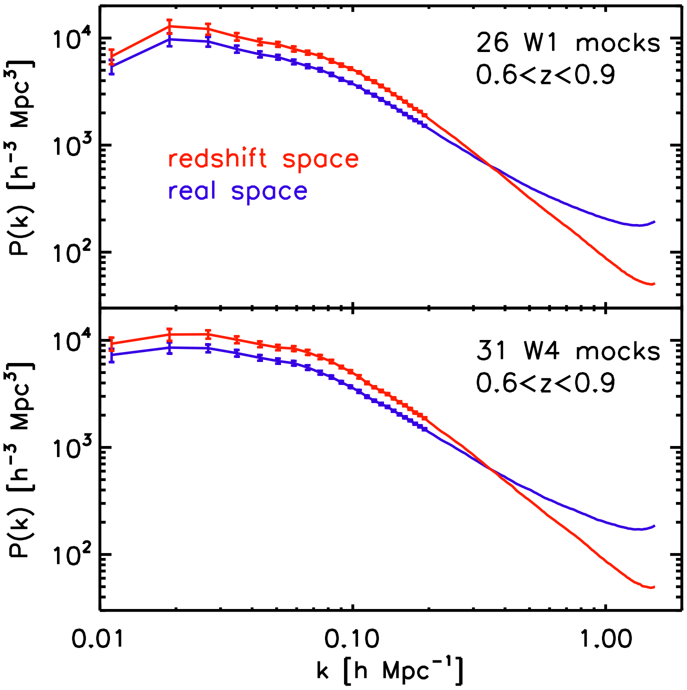

Figure 5.9: *Top panel*: measured power spectrum in real (blue line) and redshift space (red line) for the W1 VIPERS mock catalogues with the cone-like overall geometry, the angular mask applied and in a redshift range $0.6 < \mathrm{z} < 0.9$. *Bottom panel*: same as the top panel but for the field W4.

where $\mathrm{P_s}$ is the redshift-space power spectrum and μ the cosine of the angle between the vector **k** and the line-of-sight. The Kaiser factor, that accounts for the infall of galaxies toward matter-density peak, has been calculated analytically in linear theory assuming the plane-parallel approximation and is parameterised by the term $\beta = \mathrm{f/b}$ combination of the growth rate of structure f and the bias factor b. The z-direction of the FFT grid in Fourier space ($\mathrm{k_z}$) is required to be parallel to the line-of-sight in order to assume the plane-parallel approximation. Theory predicts a relation between the growth rate and the matter density: $\mathrm{f} \approx \Omega_{\mathrm{M}}^{\gamma}(\mathrm{a})$ with the γ index determinant to select between gravity theories (Steigerwald et al., 2014). Since in this work we are not interested in determining the right gravity theory, we fix the γ index to the prediction of General Relativity, $\gamma \approx 0.55$.

The exponential term in equation 5.5 describes empirically the damping of the power due to the pairwise velocity dispersion of galaxies σ_{v} in virialized structures with σ

expresses in unity of distance (Mpc)

$$\sigma = \sigma_{\rm v} \frac{1+{\rm z}}{{\rm H(z)}} \tag{5.6}$$

and accounts for the lower measured density contrast in redshift space due to Fingers-of-God. Since RSD introduce anisotropies in the power spectrum and given that the window function is not isotropic (as we have seen in the previous section), the theoretical model power spectrum $\rm P_r$ has to be corrected for RSD before doing the convolution with the VIPERS window function

$$\rm P_{CONV}(\mathbf{k}) = \int P_s(\mathbf{k}')|W(\mathbf{k}-\mathbf{k}')|^2 \frac{d^3k'}{(2\pi)^3} \, . \tag{5.7}$$

We test the RSD correction on mock catalogues comparing the measured power spectrum of figure 5.9 in redshift space with a MultiDark cosmology model generated with CAMB including RSD as in equation 5.5. The growth rate f is fixed by the MultiDark matter density and assuming General Relativity while the dispersion velocity $\sigma_{\rm v}$ is unknown a priori. Even in redshift space, the bias factor b, which is now present not only in the normalisation of the power ($\rm b^2$) but also in the β factor, is assumed to be linear and scale independent and again its value is free to vary.

The 3D convolution of equation 5.7 is performed with the same technique as in equation 4.11. After subtracting the integral constraint (4.12), we calculate the spherical average in order to obtain the monopole power spectrum. The problem due to the introduction of RSD is the nuisance parameter $\sigma_{\rm v}$ that affects the power at smaller scales. We compare in figure 5.10 the measured power spectrum from the set of MultiDark mocks (with cone-like shape and the angular mask applied) and compare it with models generated with different dispersion velocity for galaxies. For bigger value of $\sigma_{\rm v}$, the random velocity of galaxies in clustered objects is big and consequently the observed power at smaller scales is strongly damped, vice-versa for lower values of the velocity, the power is less affected.

The best fit in our modelling, in both W1 and W4 field, is $\sigma_{\rm v} = 450\,{\rm km\,s^{-1}}$ for the dispersion velocity and $\rm b \sim \sqrt{2}$ for the bias factor, which is the same value that we recovered in the real-space case. Again the inclusion of RSD are well taken into account in our modelling leading to a good match between the averaged power spectrum from mock catalogues and the theory.

5.4.2 Gaussian redshift errors

The *rms* measurement error of VIPERS redshifts has been estimated to be $\sigma_{\rm z} = 0.00047(1+{\rm z})$ or $\sigma_{\rm cz} = 141(1+{\rm z})\,{\rm km\,s^{-1}}$ which is of the order of the dispersion velocity of galaxy ($\sigma_{\rm v}/2 \sim 450/2\,{\rm km\,s^{-1}}$ for mock catalogues). The net effect of redshift errors is to smooth the clustering at small scales as the pairwise peculiar velocity of galaxies, reducing the observed power spectrum. We test the effect of adding to the redshift of particles in mock catalogues a peculiar velocity along the line-of-sight

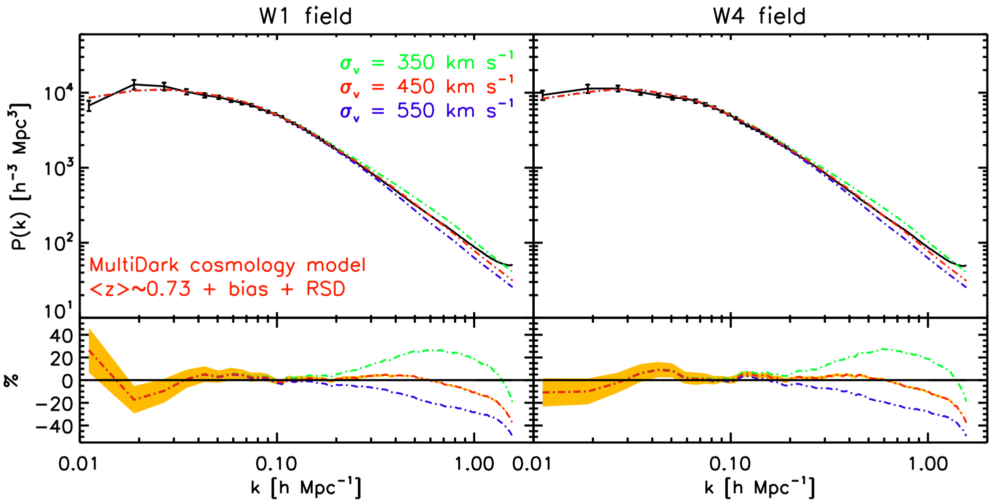

Figure 5.10: *Top left panel*: measured power spectrum from W1 VIPERS mocks with cone-like shape and angular mask applied (black solid line) compared with models characterised by different values for $\sigma_{\rm v}$ (dot-dashed lines). The model has been generated with the CAMB algorithm with the HALOFIT option, mean redshift $\langle {\rm z}\rangle \sim 0.73$ and parameters given by the MultiDark cosmology. RSD with different values for $\sigma_{\rm v}$ are included before performing the convolution with the window function. *Bottom left panel*: the ratio between the measured power spectrum and the convolved RSD models outlines the importance of the nuisance parameter $\sigma_{\rm v}$ at large scales. *Right panel*: same as the left panel but for the W4 field.

direction with intensity given by a Gaussian probability with mean 0 and deviation $\sigma_{\rm z}$. We therefore model the effect of redshift errors in the expression 5.5 by adding in quadrature the VIPERS *rms* redshift error to the peculiar velocity dispersion of galaxies $\sigma_{\rm TOT} = 2 \times \sqrt{\left(\frac{\sigma_{\rm v}}{2}\right)^2 + \sigma_{\rm cz}^2(1+{\rm z})^{-2}}$. We consider the best-fit value for the dispersion velocity of galaxies calculated in the previous subsection and, adding the value $\sigma_{\rm cz}$, we obtain $\sigma_{\rm TOT} = 530\,{\rm km\,s^{-1}}$ which we substitute to $\sigma_{\rm v}$ in equation 5.5. In figure 5.11, we plot the measured power spectrum from redshift-space mock catalogues with the addition to the position of galaxies a radial displacement drawn from a Gaussian distribution with the appropriate dispersion.

The comparison is made with the best-fit model of the previous subsection with $\sigma_{\rm v} = 450\,{\rm km\,s^{-1}}$ and a new model that takes into account the redshift error correction with $\sigma_{\rm TOT} = 530\,{\rm km\,s^{-1}}$. The correction is working well recovering the measured power spectrum up to higher modes respect to the non-corrected $\sigma_{\rm v}$ model.

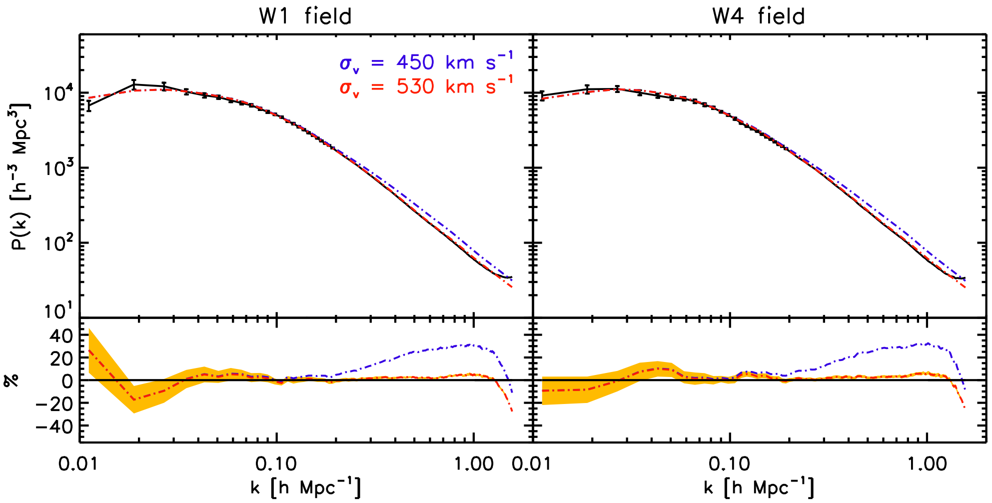

Figure 5.11: *Top left panel*: model power spectrum with (dot-dashed red line) and without (dot-dashed blue line) the inclusion of the Gaussian error in σ_v compared with the measured power spectrum from W1 VIPERS mock catalogues (solid black line). *Bottom left panel*: ratio between the measured power and the model to test by eye if the redshift error correction is working well. *Right panel*: same as the left panel but for the W4 field.

5.5 Angular sampling of galaxies

Due to the limited size of the slits, VIMOS can observe only a fraction of galaxies at each pointing. SPOC is an algorithm used to select a number of spectroscopic targets in each quadrant of the VIMOS spectrograph. Due to this practical problem, for VIPERS the fraction of observed galaxies is, in average, of the order of $\sim 40\%$ meaning that in each quadrant only 40 galaxies over 100 are selected, a priori, to obtain the spectroscopic data. The number of galaxies selected as a target for spectroscopy respect to those in the photometric catalogue is the target sampling rate and figure 2.6 shows its dependence on quadrants for both fields. To account for the angular sampling on mock catalogues and test the effect on the recovered power spectrum, we select randomly a fraction of galaxies for each quadrant proportionally to the TSR number. We compare the measured and averaged power spectrum from this new set of mock catalogues including the TSR (randomly selected) with the power recovered before correcting for the sampling. There are no differences in the two power spectra because, as one can suppose, a random sampling of galaxies should not alter the clustering of matter. This point is deeply different respect to the sampling of dark matter done by galaxies: since galaxies form in high density peaks of dark matter the clustering is enhanced by the bias factor while a random sampling does not statistically change the underlying density of matter.

The observing strategy of the SPOC algorithm is to maximise the number of objects in each quadrant. Due to this particular choice, the clustering of matter is altered because, in percentage, more objects are selected in less clustered regions giving a good sampling of the field while a lot of galaxies are missing in clustered regions: we lose angular information from the higher peaks in the density field. In this section we analyse the impact of galaxy sampling in the recovered power spectrum due to the observing strategy adopted studying the clustering of galaxies before and after applying the SPOC algorithm on mock catalogues. To test the effect of SPOC, we consider mock catalogues with all observational effects drawn from the MultiDark simulation with the VIPERS cone-like shape in a redshift bin $0.6 < \mathrm{z} < 0.9$, the angular spectroscopic and photometric mask applied, in redshift space with the inclusion of Gaussian redshift errors, TSR included and with particles selected by the SPOC algorithm. We follow again the usual steps adopted in this work to measure the power spectrum: average, separately for both fields, the redshift distribution over all the set of mocks to estimate the mean density as a function of redshift and to evaluate the radial selection, generate a random catalogue to mimic the mean density of the sample, smooth the particles over a grid (again $\mathrm{L} = 800\,\mathrm{h}^{-1}\mathrm{Mpc}$ and $\mathrm{H} = 2\,\mathrm{h}^{-1}\mathrm{Mpc}$) with the CIC assignment and measure the FKP density contrast, FFT the density field, squaring the modulus of the density in Fourier space and subtracting the shot-noise. The last step is to spherically average the 3D-grid power spectrum in shell with size $\Delta\mathrm{k} = \frac{2\pi}{\mathrm{L}}$ to obtain the monopole plotted in figure 5.12.

The comparison between the averaged power spectrum before and after applying SPOC to mock catalogues shows a suppression of the power at all considered scales. Since the selecting algorithm is not affecting the effective volume of the sample, the effect on the recovered power could not be explained by a window function effect. Even the modification of the shape is not recognisable with the effect of a window: the suppression of the power seems to be quite constant at large scales up to $\mathrm{k} \sim 0.6\,\mathrm{h}\,\mathrm{Mpc}^{-1}$ and stronger at smaller scales.

5.5.1 SPOC effects

To understand the effect of the SPOC sampling on the observed clustering, we have to ask ourselves how the selecting algorithm, in practise, is affecting the sampling. The main effects of this particular sampling are: a loss of angular area due to the size of the spectrum that covers non-negligible region of the quadrants, the anisotropy of the spectrum more extended in declination than right ascension and the cut of high density peak regions. The latter is due to the fact that in high density region fewer galaxies are selected.

The first issue should be read as a window function problem but more difficult to treat. We tried to apply SPOC to random catalogues of points in order to include this effect in the window function without any good results. The cut of regions with high density in the sky due to the sampling is the cause of the strong power suppression at smaller scales which are the most influenced by SPOC. We test this effect on simulations trying to mimic what the particular SPOC strategy is doing, cutting the density field

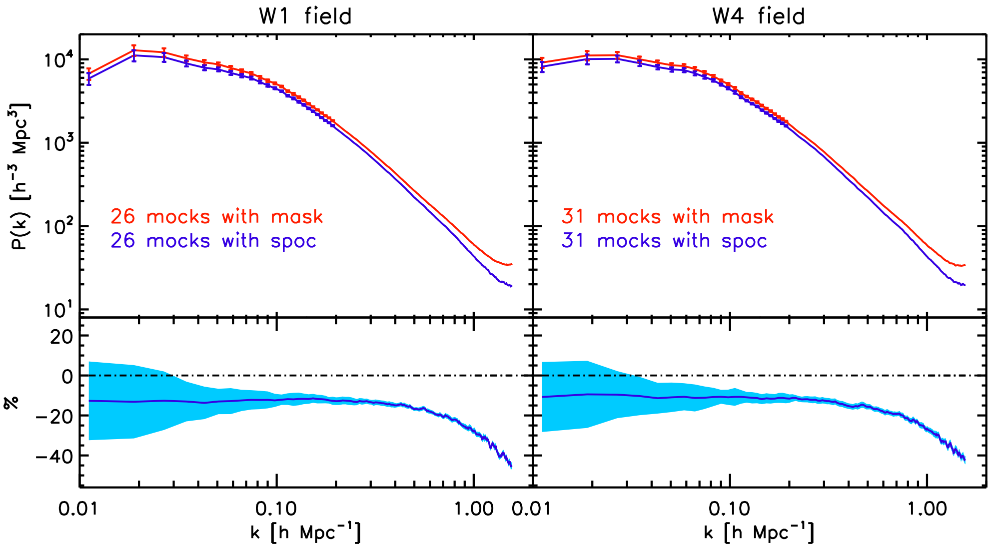

Figure 5.12: *Top left panel*: measured power spectrum before (red line) and after (blue line) applying the SPOC algorithm to the set of MultiDark mocks in redshift space. *Bottom left panel*: ratio between the two power spectra to outline the constant suppression on the power due to the SPOC strategy. *Right panel*: same as the left panel but for the W4 field.

higher than a chosen threshold for the set of MultiDark mock catalogues for W1 and W4 between $0.6 < z < 0.9$, in redshift space with the angular and photometric mask applied, with the addition of Gaussian redshift error and before applying SPOC to the samples. We measure the FKP density field of equation 3.31 with the same weight for all galaxies $w(\mathbf{x}_G) = 1$ on a grid with size $L = 800\,h^{-1}$Mpc and grid-size $H = 2\,h^{-1}$Mpc and impose $\delta_{FKP} = \delta_0$ when the normalised FKP is greater than a chosen threshold $\delta_{FKP} > \delta_0$. With this particular test, we modify the three dimensional clustering which, in principle, is different to the cut of the the projected clustering made by SPOC: however we expect to observe similar results in the recovered power. We try different values for $\delta_0 = 0.001$, 0.005, 0.008, 0.01 and for each cutting value we estimate the averaged power spectrum from mock catalogues (plotted in figure 5.13).

The constant shift of the power at large scales and the stronger suppression at smaller scales is similar to what the selection of SPOC does to the power, meaning that the main effect of the VIPERS observing strategy is the cut of high density peaks. The theory behind this is represented by the “clipped” field: ongoing works are trying to clip the highest density field in order to reduce non-linearities in the sample and deal with a more linear power spectrum (even for the VIPERS sample, Wilson et al. in preparation). The main problem related to these works is the choice of the δ_0 parameter. We conclude with this simple test that the SPOC sampling partially clips the 2D projected field making the clustering more linear.

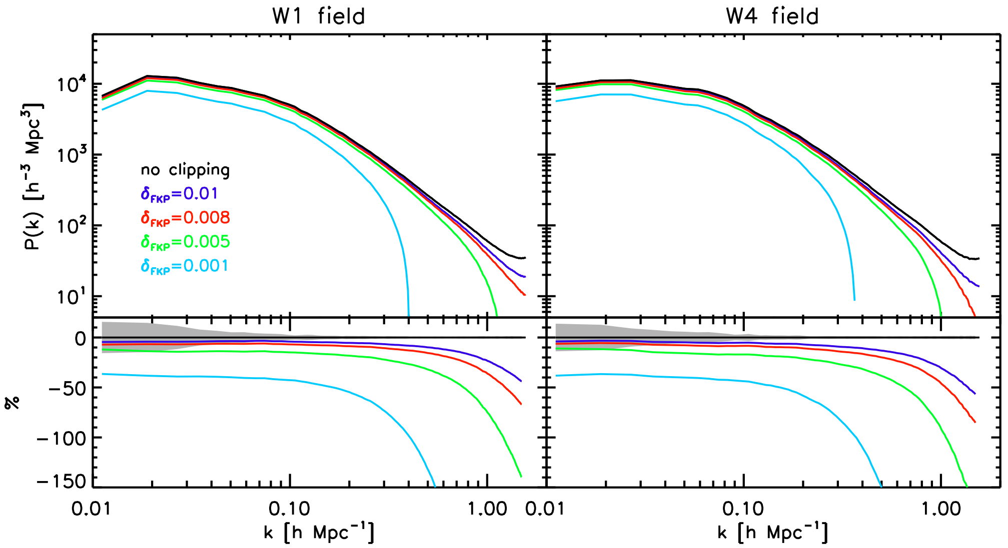

Figure 5.13: *Top left panel*: measured power spectrum from the MultiDark mocks with the W1 cone-like geometry and the angular mask applied (black line). The coloured lines represents the behaviour of the power once we clip the density field with different threshold in order to mimic the effect of SPOC in cutting the highest peak of the potential. *Bottom left panel*: the ratio between the measured power without clipping and with different threshold for the density field shows a similar behaviour on the power as the one made by SPOC. *Right panel*: same as the left panel but for the W4 field.

5.5.2 Correcting the angular SPOC selection

One way to correct for the SPOC effect on the power spectrum is to understand the theory of the “clipped” field and try to associate the cut of high density peaks to a threshold δ_0. Another problem related to this choice is to extend the “clipped” field theory to the cut of the projected clustering over the sky. And this track is solving only a part of the problem without considering the loss of volume and the anisotropy due to the slits. A second solution could be simply correct the observed power spectrum from data for the constant suppression measured from mocks. This choice depends on the particular mock samples used and in principle the clustering of matter in mock catalogues could be different respect to data.

The third solution we consider is to correct the SPOC selection up-weighting galaxies in regions where the SPOC depletion is bigger. The first trial, as applied in de la Torre et al. (2013), is to correct for the target selection up-weighting galaxies by the inverse of the TSR ($w_{SPOC} = 1/TSR$) for the quadrant in which the galaxy belongs, giving more weight to galaxies in quadrants where the effect of SPOC is stronger. Using this approach, we are trying to compensate the missing of selected galaxies in the higher density field, reconstructing the distribution of galaxies as it was before the selection

algorithm. The correcting weight is included in the estimation of the power spectrum following the power estimator of section 3.4.
We consider the final mock catalogues for both fields in redshift space and with SPOC applied and compare the measured power spectrum before and after applying this weight. The blue line of figure 5.15 shows that the correction is working in the right direction at all scales reducing the suppression of the power only by a small quantity. The problem with TSR computed in a quadrant is that it is an average over 7×8 arcmin^2. Within a quadrant there are over- and under-densities which are locally sampled differently. We can hope to get a more precise TSR by simply considering TSR in areas smaller than quadrants. In this context, Pollo et al. (2005) proposed a small scales weighting scheme that we call "local" TSR correction. They create a small circle with radius $\theta_{\rm L}$ in the angular distribution of galaxies (figure 5.14) for each SPOC target and compare the number of objects in the photometric catalogue ($\rm N_{PHOTO}$) with those selected by SPOC ($\rm N_{SPOC}$). They define a new weight for each galaxy

$$\rm w_{SPOC} = \frac{N_{PHOTO}(\theta < \theta_L)}{N_{SPOC}(\theta < \theta_L)} \; . \tag{5.8}$$

With this "local" correction one should be able to reconstruct the density field even at smaller scales. The choice of $\theta_{\rm L}$ is clearly critical and it should be comparable with the size of the slit. Keeping in mind that the size of a slit is $\Delta\rm DEC = 140''$ in declination and slightly changes in right ascension with mean $\langle\Delta\rm RA\rangle \sim 10''$, we test different choice for the angular radius $\theta_{\rm L} = 30'' - 50'' - 70'' - 90''$. We calculate the weight for each SPOC target as the ratio between the number of particles in the circle for the "mask" samples with respect to those in the SPOC samples. Results are plotted in figure 5.15 only for the best fit with $\theta_{\rm L} = 50'' - 70''$. Since the spectrum has a rectangular shape we also try with a rectangle $\Delta\rm DEC \times \Delta RA = 100 \times 60\,arcsec^2$ to compute the weight for each SPOC selected galaxy (as shown in figure 5.14). The results are encouraging showing that this choice produces a better corrected power.

The new weighting scheme is working better respect to TSR weight (as a function of quadrant) at all scales and corrects the power with an uncertainty of some percent up to scale $\rm k \sim 0.3\,h\,Mpc^{-1}$. This correction has been tested also for real-space distribution of galaxies giving the same results. The SPOC effect in data will be corrected using the "local" TSR weight of equation 5.8 considering the ratio in small rectangles with size $\Delta\rm DEC \times \Delta RA = 100 \times 60\,arcsec^2$ between galaxies in the CFHTLS photometric catalogue and those observed in VIPERS.

5.6 Covariances

The construction of a complete mock catalogues that fully characterise the selection features of the VIPERS survey is necessary to analyse statistical errors related to the power spectrum estimate. Statistical errors are calculated using the full covariance approach due to the strong correlation between bins introduced by the VIPERS window

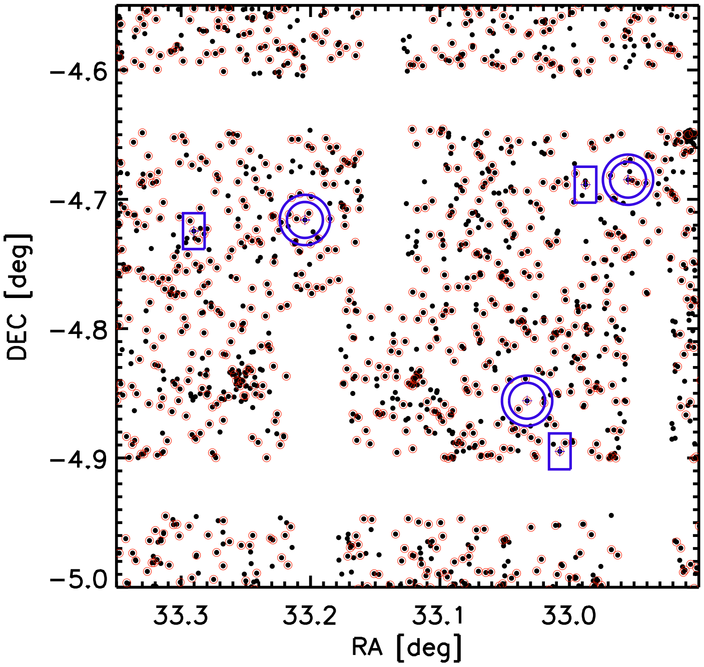

Figure 5.14: distribution of particles in the observing field W1 for a MultiDark mock sample. Black dots are galaxies in the mock catalogue with the angular mask applied while red circle points are objects selected by the SPOC algorithm. Blue lines correspond to the contour centred on SPOC galaxies for the circles with $\theta_\mathrm{L} = 50'' - 70''$ and the rectangle with $\Delta\mathrm{DEC} \times \Delta\mathrm{RA} = 100 \times 60\,\mathrm{arcsec}^2$ where the correcting weight has been computed.

function. An unbiased estimator for the covariance matrix is

$$\mathrm{C_{ij}} = \frac{1}{\mathrm{N_r} - 1} \sum_{\mathrm{l}=0}^{\mathrm{N_r}} \left[\mathrm{P_l(k_i)} - \bar{\mathrm{P}}(\mathrm{k_i})\right] \left[\mathrm{P_l(k_j)} - \bar{\mathrm{P}}(\mathrm{k_j})\right] \,, \tag{5.9}$$

calculated over a number of independent realisations $\mathrm{N_r}$ with $\mathrm{P(k)}$, the measured power spectrum of one realisation, and $\bar{\mathrm{P}}(\mathrm{k})$, the averaged power over all the realisations. Since the number of bins we are going to deal with is of the order of 10, we should estimate the covariance matrix using hundreds of mocks in order to reduce the propagation of the covariance matrix errors up to cosmological parameters (Percival et al., 2014). To avoid this kind of problems and to work with a smooth covariance without too much noise, we estimate the correlation matrix $\mathrm{r_{ij}}$ using 200 Pinocchio mocks for each field

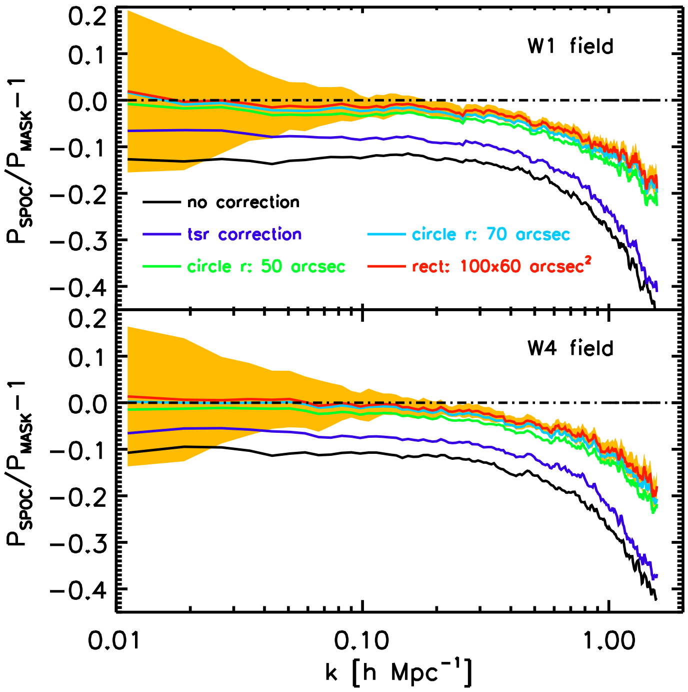

Figure 5.15: *Top panel*: ratio between the measured power spectrum before applying SPOC to W1 mock samples and after applying SPOC without correction (black line), with the TSR correction (blue line) and with the "local" TSR correction with $\theta_{\rm L} = 70''$ (green line), with $\theta_{\rm L} = 50''$ (light blue line) and $\Delta{\rm DEC} \times \Delta{\rm RA} = 100 \times 60\,{\rm arcsec}^2$ (red line). Errors are related to the dispersion of the power among all samples and are shown only for the rectangular-shape correction (orange regions). *Bottom panel*: same as the top panel but for the W4 field.

instead of considering the limited number of MultiDark mock catalogues

$$\rm r_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}\,. \tag{5.10}$$

In figure 5.16, we show the correlation matrix for both fields calculated with the final Pinocchio mock catalogues including spectroscopic and photometric masks and the distribution of galaxies in redshift space with Gaussian redshift errors included.

As a first test, we limit the covariance analysis in a k-range between $0.01 < \rm k < 0.3\,[h\,Mpc^{-1}]$ with a binning size of $\Delta \rm k = 0.01\,h\,Mpc^{-1}$ for a total of $\rm N = 29$ bins. As expected, the correlation matrix is not diagonal but strongly correlated due to the VIPERS window function that mixes power at small modes. The correlation between bins at higher modes is instead due to non-linearities. Thanks to the big number of mock catalogues, we can also notice how smooth the correlation matrix is, i.e. without too

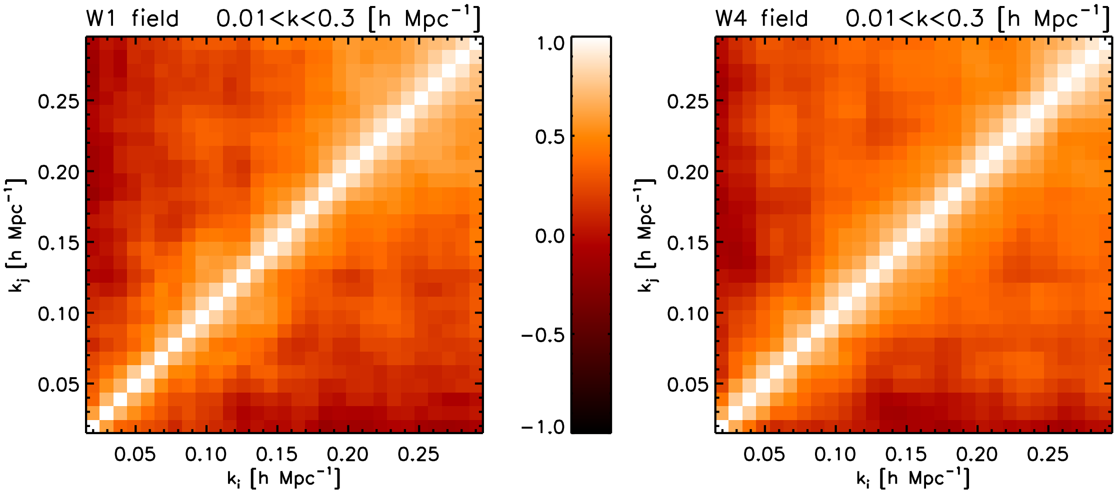

Figure 5.16: correlation matrix for both field W1 and W4 in a k-range $0.01 < \mathrm{k} < 0.3\,[\mathrm{h\,Mpc^{-1}}]$ for the estimation of the correlation between different modes computed using 200 complete VIPERS mocks generated with the Pinocchio algorithm.

many statistical fluctuations. The covariance matrix $\mathrm{C_{ij}}$ is then evaluated in a second moment multiplying the Pinocchio correlation matrix with the diagonal errors of the covariance matrix calculated from the 26 and 31 MultiDark mocks respectively for the field W1 and W4

$$\mathrm{C_{ij}} = \mathrm{r_{ij,Pin}} \times \sqrt{\mathrm{C_{ii,MD}C_{jj,MD}}}\,. \tag{5.11}$$

Having tested our modelling of RSD, the impact of the observational strategy of VIPERS on the power spectrum and finally having a good estimate of the covariance of data, we are now in position of testing on mock catalogues the cosmological inferences we expect to get from VIPERS.

Chapter 6

Recovering a known cosmology

In this section we analyse the robustness of our power spectrum estimation with simulations. Comparing different models with the actual power spectrum measured from the fully realistic mock samples, i.e. including all the angular masks, we test with a chi-square technique the goodness of our method in recovering cosmological parameters.

6.1 Likelihood method

The idea is to measure the monopole power spectrum from the final MultiDark mock catalogues, averaged over all the realisations to reduce noise and cosmic variance and fit the best theoretical model generated with the CAMB algorithm letting free different parameters.

For the moment, we limit the redshift range between $0.6 < \mathrm{z} < 0.9$ calculating an averaged redshift distribution over all the mock samples smoothed with a Gaussian filter of $\sigma = 50\,\mathrm{h}^{-1}\mathrm{Mpc}$. The density field is calculated on a grid with size $\mathrm{L} = 800\,\mathrm{h}^{-1}\mathrm{Mpc}$ and cell size $\mathrm{H} = 2\,\mathrm{h}^{-1}\mathrm{Mpc}$ smoothing particles with CIC assignment. Galaxies are weighted with the inverse of the "local" TSR estimated with a $\Delta\mathrm{DEC} \times \Delta\mathrm{RA} = 100 \times 60\,\mathrm{arcsec}^2$ rectangle on the sky in order to correct for the SPOC selection while the FKP weights are fixed to 1 for all particles. The theoretical model $\mathrm{P_M}$ is constructed with CAMB plus HALOFIT to include non-linearities fixing all the cosmological parameters to the MultiDark cosmology except for those we want to constrain and the bias factor b. The bias term is assumed to be linear and scale independent changing the amplitude of the power by a factor of b^2 (since we have fixed σ_8 to the MultiDark value and rescaled at the mean redshift of the sample $\langle \mathrm{z} \rangle \sim 0.73$). The monopole model is interpolated on a 3D k-space grid in order to add the anisotropy of RSD and perform the 3D convolution.

We account for RSD with expression 5.5 adding to the real-space power spectrum the Kaiser factor for coherent infall of galaxies toward density peaks and the exponential damping for random velocities. The velocity dispersion σ_v, with the inclusion of Gaussian redshift errors, is considered as a free parameter in our RSD modelling. We evaluated the VIPERS window function independently for the W1 and W4 field, performed the convolution with the theoretical model and subtracted the integral constraint.

Finally the resulting spherical average of the convolved 3D power spectrum can be compared through a likelihood analysis with the measured power from mock catalogues defining the chi-square as

$$\chi^2(\mathbf{p}) \equiv \sum_{\mathrm{ij}} \left[\mathrm{P}(\mathrm{k_i}) - \mathrm{P_M}(\mathrm{k_i}, \mathbf{p})\right] \mathrm{C}^{-1}_{\mathrm{ij}} \left[\mathrm{P}(\mathrm{k_j}) - \mathrm{P_M}(\mathrm{k_j}, \mathbf{p})\right] , \tag{6.1}$$

where the quantity $\mathrm{C}^{-1}_{\mathrm{ij}}$ is the inverse of the covariance matrix that has been estimated using a combination of 200 Pinocchio mocks for each field and the same set of catalogues used to recover the power spectrum, following the procedure of section 5.6. In this analysis it is important to estimate the full covariance matrix because of the correlation of modes introduced by the window function. A treatment based only on diagonal covariance matrix would lead to wrong estimation of the errors on cosmological quantities.

For these tests, we use again the mock samples used in the previous chapter to calculate the covariance matrix: catalogues generated from the MultiDark simulation with the VIPERS cone-like shape both for W1 and W4, the photometric and spectroscopic (version 4.0) angular masks applied and the sampling SPOC selection included (the SPOC effects have been corrected in the measured power spectrum for each mock). Even the Pinocchio mocks, used to measure a smooth correlation matrix given their big number, should include VIPERS properties such as the VIPERS geometry, angular masks and sampling. For the moment a set of Pinocchio catalogues updated to the current stage of observations is not available. We can deal with two sets of Pinocchio mocks: one with the spectroscopic angular mask, version 3.0, and the SPOC algorithm applied and another with the current version 4.0 angular mask but without SPOC. For the estimation of the correlation matrix, we choose the second set of mock catalogues. This choice is justified by the fact that the recovered power spectrum is mainly affected by a wrong window function (in this case due to the wrong angular mask, 3.0 instead of 4.0) with respect to the alteration due to the SPOC sampling, even further after correcting the SPOC effects with the "local" TSR approach. In figure 6.4, we will show that the absence of fully realistic Pinocchio samples does not affect much the estimation of cosmological parameters and related errors.

In statistics the likelihood is a function which tells us how likely an event is. This probability function is determined by the chi-square through equation

$$L(\mathbf{x}|\mathbf{p}) \propto \exp -\frac{1}{2}\chi^2(\mathbf{x}, \mathbf{p}) , \tag{6.2}$$

where L defined in statistics as the probability of finding some measurements $\mathbf{x}$ given some models, function of a set of parameters $\mathbf{p}$. In the bayesian statistics, constraints on parameters are given by the posterior probability distribution $\mathrm{P}(\mathbf{p}|\mathbf{x})$ through the Bayes theorem

$$\mathrm{P}(\mathbf{p}|\mathbf{x}) = \frac{L(\mathbf{x}|\mathbf{p})\mathrm{P}(\mathbf{p})}{\mathrm{P}(\mathbf{x})} , \tag{6.3}$$

where $\mathrm{P}(\mathbf{x})$ and $\mathrm{P}(\mathbf{p})$ are respectively the probability of observing the data and the prior probability distribution on parameters $\mathbf{p}$. In case of flat priors for parameters, these two last terms do not change the posterior contours making the likelihood and the posterior to contain the same information: $L(\mathbf{x}|\mathbf{p}) \propto \mathrm{P}(\mathbf{p}|\mathbf{x})$.
We sample the parameter space onto a grid over our flat prior. Finally we subtract the integral constraint and multiply the resulting model by a factor b^2.
Percival et al. (2014) proposed an unbiased estimator for the inverse covariance matrix ψ

$$\psi_{\mathrm{ij}} = \mathrm{C}^{-1}_{\mathrm{ij}} \times (1 - \mathrm{D}) \ , \tag{6.4}$$

where the term $\mathrm{D} = (\mathrm{N_b} + 1)/(\mathrm{N_s} - 1)$ dependent on the number of simulation $\mathrm{N_s}$ and the number of bins $\mathrm{N_b}$ used to estimate the covariance matrix. In case of limited number of available mock catalogues with respect to the number of bins, the error in the estimation of $\mathrm{C_{ij}}$ would propagate up to a systematic error in the parameter space (Taylor et al., 2013; Dodelson and Schneider, 2013). Following the likelihood approach to estimate the error on parameters, the variance of each parameter should be increased by a factor which depends on the number of bins $\mathrm{N_b}$, free parameters $\mathrm{N_p}$ and mocks $\mathrm{N_s}$ by a quantity

$$\mathrm{m_1} = \frac{1 + \mathrm{B}(\mathrm{N_b} - \mathrm{N_p})}{1 + \mathrm{A} + \mathrm{B}(\mathrm{N_p} + 1)} \ , \tag{6.5}$$

where A and B described in equation 8 of Percival et al. (2014). If the parameter errors are estimated as the variance from the distribution of best-fit parameter from each mock catalogue, the correcting factor is

$$\mathrm{m_2} = \mathrm{m_1}(1 - \mathrm{D}) \ . \tag{6.6}$$

Another important issue that I did not mention up to now is the choice of a fiducial cosmology that can have a strong impact on the recovered power spectrum.

6.1.1 Correcting for the fiducial cosmology

The choice of the fiducial cosmology used to convert redshifts and angles to physical coordinates could alter the shape and the amplitude of the recovered power spectrum. Since the main purpose of this paper is to fit the best power spectrum model in order to constrain cosmological parameters, the ideal case would be to recompute for each tested model, the power spectrum and the window function according to the given cosmology, but this is not computationally feasible.
Eisenstein et al. (2005) showed a practical way to incorporate these distortions in the correlation function when computed using different cosmologies by rescaling the distance scale by the factor

$$\alpha = (\alpha_{\perp}^2 \alpha_{\parallel})^{\frac{1}{3}} \ . \tag{6.7}$$

The dilation factor α is a combination of a perpendicular-plane contribution respect to the line-of-sight

$$\alpha_{\perp} = \frac{\mathrm{D_{A,model}}}{\mathrm{D_{A,fiducial}}} \ , \tag{6.8}$$

(with D_A the angular comoving distance) times a correction for the parallel direction

$$\alpha_{\parallel} = \frac{H_{\rm fiducial}(\langle z \rangle)}{H_{\rm model}(\langle z \rangle)} \ , \tag{6.9}$$

where $H(\langle z \rangle)$ the Hubble rate at the mean redshift of the sample $\langle z \rangle$. When we assume different fiducial cosmologies, comoving distances change. What we want to maintain fixed are two observables: the angle between two objects in the sky $\Delta\theta$ which leads to the rescaling factor $\alpha_{\perp}$ and their separation in redshift Δz which leads to $\alpha_{\parallel}$ (Alcock and Paczynski, 1979).

Instead of correcting for each tested cosmology the measurements of the power spectrum, the window function and the covariance matrix which depends on the assumed fiducial cosmology, we prefer to rescale the model power spectrum to the fiducial one. In case of compact survey where an isotropic treatment of the window function could be done, the rescaling factor for the pair separation r in the correlation function, once one adopts another cosmology, is $1/\alpha$, where α is the infinitesimal volume to the one third of expression 6.7. Since the power spectrum is the Fourier Transform of the correlation function, the wave-number of the model power spectrum is multiplied by a factor α and its amplitude by $1/\alpha^3$ in order to rescale it to the fiducial cosmology. Due to the anisotropy introduced by the VIPERS window function and the RSD, instead of correcting the theoretical monopole power spectrum, we apply the dilation factor both to the 3D model power spectrum before doing the convolution of equation 5.7 and to the k-space grid

$$\begin{cases} P_{\rm M,fiducial} &= \frac{P_{\rm M,model}}{\alpha_{\perp}^2 \alpha_{\parallel}} \\ k_{\rm x,fiducial} &= k_{\rm x,model} \times \alpha_{\perp} \\ k_{\rm y,fiducial} &= k_{\rm y,model} \times \alpha_{\perp} \\ k_{\rm z,fiducial} &= k_{\rm z,model} \times \alpha_{\parallel} \ , \end{cases} \tag{6.10}$$

where the z-coordinate is parallel to the line-of-sight direction. We test the modelling for the fiducial cosmology on the final MultiDark mocks where we know the input cosmology of the simulation. We measure the averaged power spectrum of MultiDark mocks assuming two completely different cosmologies: the input MultiDark one with the matter density $\Omega_M = 0.27$ and one with $\Omega_M = 0.40$. We fix again the redshift range between $0.6 < z < 0.9$ and the FFT-grid parameters to the values described in the previous section. In figure 6.1, we plot the results for the W1 field outlining, in the bottom panel, the difference in the recovered power spectrum once we assume the wrong cosmology.

This deviation in the shape and amplitude of the power spectrum could be corrected introducing in the analysis the correction factors just described. The power, measured assuming a wrong cosmology and corrected to the MultiDark one, is in fact in good agreement with the power measured assuming the right cosmology of the simulation.

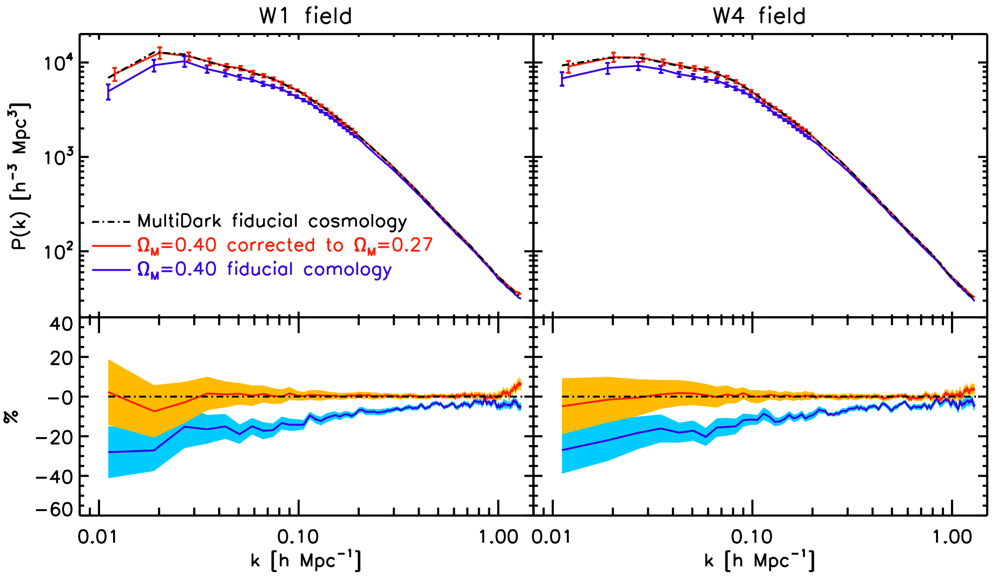

Figure 6.1: *Top left panel*: measured power spectrum from the final W1 MultiDark mocks assuming, as fiducial cosmology, the true one of the simulation (dot-dashed black line), a wrong cosmology (blue solid line) and the wrong cosmology corrected to the true one (red solid line) though 6.10. *Bottom left panel*: ratio between the measured power spectrum with (red line) and without (blue line) the wrong cosmology correction respect to the power measured with the true fiducial cosmology. *Right panel*: same as the left panel but for the W4 field.

6.2 Study of systematics

6.2.1 Modelling Redshift Distribution

One of the possible cause of systematics, in the estimation of the galaxy power spectrum, is the measure of the mean redshift distribution which should account for the selection function of the survey. The main problem is due to the fact that this mean density of galaxies is calculated directly from the data dividing, in our case, the whole redshift range of the sample between $z = 0.4$ and $z = 1.2$ in a number N of bins. Due to the noise of the galaxy redshift distribution in VIPERS, we smooth the histogram with a Gaussian filtering.

Other methods can be used to reconstruct the selection function of the sample such as the fit of the observed distribution with a smooth template. In general most of the methods, as the Gaussian filtering or the analytic fitting formula, are parametric and have to be calibrated. An alternative non-parametric method is the $V_{\max}$ approach. This method consists in randomly sampling the maximum volumes $V_{\max}$ probed by each galaxy in the survey and it is based on the galaxy luminosity function (Cole, 2011). de la Torre et al. (2013) applied the three described methods to recover the RD in the PDR-1 catalogue analysing their impact on the measurement of the projected correlation function. The difference between the three different approaches is minimal $< 5\%$. Due to this, we choose to model the RD of our sample with the simplest Gaussian filtering approach testing the impact of the kernel in the measure of the power spectrum of mock catalogues. The choice of the Gaussian kernel is critical: a big value for σ means that we mix together too many scales losing information about the mean density at a certain redshift. If σ is too small, the RD follows the clustering of galaxies which in principle we want to avoid because the goal is to extrapolate the mean density of the sample. The modelling of the RD enters, in the estimation of the power spectrum, in the radial distribution of random points used to mimic the mean density of the sample and in the normalisation of the power.

We consider a very simple set of 26 W1 MultiDark mock samples with only the spectroscopic "Nagoya" version 3.0 mask applied, no photometric mask and without any sampling selection applied, with galaxies distributed in real space in a redshift range $0.6 < \mathrm{z} < 0.9$. We also did some tests considering the distribution of galaxies in redshift space both for W1 and W4 obtaining similar results. The "true" RD has been calculated putting together all the RDs of the W1 mock catalogues and smoothing with a Gaussian filtering of $\sigma = 50\,\mathrm{h^{-1}Mpc}$. The averaged power spectrum over 26 mock samples measured assuming the "true" RD is taken as reference. To test the impact of the RD modelling, we consider three Gaussian kernels of $\sigma = 100 - 150 - 200\,\mathrm{h^{-1}Mpc}$. Figure 6.2 shows the behaviour of the RD modelled using the three different kernels compared to the measured distribution of galaxies in radial bins. For each mock, we create the random catalogue and normalise the recovered power using the proper RD measured from that particular mock, instead of using the "true" averaged RD. Then we average all the 26 power spectra calculated smoothing the RD with a particular kernel and plot the results in figure 6.3. We choose the best value for the Gaussian kernel comparing the recovered power spectra with the power obtained fixing the same "true" RD for all the mocks. We notice that the power spectrum depends on the particular RD modelling only at large scales. As pointed out before, the smallest value considered for the kernel $\sigma = 100\,\mathrm{h^{-1}Mpc}$ leads to a suppression of the power because the estimated mean density is following the clustering of galaxies. On the other hand, $\sigma = 200\,\mathrm{h^{-1}Mpc}$ is too high because mixes too many redshift bins: the mean density is less sensitive to the selection function of the sample. We conclude that the Gaussian filtering with $\sigma = 150\,\mathrm{h^{-1}Mpc}$ is a good model to the RD of our survey leading to an unbiased estimate of the power. The choice of the kernel should not significantly effect the estimation of cosmological parameters because modes more affected by the RD modelling are also the ones with the bigger statistical error. To imagine the uncertainties due to cosmic variance, errors plotted in figure 6.3 should be multiplied by a factor of $\sqrt{26} \sim 5$ leading to statistical fluctuations in the power much bigger than the systematics introduced by RD modelling.

6.2.2 2 parameters: $\Omega_\mathrm{M} - \mathrm{b}$

To analyse systematics in the estimation of cosmological parameters, we constrain the matter density parameter Ω_M from the mock catalogues through a likelihood analysis and compare it with the known MultiDark value of $\Omega_\mathrm{M} = 0.27$. As specified before, we

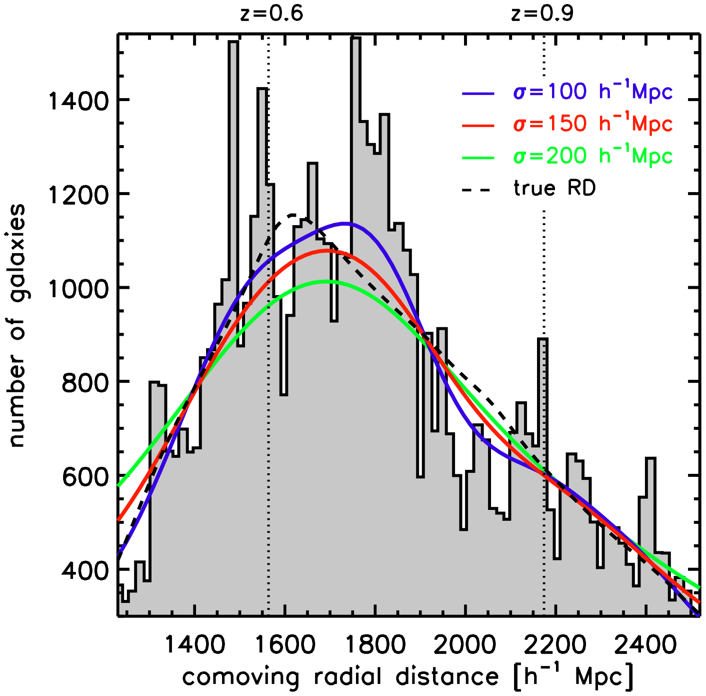

Figure 6.2: radial distribution of galaxies for one single mock measured counting the number of particles in each bin (grey histogram). The solid lines represent the Gaussian filtering with kernel $\sigma = 100\,\mathrm{h^{-1}Mpc}$ (blue line), $\sigma = 150\,\mathrm{h^{-1}Mpc}$ (red line) and $\sigma = 200\,\mathrm{h^{-1}Mpc}$ (green line) while the black dashed line shows the "true" RD averaged over 26 mock catalogues.

measure the averaged power spectrum over the 26 final MultiDark samples for W1 and 31 for W4 in redshift space between $0.6 < \mathrm{z} < 0.9$ assuming the MultiDark cosmology as the fiducial one. We fix all the cosmological parameters to the MultiDark cosmology except for the matter density Ω_M. RSD are included in our modelling before computing the convolution with the VIPERS window function. The growth rate f in the RSD Kaiser factor is fixed by the tested value of Ω_M at a certain epoch and assuming General Relativity ($\mathrm{f} = \Omega_\mathrm{M}^{0.55}(\mathrm{a})$). For the moment, the pairwise velocity parameter σ_v is fixed to the best fit of figure 5.11: $\sigma_\mathrm{v} = 530\,\mathrm{km\,s^{-1}}$. The model is then convolved with the VIPERS window function, rescaled to the MultiDark fiducial cosmology and normalised by a factor b^2 where the bias factor b is let free. We compare the measurements with theory using a chi-square technique. We generate a two dimensional grid-space parameter with flat priors

$$\begin{gathered} 0.20 < \Omega_\mathrm{M} < 0.40 \\ 1 < \mathrm{b} < 2\,, \end{gathered} \tag{6.11}$$

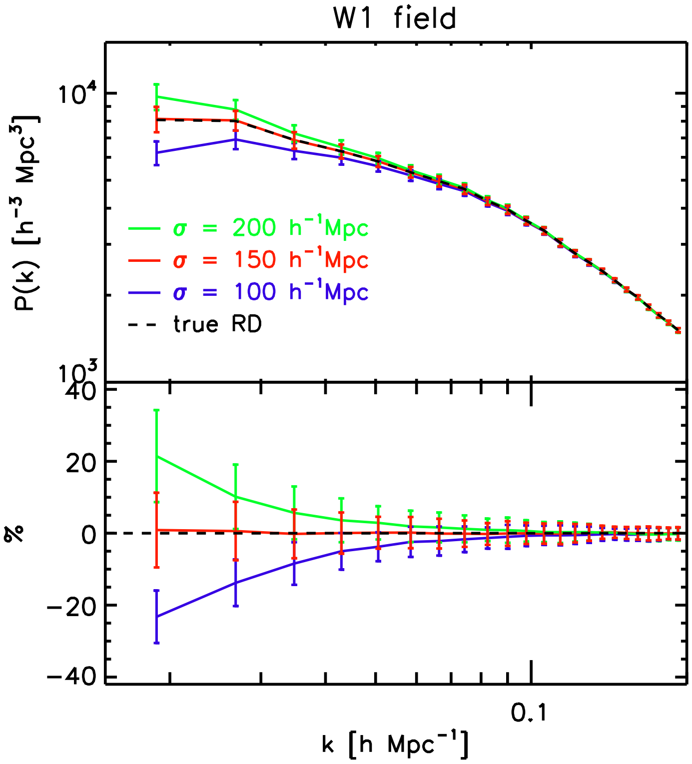

Figure 6.3: *Top panel*: impact of RD modelling on the recovered power spectrum. The plotted power spectra are related to different kernels in the Gaussian smoothing: $\sigma = 200\,\mathrm{h^{-1}Mpc}$ green line, $\sigma = 150\,\mathrm{h^{-1}Mpc}$ red line and $\sigma = 100\,\mathrm{h^{-1}Mpc}$ blue line while the dashed black line represents the power measured with the 'true" RD. *Bottom panel*: ratio between the recovered power spectrum with the RD calculated directly from the single mock and the reference power with the RD calculated using an averaged over all the mock catalogues.

where each point of the grid corresponds to a set of parameters $\Omega_{\mathrm{M}} - \mathrm{b}$ with steps $\Delta\mathrm{b} = 0.05$ and $\Delta\Omega_{\mathrm{M}} = 0.01$. The model power spectrum is constructed for each point of the 20×20 parameter-space grid. Since the step in the grid is quite large (to reduce the computational time), each model, after the inclusion of RSD, fiducial cosmology, bias factor and window function, is interpolated to a finer 1000×1000 grid reducing the step to $\Delta\mathrm{b} = 0.001$ and $\Delta\Omega_{\mathrm{M}} = 0.0002$. The chi-square χ^2 as a function of the two parameters $\Omega_{\mathrm{M}} - \mathrm{b}$ is evaluated at each point of the new grid. Since we want to study systematics, the variance computed among the mocks is rescaled to obtain the error on their averaged. As a result, before applying the correction of expression 6.4 to the inverse of the covariance matrix, we divide each element $\mathrm{C_{ij}}$ by the number of MultiDark mocks.

The fit is performed in a k-range between $0.01 < \mathrm{k} < 0.3\,[\mathrm{h\,Mpc^{-1}}]$ with a linear binning step of $\Delta\mathrm{k} = 0.01\,\mathrm{h\,Mpc^{-1}}$.

The minimum frequency of $\mathrm{k_{min}} = 0.01\,\mathrm{h\,Mpc^{-1}}$ is the mode related to the maximum wavelength sampled in the redshift direction even if its contribution to the estimate of the chi-square is very small due to the large statistical errors at those scales.
The maximum frequency instead is quite important for the best-fit parameter estimation because errors are very small at higher modes. We choose a maximum frequency of $\mathrm{k_{max}} = 0.3\,\mathrm{h\,Mpc^{-1}}$, large enough to have a sufficient number of bins $\mathrm{N_b} = 29$ and small enough to avoid scales strongly affected by non-linearities in the power, non-linear bias and SPOC sampling. Given the set of Pinocchio mocks $\mathrm{N_s} = 200$ used to estimate the correlation matrix and the number of bins $\mathrm{N_b} = 29$, the correcting value for the inverse covariance matrix is of the order of $1 - \mathrm{D} \sim 85\%$. In figure 6.4, we compare the chi-square contours in the $\Omega_\mathrm{M} - \mathrm{b}$ plane for the W1 and W4 field: in one case we measure the monopole power spectrum from mock catalogues with only the angular mask applied, while the other two contours are calculated from mocks with also the SPOC sampling applied to our samples. The difference between the last two is that, in one case we measure the chi-square from the raw power spectrum while in the other we correct the power for the SPOC effect using the rectangular "local" TSR.

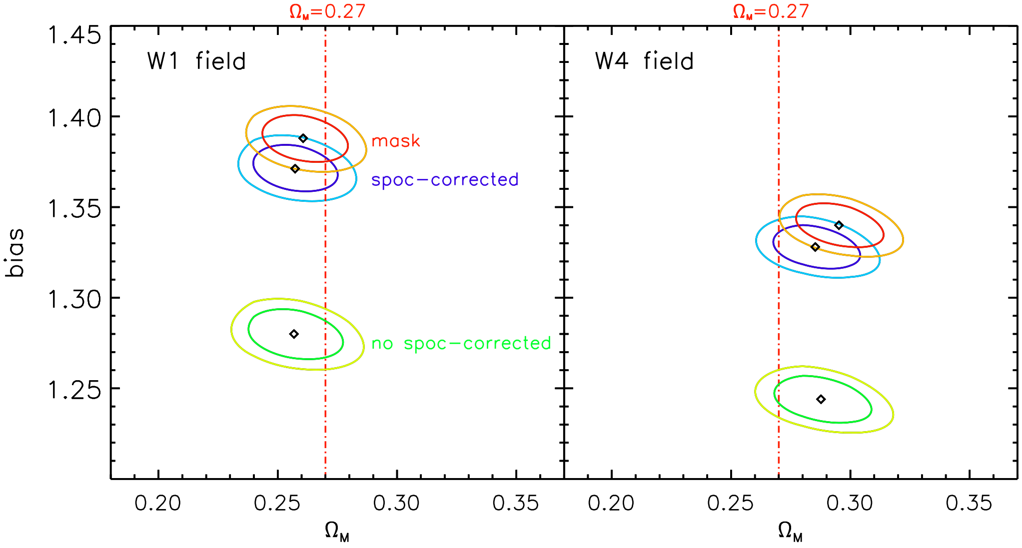

Figure 6.4: *Left panel*: contours $\chi^2 = \chi^2_{\mathrm{min}} + 2.3$ and $\chi^2 = \chi^2_{\mathrm{min}} + 4.61$ for the chi-square, function of two parameters Ω_M, b. The red line corresponds to the chi-square evaluated from W1 MultiDark mocks with the cone-like geometry and the angular mask applied. The blue and green lines show the case for samples with also the SPOC algorithm applied, respectively with and without the correction of the power for the "local" TSR (using the rectangular shape for the estimation of the weight). *Right panel*: same as the left panel but for the W4 field.

The chi-squares evaluated for the "mask" catalogues and the SPOC samples with the "local" TSR correction included give similar results for both Ω_M and b, meaning that the SPOC correction of section 5.5.2 is working quite well. The small difference between the chi-square contours, in this case, could be related to the "local" TSR correction applied to the measured power spectrum in SPOC samples, which does not perfectly correct the power at the highest modes fitted $k \sim 0.3\,h\,Mpc^{-1}$ (figure 5.15). Another source of systematic, that could explain this small discrepancy, is the incorrect correlation matrix estimated using a set of Pinocchio mocks without the SPOC sampling. However, since we divide the covariance matrix by the number of MultiDark mock samples in order to study systematics, these small fluctuations of the chi-square in the parameters space will not have a significant impact in the estimation of cosmological quantities in VIPERS data.

The difference between the chi-square contours for the SPOC corrected power spectrum and the one without correction is in both cases negligible for Ω_M and more evident for the bias. This is easily understood looking at figure 5.12 where the difference between the two power spectra in the k-range $0.01 < k < 0.3\,[h\,Mpc^{-1}]$ is just a normalisation factor which is included in our modelling through the bias term. The parameter Ω_M is well recovered within statistical uncertainties in both W1 and W4 field. In figure 6.5, we focus on the chi-square calculated with the final MultiDark mocks with the selecting sampling included and with the power corrected for the SPOC effect. In the bottom

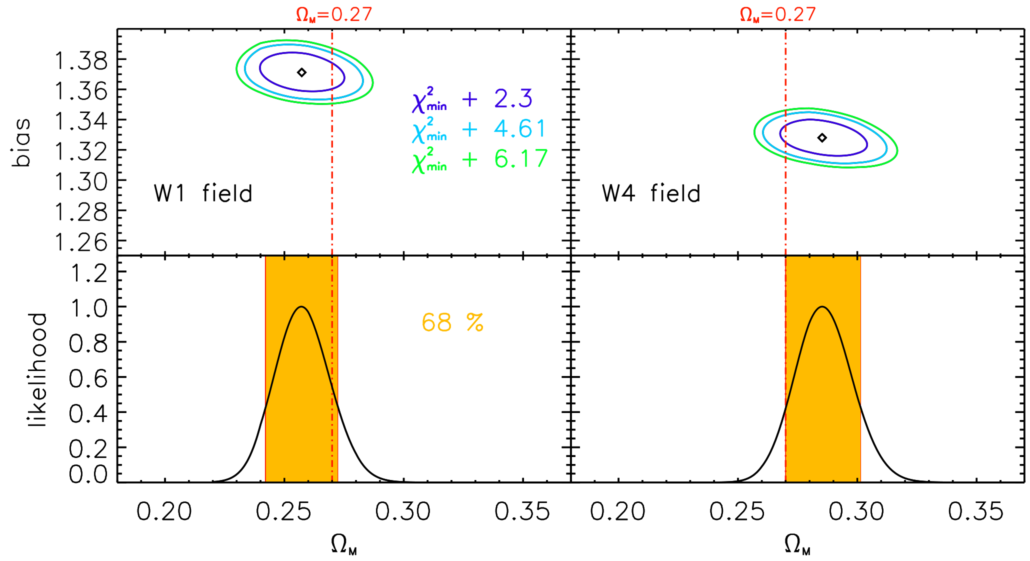

Figure 6.5: *Top left panel*: chi-square contours in the Ω_M, b parameter-plane for the power measured from the final MultiDark W1 mocks corrected for the SPOC effect. *Bottom left panel*: marginalised likelihood over the bias with the coloured region representing the 68 % of probability. The best fit is consistent with the value of the MultiDark matter density shown by the $\Omega_M = 0.27$ red-dashed line. *Right panel*: same as the left panel but for the W4 field. Even for W4, the matter density is well constrained to the expected value within statistical uncertainties.

panel, the marginalised likelihood over the bias, calculated from equation 6.12, is plotted to analyse systematics in the constraint of Ω_{M},

$$\mathrm{L}(\Omega_{\mathrm{M}}) = \frac{\int \mathrm{L}(\Omega_{\mathrm{M}}, \mathrm{b})\,\mathrm{db}}{\int\int \mathrm{L}(\Omega_{\mathrm{M}}, \mathrm{b})\,\mathrm{db}\,\mathrm{d}\Omega_{\mathrm{M}}}\,. \tag{6.12}$$

The best value for the matter density is consistent with the value of the MultiDark simulation within the statistical errors. The error on the Ω_{M} parameter, plotted in the bottom figure 6.5, has been corrected by a factor of $\sqrt{\mathrm{m}_1} = 1.07$ accounting for the propagation of the error in the estimation of the covariance matrix.
In the following, errors of all estimated parameters will be corrected by the m_1 factor which, in all cases, is very small due to the large number of Pinocchio mock catalogues used to estimate the covariance matrix with respect to the number of bins and parameters constrained.

Monte Carlo Markov Chain

The estimation of the chi-square on a N-dimension parameter grid, when the number of parameters to be constrained is high, is very time-consuming. In this cases, it is useful to sample the N-dimensional likelihood surface through a Monte-Carlo technique. The Monte Carlo Markov Chain (MCMC) is a fast algorithm, often used in statistics, to collect samples of a target distribution. In our analysis, we use the free software package CMBeasy MCMC code designed by Doran and Müller (2004).
A simple description of the main algorithm (Metropolis-Hastings algorithm, Metropolis et al., 1953; HASTINGS, 1970) used to run an MCMC is given here: it works starting from the measurement of the likelihood L_1 of a random point $\mathbf{p}_1$ in the parameter space and then it jumps, with a given step, to another one $\mathbf{p}_2$ measuring again the likelihood L_2: if $\mathrm{L}_2/\mathrm{L}_1 > 1$ the new starting point of the chain will be $\mathbf{p}_2$ because the set of parameters in that point are more likely than the ones in $\mathbf{p}_1$, while if $\mathrm{L}_2/\mathrm{L}_1 < 1$ there is a probability proportional to the ratio of the two likelihoods to move to the new point $\mathbf{p}_2$ or maintain the $\mathbf{p}_1$ position. With this simple idea the chain tends to sample more the peak of the likelihood distribution with the quality of the sample improving as a function of the number of steps. The important part is that, even in the case $\mathrm{L}_2/\mathrm{L}_1 < 1$, there is a probability different from 0 to move to the new unlikely point giving a true sampling of the underlying likelihood distribution. Without that condition, the chain would reach quickly the parameters with the maximum likelihood giving only the best fit without any errors. One advantage of the MCMC is that the marginalised likelihood for each parameter is proportional to the number density of points sampled: so we do not have to perform equation 6.12 for each parameter. For the rest of the work we are going to use the MCMC algorithm to estimate the likelihood for more than two parameters. We test the accuracy of the MCMC method constraining the parameters Ω_{M} and b from the final set of mock catalogues including the SPOC correction in the estimation of the power spectrum. In figure 6.6 we over-plot to the contours of the chi-square calculated in the previous section the MCMC points with $\chi^2 < \chi^2_{\mathrm{MIN}} + 2.3$ obtaining a good match between the two. The bottom panel shows the comparison between the marginalised

likelihood calculated with equation 6.12 and the number density of points given by the MCMC in each $\Omega_{\rm M}$-bin.

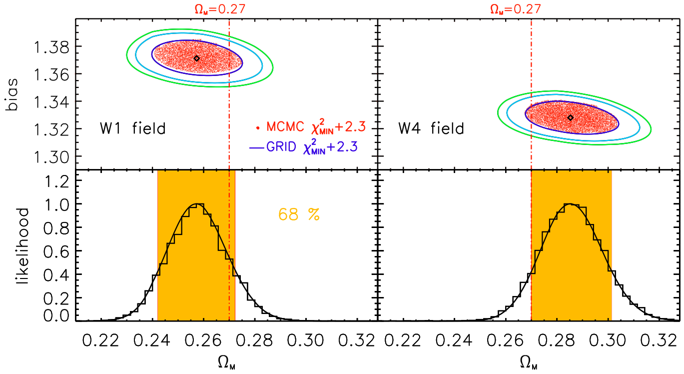

Figure 6.6: *Top panels*: same as figure 6.5 but with the chi-square points calculated with the MCMC technique over-plotted (red points). *Bottom panels*: same as figure 6.5 with the histogram showing the number density of points in each $\Omega_{\rm M}$-bin evaluated from the MCMC steps.

6.2.3 3 parameters: $\Omega_{\rm M} - {\rm b} - \sigma_{\rm v}$

Since for data we will use a Gaussian prior for the pairwise velocity dispersion of galaxies, we recalculate the chi-square with the MCMC approach for a set of three free parameters: $\Omega_{\rm M}$, b and $\sigma_{\rm v}$. We assign to the matter density and the bias the same flat priors of expression 6.11 while we add, for the $\sigma_{\rm v}$ parameter, a Gaussian prior of the same order to the one for data. To add the Gaussian prior information on the $\sigma_{\rm v}$ parameter, we recall the Bayes theorem of equation 6.3. The posterior distribution is no longer proportional to the likelihood and becomes

$$L(\mathbf{p}|\mathbf{x}) \propto L(\mathbf{x}|\mathbf{p}) \times \exp^{-\frac{(\sigma_{\rm v}-\bar{\sigma}_{\rm v})^2}{2\sigma^2}} , \tag{6.13}$$

where $\bar{\sigma}_{\rm v}$ and σ^2 are respectively the mean and the variance associated to the Gaussian prior. For the Gaussian prior mean, we choose the best fit for the pairwise dispersion velocity obtained in figure 5.11, $\bar{\sigma}_{\rm v} = 530\,{\rm km\,s^{-1}}$ with deviation $\sigma = 30\,{\rm km\,s^{-1}}$ to be consistent with the value found in the work by de la Torre et al. (2013) for the VIPERS PDR-1 catalogue and mentioned in Bel et al. (2014).

In figure 6.7 for the W1 field and 6.8 for the W4 field, we plot a combination between the three free parameters in our power spectrum modelling. The matter density is again well constrained to the MultiDark value within statistical uncertainties in both fields. However, the pairwise dispersion velocity seems to be higher in the case of W1 reaching a best fit of $\sigma_{\rm v} \sim 570\,{\rm km\,s^{-1}}$, while in the case of W4 the best fit is again of the order of $\sigma_{\rm v} \sim 530\,{\rm km\,s^{-1}}$. Figure 6.9 shows a comparison for the chi-square contours in the $\Omega_{\rm M} - {\rm b}$ plane between the case in which the pairwise velocity is fixed and the one with Gaussian prior on $\sigma_{\rm v}$.

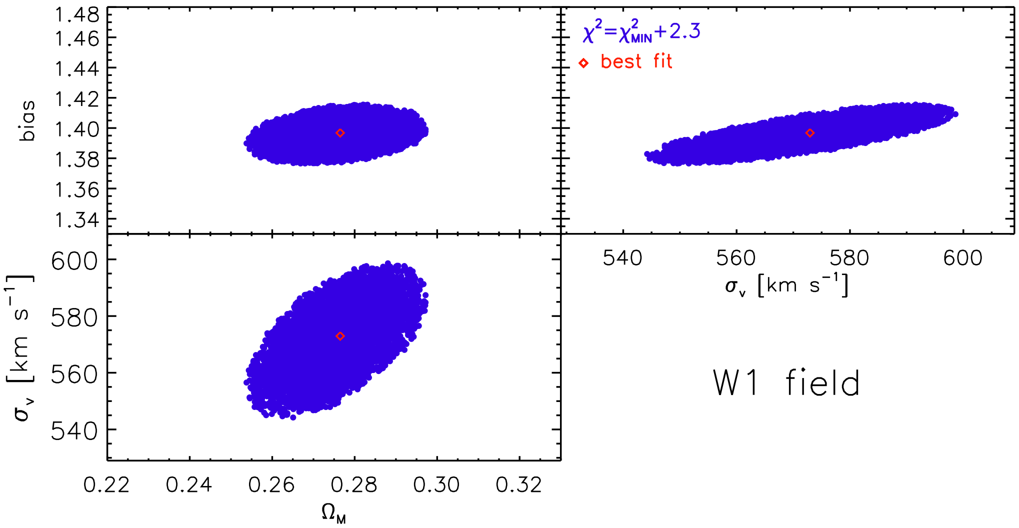

Figure 6.7: best fit for the cosmological parameters $\Omega_M - b - \sigma_v$ showing the 68 per cent confidence levels calculated using the averaged power spectrum from the 26 final MultiDark mocks for the field W1 in a redshift bin $0.6 < z < 0.9$ and k-range $0.01 < k < 0.3$ [h Mpc^{-1}]. The red diamond points show the best fit for the parameters.

As expected the red contours, associated with the three free parameter model, are larger than the two parameter chi-square due to the introduction of a new degree of freedom. In the W4 field the introduction of a Gaussian prior to σ_v does not change too much in the chi-square as function of Ω_M and b. In the W1 field, the best fit chi-square contours slightly move to higher matter density values in order to compensate the enhancement of the pairwise velocity. The higher value of σ_v, in fact, increase the damping of the model power spectrum which could be compensated in two ways: shifting the power toward higher k-modes or reducing the damping of baryon. The increase of the matter density has the two effects on the power: shift the k_{eq} to higher modes, since the equivalence between radiation and matter occur before in the evolution of the Universe at smaller scale for the horizon, and reduce the baryonic fraction ($f_B = \Omega_B/\Omega_M$) because, during all the process, the baryon density is maintained constant to the MultiDark value.

6.2.4 4 parameters: $\Omega_M - f_B - b - \sigma_v$

To anticipate what the power spectrum analysis could constrain from real VIPERS data, we extract cosmological information from VIPERS simulation on the two parameters that mostly affect the power spectrum shape: the baryonic fraction $f_B = \Omega_B/\Omega_M$ and the matter density Ω_M. Since we cannot break the degeneracy between the bias and σ_8 in real measurements, we use the overall shape information to constraint cosmology marginalising over the amplitude. We fix σ_8 to the MultiDark value of $\sigma_8 = 0.82$ nor-

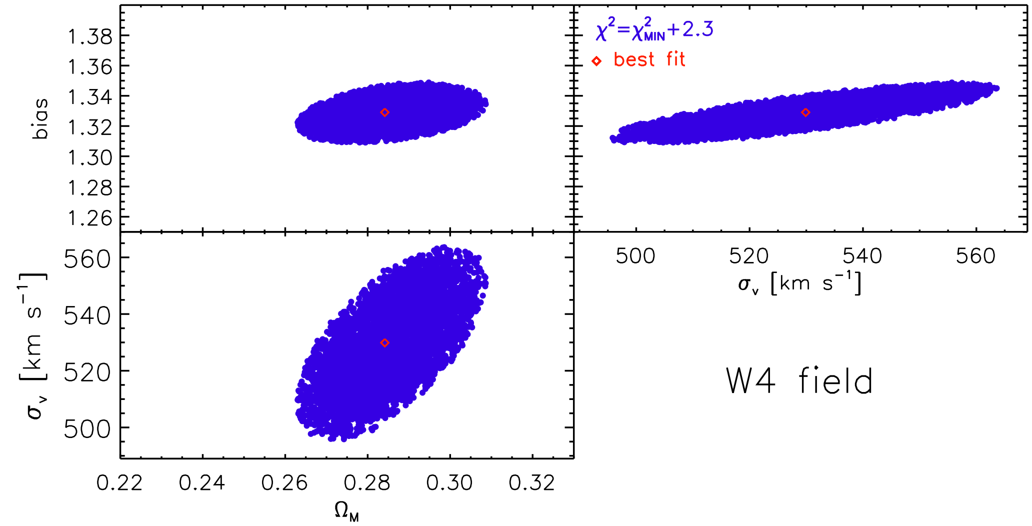

Figure 6.8: best fit for the cosmological parameters $\Omega_M - b - \sigma_v$ showing the 68 per cent confidence levels calculated using the averaged power spectrum from the 31 final MultiDark mocks for the field W4 in a redshift bin $0.6 < z < 0.9$ and k-range $0.01 < k < 0.3\ [h\,\mathrm{Mpc}^{-1}]$. The red diamond points show the best fit for the parameters.

malising the amplitude of each model to that value rescaled to the mean redshift of the sample. We impose flat priors on the matter density, baryonic fraction and bias and a Gaussian prior on the pairwise velocity dispersion

$$\begin{aligned} 0.01 < \Omega_M &< 0.90 \\ 0.01 < f_B &< 0.50 \\ 1 < \ b &< 2 \\ \sigma_v &= 530 \pm 30\,\mathrm{km\,s^{-1}}\,. \end{aligned} \tag{6.14}$$

Again we measure the averaged power spectrum from the final set of mocks both for W1 and W4 in a redshift range $0.6 < z < 0.9$ and, after correcting for SPOC, compare it with theoretical models. To perform the comparison between data and theory, we use an MCMC to sample the likelihood surface, which is a really faster approach than generate a 4-dimensional parameter space grid. We show for the field W1 and W4, respectively in figure 6.10 and 6.11, a combination between the chi-square contours on the $\Omega_M - f_B - b$ parameters. Again errors in the $\Omega_M - b$ plane are bigger respect to the two (and three) free parameter case due to the introduction of another degree of freedom given by f_B. The main results plotted in figure 6.12 are the chi-square contours in the $\Omega_M - f_B$ plane. The results for both fields are consistent between each other with the same degeneracy. The confidence levels of $68 - 90 - 98$ per cent are larger in the W4 case due, in part, to the limited angular aperture which translates into a smaller

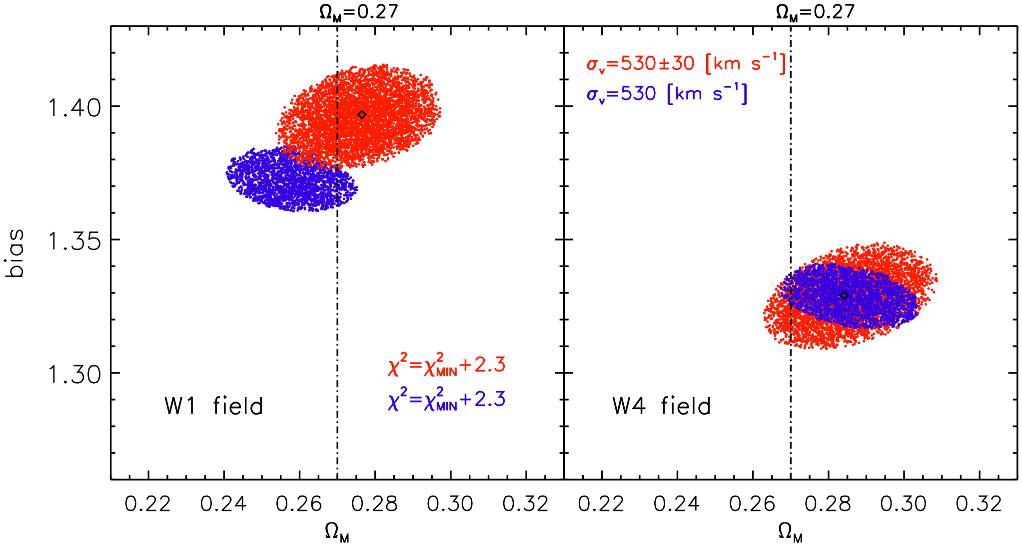

Figure 6.9: *Left panel*: comparison for the field W1 between the 68 per cent confidence level chi-square in the $\Omega_{\rm M} - {\rm b}$ plane for a modelling of the theoretical power spectrum with two free parameters ($\Omega_{\rm M} - {\rm b}$, blue contour) and a model with three free parameters ($\Omega_{\rm M} - {\rm b} - \sigma_{\rm v}$, red contour) with the best fit represented by a black diamond point. *Right panel*: same as the left panel but for the W4 field.

volume sampled. Further in both fields the expected value of the MultiDark cosmology $\Omega_{\rm M} = 0.27$ and $f_{\rm B} = 0.0469/0.27$ is within the 68 per cent confidence level meaning that we are able to constraint cosmological information from the VIPERS catalogue with systematic errors lower than statistical ones.

Note that errors plotted in these tests with mocks are a factor of $\sim 5 - 6$ smaller that the ones expected for data because, in order to study systematics, the covariance matrix has been divided by the number of MultiDark mocks.

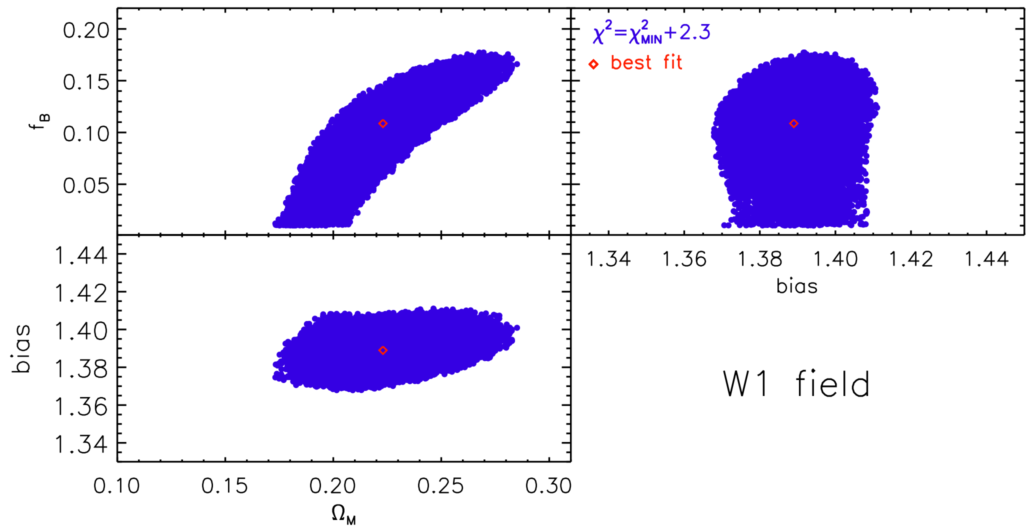

Figure 6.10: best fit for the cosmological parameters $\Omega_{\rm M} - {\rm f_B} - {\rm b}$ showing the 68 per cent confidence levels calculated using the averaged power spectrum from the 26 final MultiDark mocks for the field W1 in a redshift bin $0.6 < z < 0.9$ and k-range $0.01 < k < 0.3$ $[{\rm h\,Mpc^{-1}}]$. The red diamond points show the best fit for the parameters.

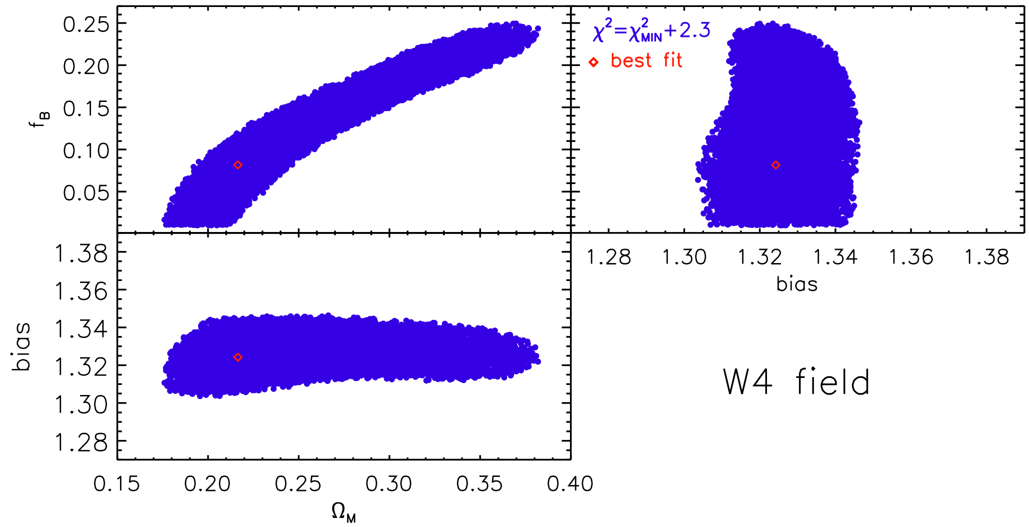

Figure 6.11: best fit for the cosmological parameters $\Omega_{\rm M} - {\rm f_B} - {\rm b}$ showing the 68 per cent confidence levels calculated using the averaged power spectrum from the 31 final MultiDark mocks for the field W4 in a redshift bin $0.6 < z < 0.9$ and k-range $0.01 < k < 0.3$ $[{\rm h\,Mpc^{-1}}]$. The red diamond points show the best fit for the parameters.

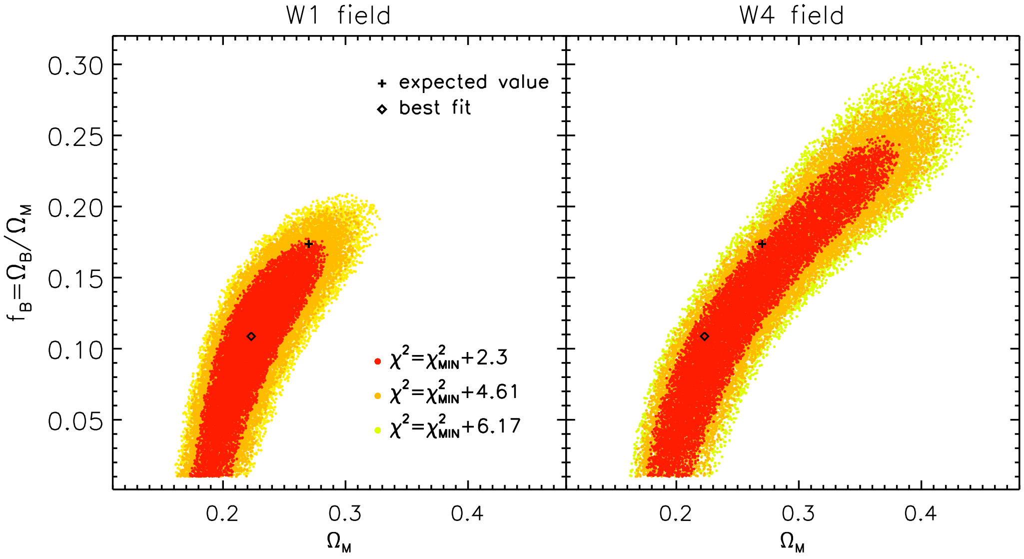

Figure 6.12: *Left panel*: 68 − 90 − 95 per cent confidence levels for the chi-square in the $\Omega_M - f_B$ plane calculated using the power spectrum recovered from the 26 final MultiDark mocks for the field W1. The best fit is represented by a black diamond point while the black cross shows the expected value for the MultiDark cosmology. *Right panel*: same as the left panel but for the W4 field.

Chapter 7

VIPERS power spectrum

The two independent fields and the redshift depth of VIPERS allow us to split the current VIPERS data set into four independent volumes, i.e. two redshift bins $0.6 < \mathrm{z}_1 < 0.9$ and $0.9 < \mathrm{z}_2 < 1.1$ with more or less the same volume, in both W1 and W4. These four data sets will be characterised by two slightly different bias values and growth factors (at different redshifts) and two slightly different window functions (W1 and W4). We can handle these differences at best by treating each of them separately. The division has been made in order to avoid the comparison between the clustering of galaxies characterised by a strongly different bias since galaxies at smaller redshift have smaller clustering properties with respect to those more luminous at higher redshifts. We do not consider data at very low redshifts $0.4 < \mathrm{z} < 0.6$ due to uncertainties in the colour-cut galaxy selection. We also avoid the most distant regions of the survey because they are shot-noise dominated. The mean redshift for the two redshift bins is $\langle \mathrm{z}_1 \rangle \sim 0.73$ and $\langle \mathrm{z}_2 \rangle \sim 0.98$.

The number of galaxies used for the estimation of the power spectrum in each subsamples and the effective volume is shown in table 7. The redshifts and angles of galaxies are converted into physical coordinates with a MultiDark fiducial cosmology with $\Omega_{\mathrm{M}} = 0.27$, $\Omega_\Lambda = 0.73$ and $\mathrm{h} = 1$ since the covariance matrix is estimated using MultiDark mock samples. The redshift distribution has been calculated counting the number of galaxies in $\mathrm{N} = 100$ radial bins which divide the whole redshift range $0.4 < \mathrm{z} < 1.2$ in slices of size $\Delta \mathrm{r} \sim 16\,\mathrm{h}^{-1}\mathrm{Mpc}$ both for W1 and W4. In order to minimise the noise of the RD and since the two VIPERS fields select galaxies in the same way, we sum in each radial bin the number of galaxies observed in W1 with those in W4. We then divide $\mathrm{N}_{\mathrm{W1}} + \mathrm{N}_{\mathrm{W4}}$ by the sum of the volume of the bin in W1 with the one in W4 in order to obtain the mean density of the sample

$$\bar{\mathrm{n}}(\mathrm{r}) = \frac{\mathrm{N}_{\mathrm{W1}} + \mathrm{N}_{\mathrm{W4}}}{\mathrm{V}_{\mathrm{W1}} + \mathrm{V}_{\mathrm{W4}}}(\mathrm{r}) \,. \tag{7.1}$$

From tests on mocks, we established that a Gaussian filtering kernel with characteristic size $\sigma \sim 150\,\mathrm{h}^{-1}\mathrm{Mpc}$ allows us to recover an unbiased power spectrum with

	W1	W4
0.4-1.2	30705 / 1.3 $\times 10^7$	22622 / 9.5$\times 10^6$
0.6-0.9	18077 / 4.6 $\times 10^6$	13140 / 3.4 $\times 10^6$
0.9-1.1	4179 / 4.2 $\times 10^6$	2637 / 3.1 $\times 10^6$

Table 7.1: Number of high-quality spectroscopic redshifts and volume in $\mathrm{h^{-3}Mpc^3}$ for each VIPERS subsample in W1 and W4 generated dividing the VIPERS sample in three redshift bins: the whole redshift range of VIPERS ($0.4 < z < 1.2$), and two smaller redshift bins ($0.6 < z_1 < 0.9$ and $0.9 < z_2 < 1.1$) used to measure the galaxy power spectrum.

$\sigma \sim 150\,\mathrm{h^{-1}Mpc}$ (section 6.2.1). The situation for data is slightly different with respect to the one tested with mocks. In the course of this work, in fact, the VIPERS sample improved, passing from the version 3.0 angular mask to the version 4.0, completing more pointings on the sky with more galaxies selected: more galaxies means less noise in the radial redshift distribution. Moreover for data, we compute the averaged RD between the two fields, W1 and W4, as in equation 7.1 reducing the noise by a factor of $\sim \sqrt{2}$ with respect to the test with mocks. Due to these reasons and in order to correctly account for the VIPERS selection function without mixing too much information from very far radial bins, we decide to apply a Gaussian filter to smooth the radial distribution of galaxies with a lower value for the kernel $\sigma = 100\,\mathrm{h^{-1}Mpc}$. As shown in figure 7.2, the smoothed mean density is not recovering the radial distribution of galaxies at very low redshift $z \sim 0.4$ but follows well the behaviour of the histogram at the redshift bins of the two analysed subsamples without being affected too much by the underlying clustering.

We generate two random catalogues with W1 and W4 peculiar geometries including the spectroscopic “Nagoya” 4.0 and the photometric mask and with the same redshift distribution modelled with the Gaussian filtering. The number of unclustered points is ~ 200 times the number of galaxies in all four regions in order to reduce the random shot-noise contribution and cover the VIPERS volume with high resolution. We choose the same FFT grid for each subsample with size $\mathrm{L} = 800\,\mathrm{h^{-1}Mpc}$, big enough to inclose all the galaxies, and cell size $\mathrm{H} = 2\,\mathrm{h^{-1}Mpc}$. The minimum frequency of the k-space grid in which we sample the galaxy power spectrum is $\mathrm{k_{min}} = \frac{2\pi}{\mathrm{L}} \sim 0.008\,\mathrm{h\,Mpc^{-1}}$, the maximum $\mathrm{k_{max}} = \frac{\pi}{\mathrm{H}} \sim 1.57\,\mathrm{h\,Mpc^{-1}}$ with step $\Delta\mathrm{k} = \mathrm{k_{min}}$. The density field in each grid point is calculated with the CIC assignment and Fourier Transformed with the FFT al-

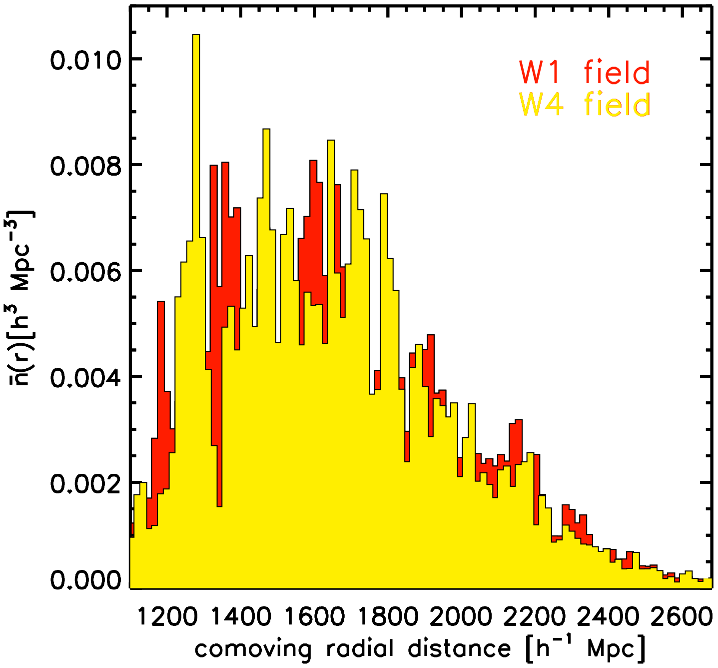

Figure 7.1: redshift distribution of galaxies in the VIPERS catalogue as a function of radial distance for both field, W1 (red histogram) and W4 (yellow histogram), measured dividing the whole redshift range into N = 100 slices, counting the number of galaxies in each bin and dividing by the volume of the bin to obtain a density.

gorithm. Smoothing the density field with the CIC scheme onto a grid biases the power spectrum estimator at smaller scales. We adopt equation 3.22 to correct for this particular choice and then subtract for the shot-noise contribution. The shot-noise term is of the order of $\mathrm{P_{SN}} \sim 250\ \mathrm{h^{-3}Mpc^3}$ for the low redshift bin in both fields. Due to the low number of galaxies at high redshift, typical of magnitude-limited survey as VIPERS, the shot-noise term ($\mathrm{P_{SN}} \sim 800\ \mathrm{h^{-3}Mpc^3}$) becomes dominant for the $\mathrm{z_2}$ power spectrum at scales of the order of $\mathrm{k} \sim 0.3\,\mathrm{h\,Mpc^{-1}}$. Finally we apply a spherical average to obtain the monopole power spectrum in linear bins with size $\Delta \mathrm{k} = \frac{2\pi}{\mathrm{L}}$.
The errors in the estimation of the VIPERS power spectrum are calculated as in section 5.6. We estimate the correlation matrix using the final 200 Pinocchio mock catalogues for each subsample (figure 7.3) and multiplying each element of the matrix, as in equation 5.11, by the variance evaluated through the final MultiDark mock samples. Again these mock samples should include all the selection functions of VIPERS data. The set of Pinocchio mocks available and used to estimate $\mathrm{r_{ij}}$ does not include yet the SPOC sampling. We verified that this does not affect much the estimation of cosmological parameters.

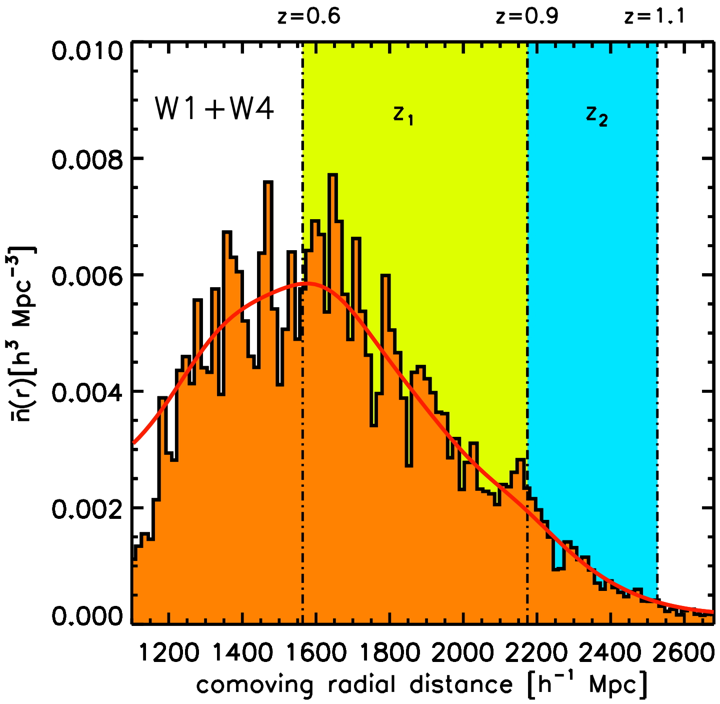

Figure 7.2: redshift distribution of galaxies in the VIPERS sample averaged as in equation 7.1 combining the two fields (orange histogram). The red line shows our best modelling of the mean density distribution obtained Gaussian filtering the radial distribution of galaxies with a kernel of $\sigma = 100\,\mathrm{h^{-1}Mpc}$. The coloured regions correspond to the two redshift bins ($0.6 < z_1 < 0.9$ yellow region, $0.9 < z_2 < 1.1$ light blue region) used in the analysis to recover the power spectrum.

The VIPERS galaxy power spectrum is finally shown in figure 7.4 with the diagonal elements of the covariance matrix $\sqrt{C_{ii}}$ plotted as statistical errors. The minimum frequency plotted $k_{min} \sim 0.01\,\mathrm{h\,Mpc^{-1}}$ is related to the largest scale we are able to sample in the redshift direction and in each sub-volume corresponding to $\lambda_{max} \sim \Delta z \sim 600\,\mathrm{h^{-1}Mpc}$. At $k > 0.1\,\mathrm{h\,Mpc^{-1}}$, where the effect of the window function is negligible, the combined effect of bias/growth is visible, with a slightly larger amplitude for the two higher-z samples (higher bias due to larger mean luminosity). On large scales ($k < 0.1\,\mathrm{h\,Mpc^{-1}}$) the four power spectra are all compatible to each other within statistical errors, showing that small differences between the four window functions are smaller than statistical fluctuations. As we have seen in the previous section, the selection effects due to the observing strategy and instrumental limitations affect the properties of clustering and the relative power spectrum.

We correct for the SPOC effect using the “local” TSR weight of equation 5.8 considering the ratio in small angular rectangle with size $\Delta RA \times \Delta DEC = 60 \times 100\,\mathrm{arcsec^2}$

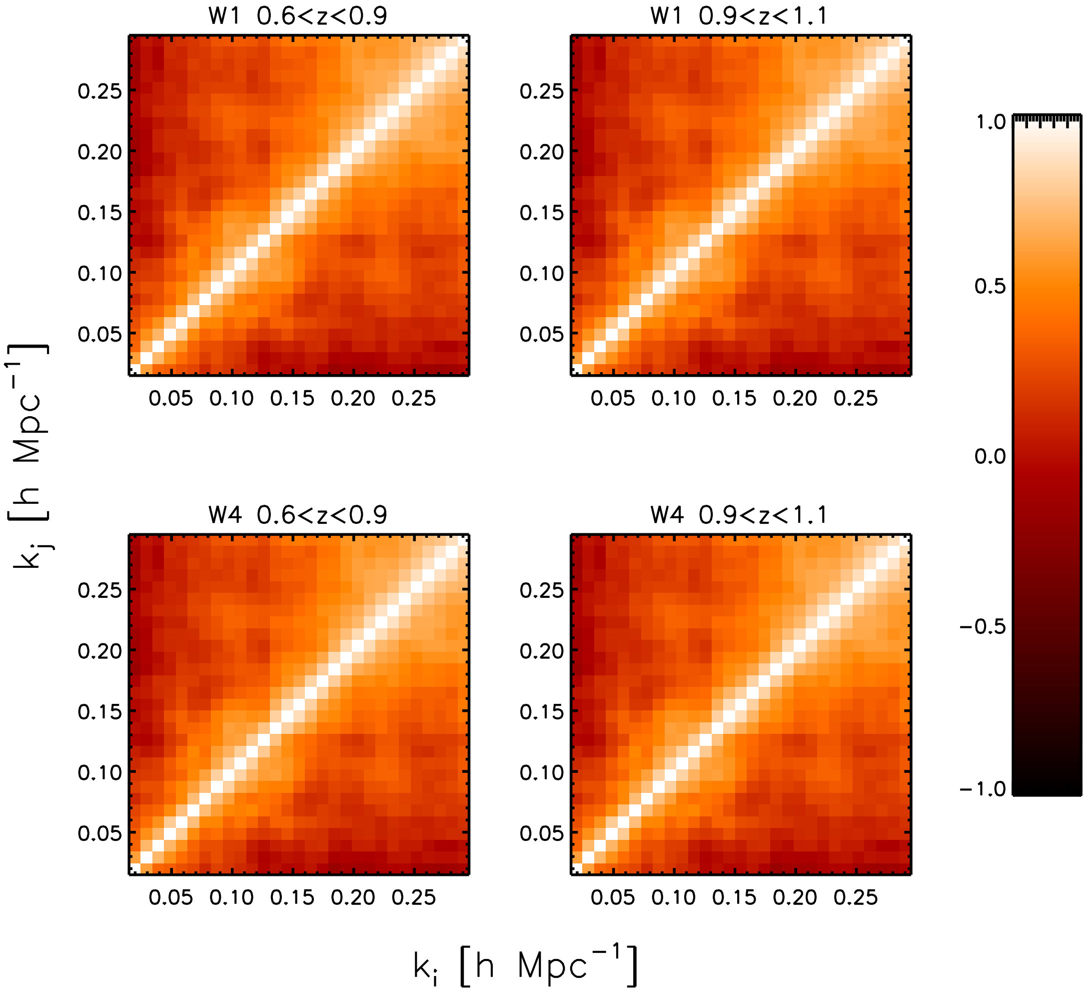

Figure 7.3: four correlation matrix in the two redshift bins in each field (W1 and W4) measured with the Pinocchio mock catalogues as described in section 5.6. The k-space range used to estimate r_{ij} is between $0.01 < k < 0.3$ $[h\,\mathrm{Mpc}^{-1}]$ and will be motivated later in the text.

between galaxies in the CFHTLS photometric catalogue and those observed in VIPERS. In each subsample, the difference in the recovered power spectrum with and without the SPOC correction is very small and, contrarily to what we observe in mock samples, does not seem to show any systematics. The reason could be that in mock samples we average the power over 26 (W1) and 31 (W4) realisations highlighting the systematic effect of the SPOC correction while, for one realisation, the statistical fluctuations are

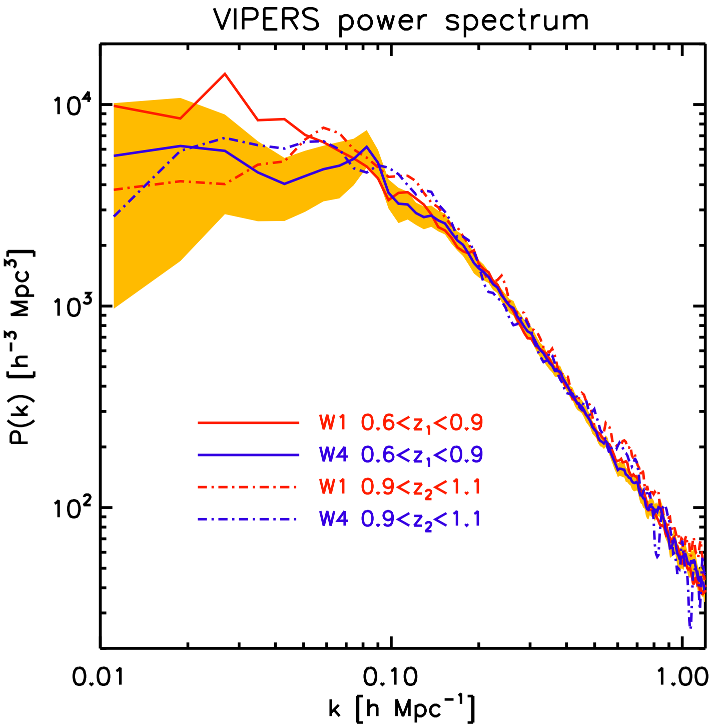

Figure 7.4: comparison between the four galaxy power spectra measured in each VIPERS subsample with errors measured from the dispersion of the power in the MultiDark mock catalogues and plotted for the W4 field in the lower z_1 redshift bin.

still dominant. We show in figure 7.5 only the case for the W1 field in the lower redshift bin z_1 where the small difference between the power spectrum with and without the SPOC correction is outlined in the bottom panel. Also, SPOC may have a larger effect in mocks due to a slightly different number density of galaxies.

7.1 Cosmological results

Since the total volume of VIPERS is comparable with the 2dFGRS we should be able to recover similar constraint on cosmological parameters. Except for the BAO signal observed in the 2dFGRS power spectrum but difficult to detect in VIPERS due to the anisotropy of the window function (as showed in figure 5.4), we try to constrain the matter density $\Omega_{\rm M}$ and the baryonic fraction $f_{\rm B} = \Omega_{\rm B}/\Omega_{\rm M}$ as Percival et al. (2001) and later Cole et al. (2005) did for the 2dFGRS.
Again to compare the four measured power spectra with a model, we need to include RSD, correct for the assumed fiducial cosmology, convolve with the VIPERS window function and subtract the integral constraint.

We assume a ΛCDM cosmology with flat curvature $\Omega_{\rm k} = 0$. The theoretical model

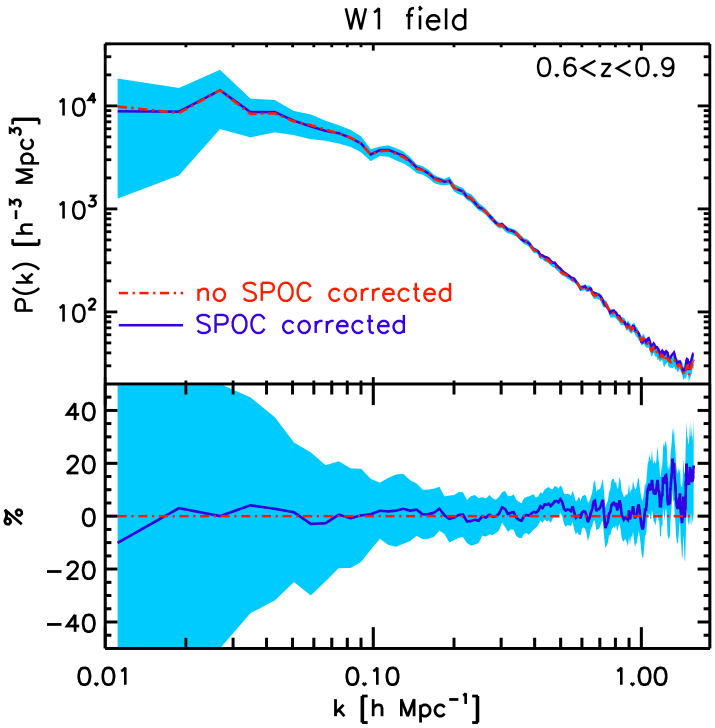

Figure 7.5: *Top panel*: measured power spectrum in the W1 field and in the first z_1 redshift bin with (blue dashed line) and without (red solid line) the inclusion, in the estimation of the power, of the SPOC correction with the "local" TSR. The weight for each galaxy has been calculated as the ratio between the number of points in the photometric catalogue respect to those in the VIPERS sample in a rectangular shape with size $\Delta\mathrm{DEC} \times \Delta\mathrm{RA} = 100'' \times 60''$ arcsec^2 centered at each galaxy selected by SPOC. *Bottom panel*: ratio between the recovered power without SPOC correction and the one corrected, plotted in order to outline the small difference between the two contrarily to what we observe in mock catalogues.

power spectrum is generated with CAMB plus HALOFIT imposing flat priors on the baryonic fraction between $1\% < \mathrm{f_B} < 50\%$, the matter density, $0.1 < \Omega_\mathrm{M} < 0.9$ and bias $0 < \mathrm{b} < 2$. The other parameters, as the Hubble constant and the primordial index, are fixed to the Planck values (Planck Collaboration et al., 2013): $\mathrm{h} = 0.6711$ and $\mathrm{n_s} = 0.9624$. As opposed to the approach adopted for the mocks, where we fix the amplitude of the power spectrum with the σ_8 of the simulation breaking the degeneracy with the bias, for data we fix the scalar amplitude $\mathrm{A_s}$ to the best-fit Planck prior ($\mathrm{A_s} = 2.21536 \times 10^{-9}$) because this quantity is well constrained by CMB observations.

Before convolving the model with the VIPERS window function, we interpolate the monopole theoretical power spectrum on a 3D k-space grid in order to introduce the anisotropy of RSD. RSD are modelled with the dispersion model of equation 5.5, assuming General Relativity for the Kaiser term and letting the β parameter to vary as a function of matter density and bias. We impose, for the dispersion of pairwise velocities and redshift errors, the Gaussian prior mentioned in Bel et al. (2014) with $\bar{\sigma}_{\rm TOT} = 514\,{\rm km\,s^{-1}}$ and deviation $\sigma_{\sigma_{\rm TOT}} = 24\,{\rm km\,s^{-1}}$. The $\sigma_{\rm TOT}$ used takes into account the Gaussian dispersion for redshift errors of the order of $\sigma_{\rm cz} = 141(1+{\rm z})\,{\rm km\,s^{-1}}$.

For each set of parameters tested, we rescale the model to the MultiDark fiducial cosmology correcting the amplitude of the power and the k-space grid as in equation 6.10. The four VIPERS window functions have been calculated with four very dense random catalogues with the same angular and redshift properties of the analysed VIPERS subsamples and the same redshift distribution calculated with equation 7.1 to mimic the mean density of VIPERS 4.0. In figure 7.6, we show the behaviour of the window functions projected over the main coordinates, after rotating the unclustered catalogue. The different features in the window at higher modes is due to the particular pattern introduced by the angular mask (gaps and quadrants) which generates different periodic scales in the distribution of galaxies dependent on redshift. The overall behaviour of the windows, however, is very similar for both redshift bins and both fields because the four VIPERS subsamples are very similar in terms of volume and shape. This explains why we observe four compatible power spectra in different VIPERS sub-volumes.

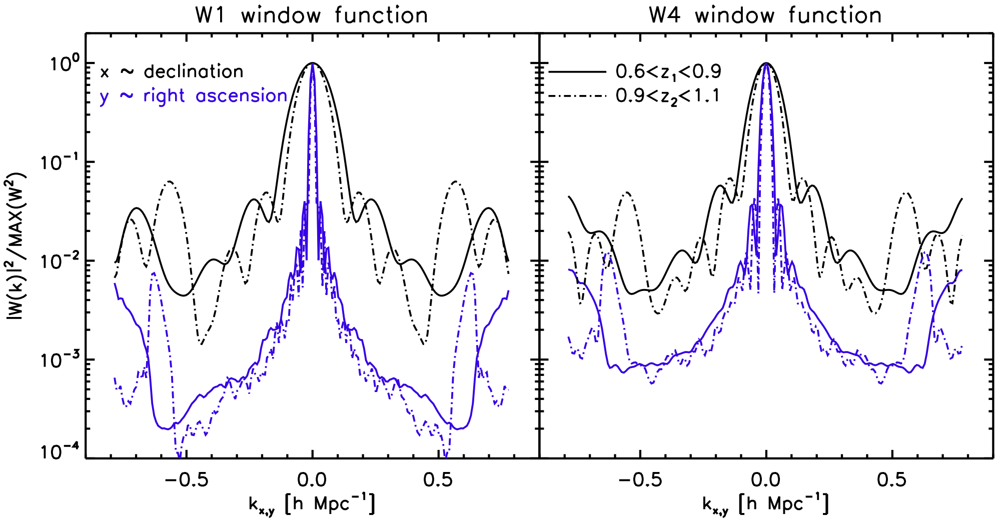

Figure 7.6: *Left panel*: comparison between the projected 3D window function in the W1 field along the ~right ascension direction (black line) and the ~declination one (blue line). Solid and dashed lines represent respectively the lower redshift bin z_1 window and the higher z_2 one. *Right panel*: same as the left panel but for the W4 field.

The posterior likelihood distribution is estimated running an MCMC on the combined W1 and W4 measurements (accounting for the different window function), but treating separately the two redshift bins due to the different bias parameter in z_1 and z_2. The fit is performed in a k-range between $0.01 < {\rm k} < 0.3$ $[{\rm h\,Mpc^{-1}}]$ with linear binning size $\Delta k = 0.01\,h\,Mpc^{-1}$. This choice is explained because we want to use all the information available in the power spectrum estimation limiting the minimum mode to the largest possible scale we can sample in the redshift direction and extending the maximum fitting mode to the frequency $k = 0.3\,h\,Mpc^{-1}$. Higher modes, in fact, are dominated by shot-noise in the high redshift bin, the power is not properly corrected for the SPOC effect, we have to deal with strong non-linearities which are difficult to model and the scale independence of the bias is no more true.

The chi-square in the $\Omega_M - f_B$ plane, estimated with the covariance matrix of figure 7.3, is shown in figure 7.7 with the red contours representing the degeneracy between the two parameters in the lower redshift bin z_1 and the blue ones showing results for the redshift bin z_2. The two redshift bins give consistent results.

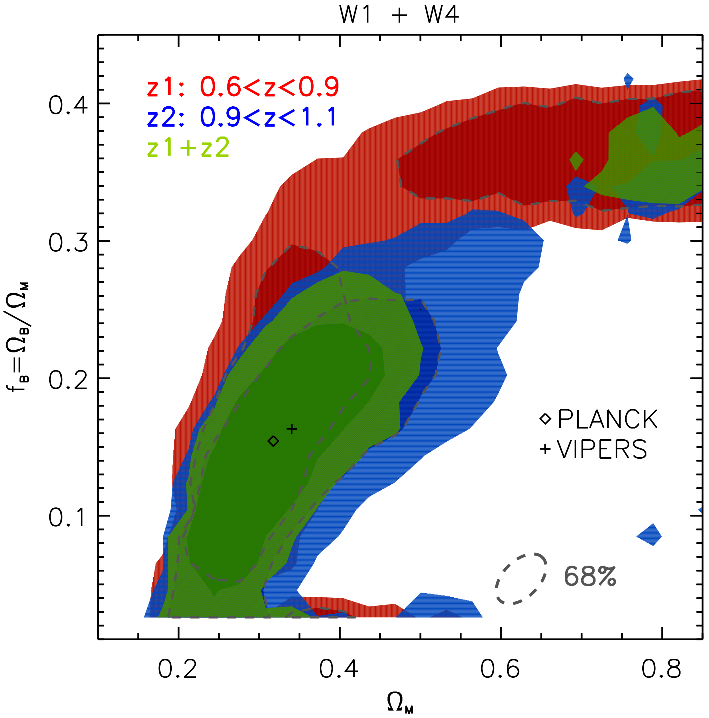

Figure 7.7: chi-square contours in the $\Omega_M - f_b$ plane for a combined W1+W4 analysis in the two redshift bins: red contours are showing the 68% − 90% confidence levels measured with the recovered power spectrum with galaxies in $0.6 < z_1 < 0.9$, while blue contours show the case for $0.9 < z_2 < 1.1$. The green region corresponds to the joint likelihood between the chi-squares in the two redshift bins. The cross point is the position of the best fit calculated with the current analysis while the diamond one shows the best prior estimated with the Plank analysis.

We also combine the posterior likelihoods in the two redshift bins in order to estimate the matter density and the baryonic fraction (green region in figure 7.7) obtained from the full VIPERS sample. The best fit, estimated with the VIPERS power spectrum, is consistent with the Planck value of $\Omega_{\rm M} = 0.3175$ and $f_{\rm B} = \Omega_{\rm B}/\Omega_{\rm M} = 0.049/0.3175$ (Planck Collaboration et al., 2013) within statistical uncertainties.

A simple explanation for the shape of the chi-square contours between the matter density and the baryonic fraction could be given considering two points in the parameter space, one with low $\Omega_{\rm M}$ and one with a high value for the matter density. In the low-$\Omega_{\rm M}$ model, the peak at $k_{\rm eq}$ is at low modes because the time of equivalence happens at late time when the size of the horizon is quite large. At the scales sampled by the VIPERS catalogue, the amplitude at various modes of the power is well constrained by this kind of model without adding too much suppression due to the baryonic damping of the power (low $f_{\rm B}$). If the matter density is quite high, the theoretical model shifts toward higher modes increasing the power at the VIPERS measured scales. In this case to compensate the increase of power, the good fit is given by models with high baryonic fraction necessary to suppress a bit the power. Therefore we have an idea of what is happening in the $\Omega_{\rm M} - f_{\rm B}$ chi-square where the best fit models are the ones with low matter density and low baryonic fraction and (vice-versa) high $\Omega_{\rm M}$ and high $f_{\rm B}$. Avoiding the high matter density values, we plot in figure 7.8 only the low-$\Omega_{\rm M}$ best fit. Again the consistency between the chi-square in the two redshift bins is well visible. We also notice the higher value for the bias in the z_2 case which is not so clear in figure 7.4: the amplitude of dark matter structures is in fact lower at higher redshift where however the effect of the bias is stronger. Accounting for the different growth factor in the theoretical power spectrum for the two redshift bins, it is possible to extract the right value for the bias which is higher in z_2, since at higher redshift we observe more luminous, and so more biased, galaxies.

7.1.1 Comparison with 2dFGRS

Since VIPERS and 2dFGRS are very similar in term of volumes, we can combine the results looking at the VIPERS sample as an extension of the 2dF survey at higher redshift. Percival et al. (2001) measured the galaxy power spectrum in the local Universe with an incomplete 2dF catalogue composed of $\sim 150,000$ galaxy redshifts. The theoretical power spectrum has been computed using the transfer function fitting formulae of Eisenstein and Hu (1998) and convolved with a mono-dimensional window function, justified by the compactness and isotropy of the 2dFGRS volume. Since the fit has been performed in a k-range $0.02 < k < 0.15$ $[h\,{\rm Mpc}^{-1}]$, RSD and non-linear effects (as the non linearity of the bias) could be rightly considered to have a negligible impact on the power spectrum shape. The likelihood analysis is then performed letting the baryonic fraction, the matter density and the Hubble factor free. Cole et al. (2005) measured the galaxy power spectrum of the full 2dFGRS sample using the complete catalogue with $\sim 220,000$ galaxies constraining the same cosmological parameters but fixing the Hubble constant to $h = 0.72$. We run again an MCMC for the two redshift bins fixing the Hubble constant to $h = 0.72$ (while the other parameters fixed to Planck) to be consistent with Cole et al. (2005) and plot the chi-square contours in figure 7.9 where we rescale the matter density, in the abscissa coordinate, with the Hubble constant in

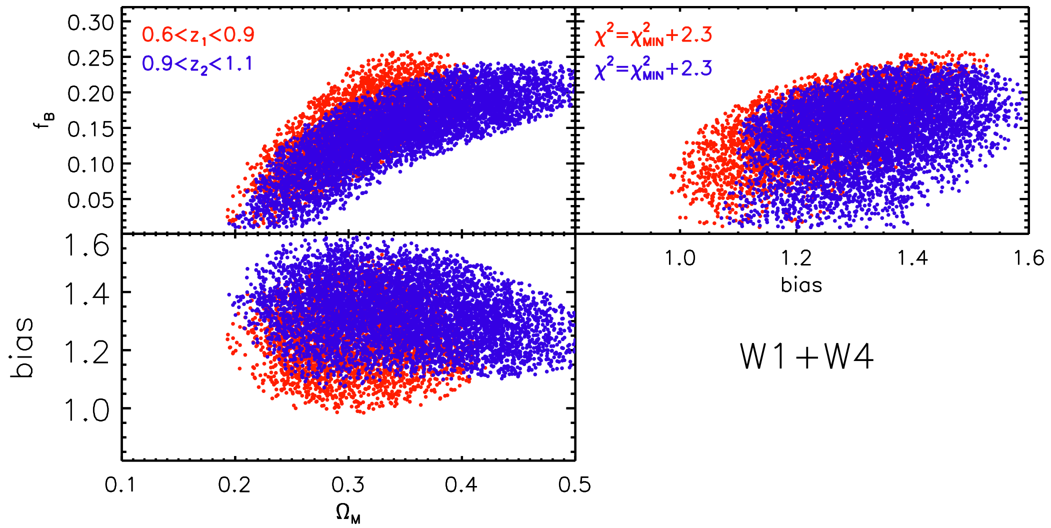

Figure 7.8: best fit for the cosmological parameters $\Omega_{\rm M}-{\rm f_B}-{\rm b}$ showing the 68 per cent confidence level calculated using a combined W1+W4 MCMC analysis in the two redshift bins ($\rm z_1$ red points, $\rm z_2$ blue points. We avoid in these plots the degeneracy, in the lower redshift bin $\rm z_1$, for high $\Omega_{\rm M}$ values.)

order to have a fair comparison with the 2dFGRS results.

The degeneracy in the $\Omega_{\rm M}-{\rm f_B}$ plane, observed in Percival et al. (2001) (and re-measured by Cole et al. (2005) using the same data of 2001 but with a new methodology and fixing h, dotted lines in figure 7.10) is similar to what we recover in VIPERS for the lower redshift bin. In the 2dFRGS, the larger volume analysed for the complete sample allows Cole et al. (2005) to break the degeneracy as shown by solid lines in figure 7.10. In VIPERS, the degeneracy is broken, in addition to a strong reduction of statistical errors, thanks to a joint likelihood between the two redshift bins (green chi-square region in 7.9). The 2dFGRS errors in the $\Omega_{\rm M}-{\rm f_B}$ chi-square from Percival et al. (2001) are similar to our measurements due to a combination of effects. The bigger volume of the 2dFGRS strongly reduces statistical cosmic variance even though the closeness of the sample, which surveys the local Universe, constrains their fit to a maximum mode of $\rm k_{max} \sim 0.15\,h\,Mpc^{-1}$ to avoid non-linearities. On the contrary, VIPERS is smaller in terms of volume but the high redshift observed allows us to push the maximum fitting mode up to $\rm k_{max} = 0.3\,h\,Mpc^{-1}$ using more information from the power spectrum analysis. The power spectrum at higher redshift with respect to the local Universe is, in fact, less affected by non-linearities as shown in figure 1.3.

We compare our measurement (figure 7.7) also with the WiggleZ result in Blake et al. (2010) where they find a similar behaviour in the chi-square contours between the matter density and the baryonic fraction (figure 7.11). Blake et al. (2010) performed an approach in measuring and modelling the galaxy power spectrum similar to our, using 56,159 redshifts of bright emission-line galaxies at effective redshift $\langle z\rangle \sim 0.6$ and

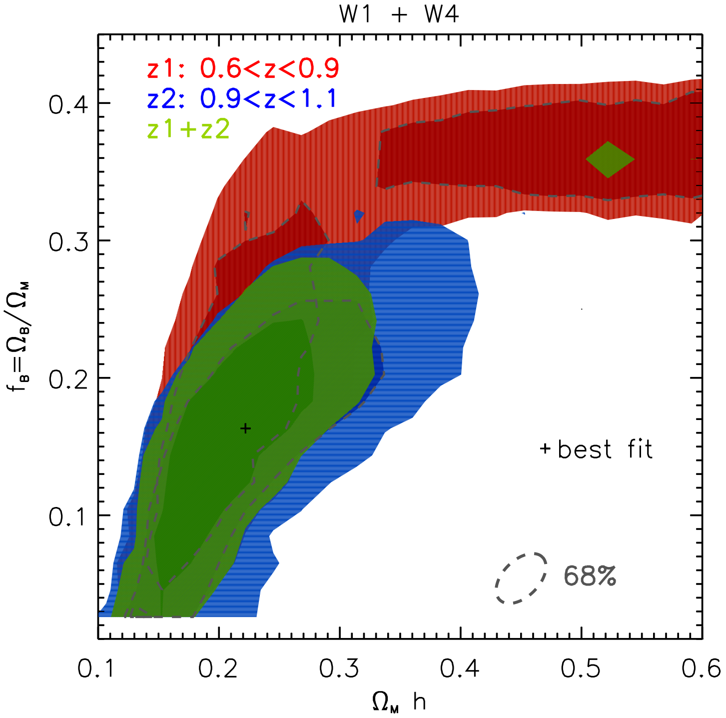

Figure 7.9: chi-square contours similar to figure 7.7 but obtained imposing a value of $h = 0.72$ for the Hubble constant. For the abscissa coordinate, $\Omega_M h$, we simply multiply the matter density with the Hubble constant $h = 0.72$ in order to have a fair comparison with the 2dFGRS results.

fitting the power at scales $k_{max} < 0.4\,h\,Mpc^{-1}$ fixing values for the Hubble parameter $h = 0.72$, scalar index of spectral fluctuations $n_s = 0.96$ and normalisation $\sigma_8 = 0.8$, and vary only the matter density Ω_M and the baryon fraction f_B. Their confidence contours are very small due to the WiggleZ enormous volume. Moreover, they can extract the maximum information from the power spectrum (extending the maximum fitting mode up to $k_{max} = 0.4\,h\,Mpc^{-1}$) given the high redshift sample and consequently pushing non-linearities toward smaller scales.

7.1.2 Consistency with Planck

The power spectrum shape information encoded in the transfer function depends mainly on the combination of the matter density with the Hubble constant, $\Omega_M h^2$, as well as on the combination of the baryonic density and the Hubble constant, $\Omega_B h^2$.
Letting free the baryonic fraction is a coherent choice because combination of these two cosmological parameters: $f_B = \frac{\Omega_B}{\Omega_M} = \frac{\Omega_B h^2}{\Omega_M h^2}$. The power spectrum observed today in the local Universe, or at any other redshift, is also characterised by the growth factor D(z) (equation 1.20), which depends independently on the matter density and the Hubble

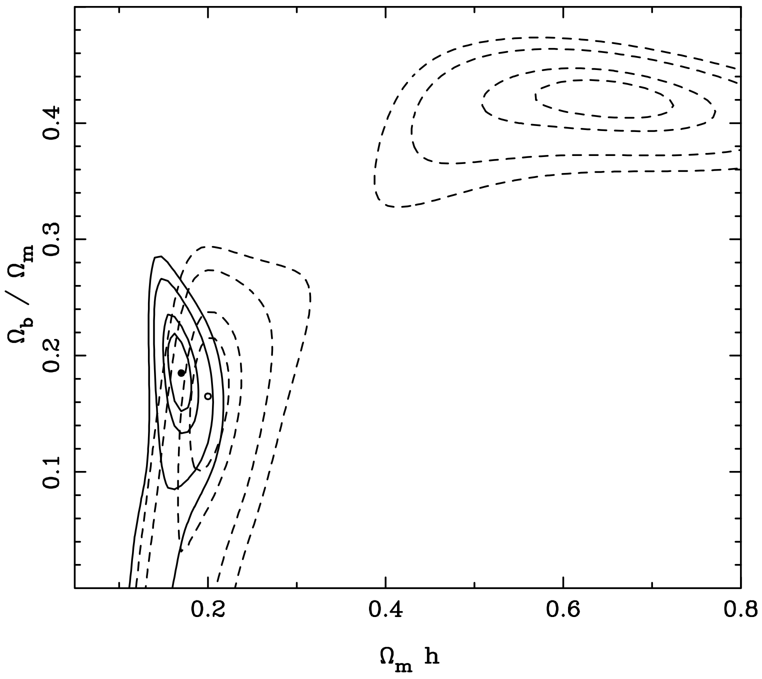

Figure 7.10: degeneracy between the baryonic fraction and the matter density (times the Hubble factor) in the power spectrum analysis of the 2dFGRS. The dashed contours represent the confidence levels re-calculated by Cole et al. (2005) with the intermediate sample used for the analysis in Percival et al. (2001). Solid lines show the chi-square levels computed by Cole et al. (2005) with the complete 2dFGRS catalogue.

constant. Moreover, even the growth rate f of structures, which enters in the modelling of the observed power spectrum through the Kaiser term in redshift-space distortions, depends on the matter density parameter alone with equation $f \propto \Omega_M^\gamma(a)$. For these reasons, we choose to separate the two parameters in $\Omega_M h^2$ fixing h and constraining Ω_M. The Hubble constant is however fixed to the Planck value, h = 0.6711, a quantity which is not directly measured by Planck through the angular power spectrum analysis but inferred assuming the Planck cosmology. We therefore want to understand how much our result is consistent with Planck, i.e. how the assumption of the Planck best fit Hubble constant determines our recovered best fit for the matter density and the baryonic fraction.

We compare again the four measured power spectra, in the two redshift bins (z_1 and z_2) and for each field (W1 and W4), with different theoretical models through a chi-square technique to constrain Ω_M and f_B. Assuming a ΛCDM flat cosmology with a cosmological constant $w = -1$ and imposing the scalar amplitude and spectral index to the Planck best fit, we consider an Hubble constant of $h = 0.72$ different with respect to the Planck derived value of $h = 0.67$.

The resulting chi-square contours combination of the two fields, W1 and W4, and for each redshift bin are plotted in figure 7.12 and compared to the results obtained

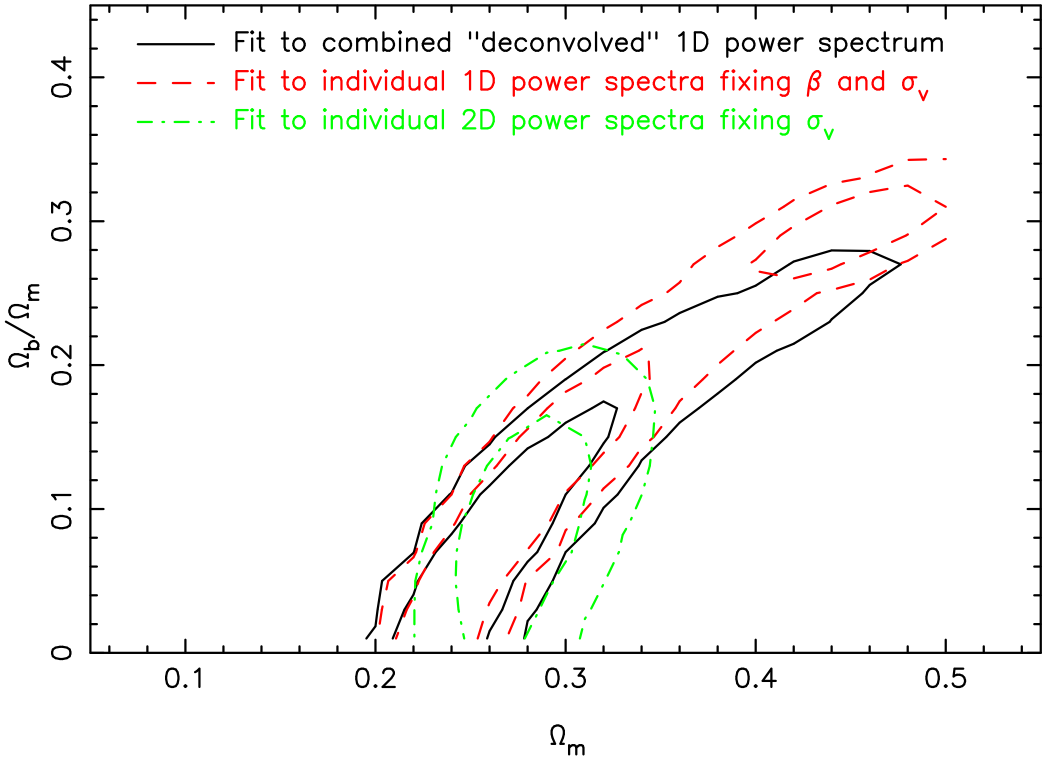

Figure 7.11: probability contours of Ω_M and f_B fitting the WiggleZ survey power spectra in different redshift bins using three different approaches. For the black solid contours, they have stacked all the monopole power spectra measured in different redshift bins and in different regions on the sky and deconvolved the result (made simpler by the isotropy of their window), red dashed contours show the joint likelihood between each single power spectrum compared with a convolved model, while dot-dashed green contours are the fit between the 2D power spectrum analysed in the tangential and radial component of the wavenumber respect to the line-of-sight and a convolved model.

assuming the Planck best fit for the Hubble constant. We plot the $1-2\,\sigma$ contours as a function of f_B and a combination of Ω_M and h, $\Omega_M h^2$ (obtained simply multiplying the matter density with the assumed Hubble constant).

It is important to remark here that in linear theory, if distances are measured in unity of h^{-1}Mpc (and consequently frequencies in $h\,\mathrm{Mpc}^{-1}$), the degeneracy in the $\Omega_M - h$ plane roughly develops along the line $\Omega_M h = $ constant as shown by equation 1.22, where the transfer function depends on $q = k/(\Omega_M h^2 \mathrm{Mpc}^{-1}) = k[h\,\mathrm{Mpc}^{-1}]^{-1}/(\Omega_M h)$. The non-linear power spectrum is, instead, affected by the variance σ_8^2 (equation 1.26) which depends on the combination $\Omega_M h^2$ due to the growth factor. Bel et al. (2014) clearly show, in their figure 16, that the degeneracy between the matter density and the Hubble constant for the power spectrum follows mainly the degeneracy $\Omega_M h^2 = $ constant, justifying our choice of comparing results as a function of the quantity $\Omega_M h^2$.

In both redshift bins, we observe a small, almost negligible, shift of the chi-square contours once we vary the assumed Hubble constant, meaning that the modelling of the power spectrum shows a small residual dependence on h. In figure 7.13 we plot the combined chi-square for z_1 and z_2 showing a good agreement between our result and the Planck best fit even in the case of $h = 0.72$, deducing that our measurement of cosmological parameters depends weakly on the choice of the Hubble constant. We conclude that our results are totally consistent with the Planck ones without being affected by the assumed priors.

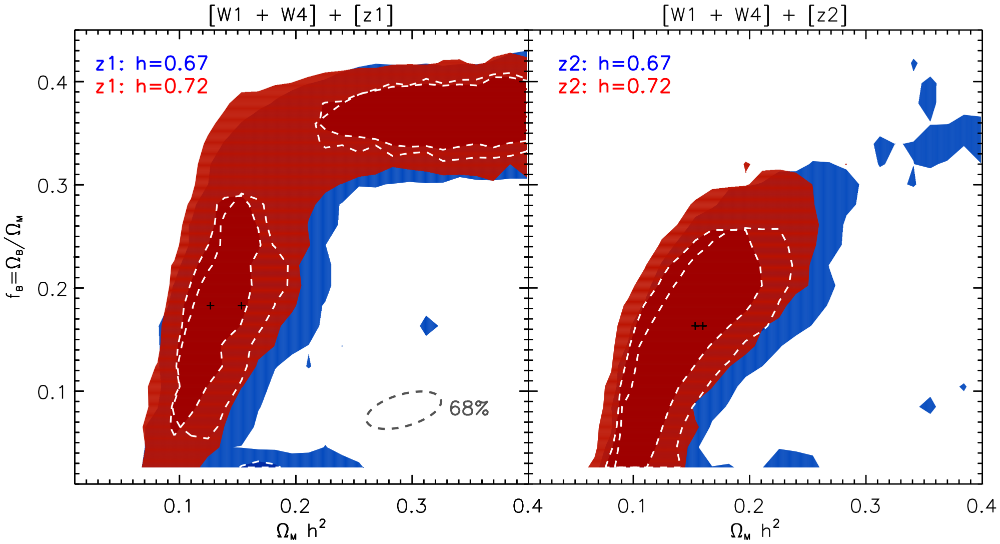

Figure 7.12: *Left panel*: combined $1-2\sigma$ chi-square contours in the $\Omega_M - f_B$ plane evaluated with the measured power spectra in the lower redshift bin z_1 for the field W1 and W4. The blue region corresponds to chi-square measured fixing the Hubble constant to the Planck value of $h = 0.67$, while the red region shows the same contours but with the Hubble constant fixed to $h = 0.72$ (same prior used by Cole et al. (2005)). *Right panel*: same as the left panel but for the higher redshift bin z_2.

7.1.3 Varying the maximum fitting mode

The minimum mode $k_{min} = 0.01 \, h \, Mpc^{-1}$, related to the largest scale sampled with the VIPERS distribution of galaxies and chosen for the fit, does not affect the chi-square contours and the best fit since errors introduced by cosmic variance at those scales are big. In this subsection, we analyse the dependence of the cosmological parameters constrained with the VIPERS power spectrum from the maximum mode sampled. The value chosen for the maximum frequency k_{max} used for the fit of the measured power spectrum could alter the estimation of the matter density and the baryonic fraction due to the tiny errors recovered at small scales.

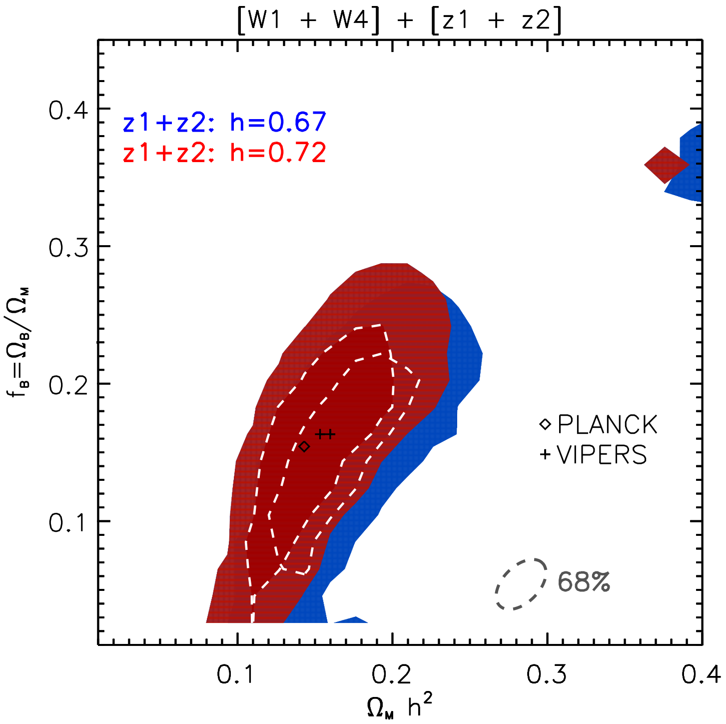

Figure 7.13: same as figure 7.12 but with the chi-square in the two redshift bins combined. Note the consistency with the Planck best fit even when we evaluate the chi-square fixing the Hubble constant to a different value respect to the one recovered by Planck.

We want to estimate how robust is our choice of $\mathrm{k_{max}} = 0.3\,\mathrm{h\,Mpc^{-1}}$ considering two other maximum frequencies, $\mathrm{k_{max}} = 0.2\,\mathrm{h\,Mpc^{-1}}$ and $\mathrm{k_{max}} = 0.25\,\mathrm{h\,Mpc^{-1}}$. The fit has been estimated with the same minimum frequency as before and the same step in k-space $\Delta\mathrm{k} = 0.01\,\mathrm{h\,Mpc^{-1}}$ obtaining $\mathrm{N} = 19$ bins for $\mathrm{k_{max}} = 0.2\,\mathrm{h\,Mpc^{-1}}$ and $\mathrm{N} = 24$ for $\mathrm{k_{max}} = 0.25\,\mathrm{h\,Mpc^{-1}}$.

Results for the $68\% - 95\%$ chi-square confidence levels are plotted in figure 7.14 considering separately the two redshift bins. When $\mathrm{k_{max}}$ is lower, the errors increase since, decreasing the maximum mode fitted, the information is lower. Further, for smaller $\mathrm{k_{max}}$ the chi-square contours are shifting toward higher Ω_M best fit. The estimation of Ω_M and $\mathrm{f_B}$ seems to be more robust for the high redshift bin with the best fits consistent with each other within statistical uncertainties. This could be explained by the fact that the power spectrum, at higher redshift, is linear up to smaller scales because the clustering of matter was less strong in the past. The mode where the power becomes dominated by non-linearities increases at higher redshift giving, for the high redshift bin, a more reliable modelling of the power in linear theory. Moreover, the shapes of the chi-square contours in the $\Omega_\mathrm{M} - \mathrm{f_B}$ plane are similar in both redshift bins and for all

the k_{max} considered except for the case z_1 and $k_{max} = 0.3\,h\,Mpc^{-1}$: at the increase of the value of Ω_M does not correspond an increase of the baryonic fraction in the high Ω_M regime (a similar behaviour has been observed by Percival et al. (2001) in the 2dF survey, figure 7.10). This could be again due to non-linearities which are stronger at lower redshift z_1 and at smaller scales k_{max}. In this case, our modelling of non-linearities is far from being correct in particular for higher values of the matter density: the shift of the model power spectrum toward higher k-modes, due to the increasing of Ω_M, is no more compensated by an increase of f_B but by an insufficient correction for non-linearities.
The combination of the chi-square estimated in the two redshift bins is plotted in figure 7.15. Despite the shift toward higher matter density related to the decreasing of the maximum mode where the fit has been performed, the best fits are consistent between each other. A possible explanation for the discrepancy between chi-square contours associated to the maximum fitting mode could be the assumption of the galaxy bias to be constant which is not true once we study very small scales.

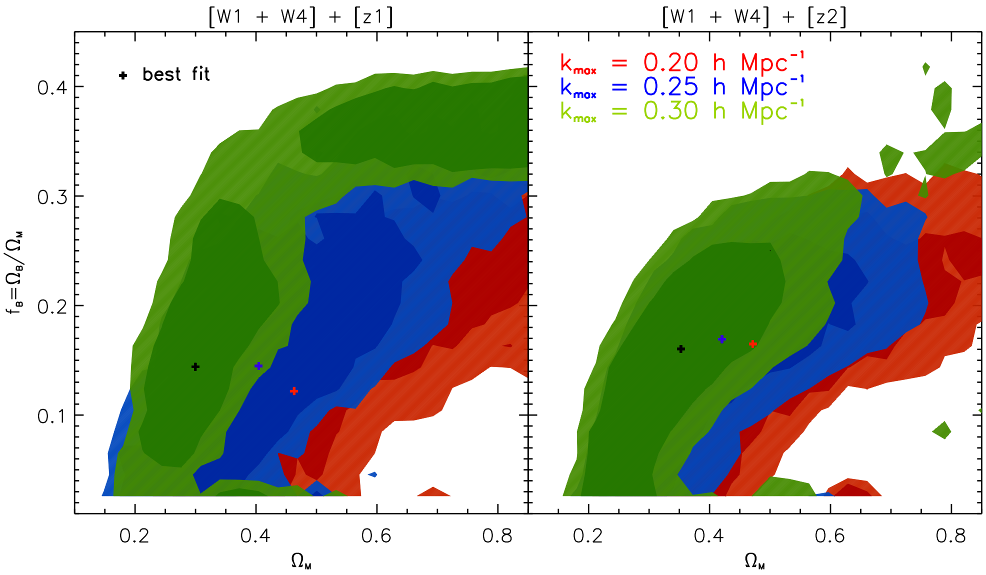

Figure 7.14: *Left panel*: 68% − 95% confidence levels for the chi-square estimated, in the first redshift bin $0.6 < z_1 < 0.9$ and for both fields (W1 and W4), varying the maximum mode where the fit has been performed. The green contours are associated to a maximum fitting mode $k_{max} = 0.3\,h\,Mpc^{-1}$, the blue ones to $k_{max} = 0.25\,h\,Mpc^{-1}$ and red to $k_{max} = 0.2\,h\,Mpc^{-1}$. *Right panel*: same as the left panel but with the power measured in the second redshift bin $0.9 < z_2 < 1.1$.

7.1.4 Estimating Ω_M

Given the big degeneracy between the baryonic fraction and the matter density in the determination of the overall shape of the matter power spectrum, we decide to extract only the Ω_M parameter fixing the f_B quantity to known prior. For the estimation of

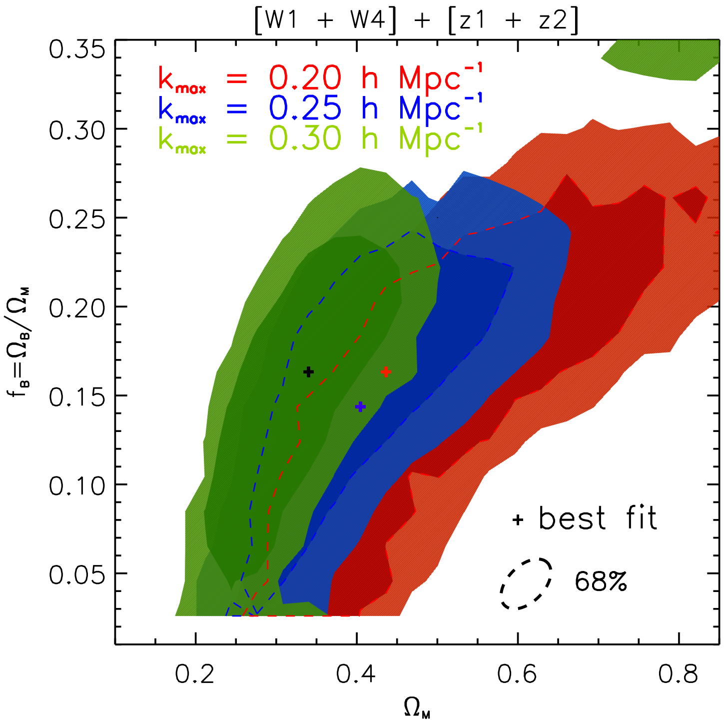

Figure 7.15: same as figure 7.14 but with a joint likelihood between the two redshift bins.

Ω_M, we run an MCMC for each redshift bin, z_1 and z_2, combining the chi-square of the two fields. We assume a ΛCDM cosmology with the the accelerated expansion caused by a cosmological constant term with $w = -1$ and flat space $\Omega_k = 0$. We allow to vary, in our power spectrum modelling, some cosmological parameters fixing Gaussian prior to the best available estimation as the spectral index, $n_s = 0.9616 \pm 0.0094$, and scalar amplitude, $\ln(10^{10} A_s) = 3.103 \pm 0.072$, from Planck results (Planck Collaboration et al., 2013), the Hubble constant $h = 0.738 \pm 0.024$ from the measurements of the Hubble Space Telescope (HST, Riess et al., 2011) and the baryonic matter density parameter $\Omega_B h^2 = 0.0213 \pm 0.0010$ from the Big Bang Nucleosynthesis (BBN, Pettini et al., 2008). The bias is again considered linear and scale independent and we assume a Gaussian prior even for the effective pairwise dispersion velocity $\bar{\sigma}_{TOT} = 514 \pm 24\,\mathrm{km\,s^{-1}}$, as estimated from the VIPERS data themselves.

The marginalised likelihood for the matter density estimated in z_1 and z_2 is plotted in figure 7.16: results are consistent between the two redshift bins. A joint likelihood between the two redshift bins has been done in order to measure the best fit with the related 1σ error for the matter density obtaining $\Omega_M = 0.272^{+0.027}_{-0.030}$ perfectly in agreement with the result of $\Omega_M = 0.270^{+0.029}_{-0.025}$ obtained with the VIPERS clustering ratio in Bel et al. (2014) and with the WMAP7 determination of $\Omega_M = 0.267 \pm 0.029$ (Larson et al.,

2011).
The difference between our measurements and the Planck one for the matter density is due to the different prior assumed for the Hubble parameter. Rescaling our Ω_{M} best fit with the Planck Hubble constant of $\mathrm{h} = 0.6711$, we obtain:
$\Omega_{\mathrm{M}} = 0.272 \times \left[\frac{0.738}{0.6711}\right]^2 = 0.329^{+0.032}_{-0.036}$ consistent now with the Planck best fit of
$\Omega_{\mathrm{M}} = 0.3175$ (the rescaling factor is justified by the dependence of the power spectrum on $\Omega_{\mathrm{M}}\mathrm{h}^2$ described in section 7.1.2). Best-fit parameters with the 1σ error are shown in table 7.1.4 for a joint likelihood between the two redshift bins. In all cases, the measured parameters are consistent with the Gaussian prior. The bias factor is, as expected from the previous estimation, slightly higher in the z_2 sample.

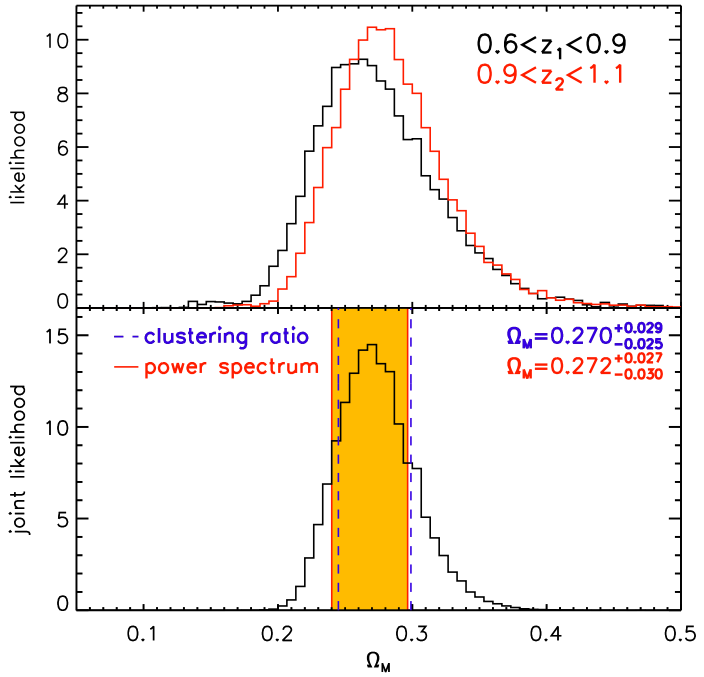

Figure 7.16: *Top panel*: number density in each Ω_{M}-bin computed with a combined chi-square for the two fields, W1 and W4, for the matter density in the two redshift bins: $0.6 < \mathrm{z}_1 < 0.9$ black histogram, $0.9 < \mathrm{z}_2 < 1.1$ red histogram. *Bottom panel*: joint number density combining the results in the two redshift bins. The shaded area corresponds to the 68% confidence level measured with this analysis while the dashed blue lines correspond to the results obtained with the clustering ratio technique.

	Ω_M	$\Omega_\mathrm{b}\mathrm{h}^2$	h	$\mathrm{n_s}$	$\ln(10^{10}\mathrm{A_s})$	σ_TOT [$\mathrm{km\,s^{-1}}$]	$\mathrm{b}\,(\mathrm{z_1}\,/\,\mathrm{z_2})$
prior	$0.1-0.9$	0.0213 ± 0.0010	0.738 ± 0.024	0.9616 ± 0.0094	3.103 ± 0.072	514 ± 24	$0-2$
best fit	$0.272^{+0.027}_{-0.031}$	$0.0211^{+0.0010}_{-0.0004}$	$0.735^{+0.018}_{-0.016}$	$0.9630^{+0.0054}_{-0.0088}$	$3.096^{+0.046}_{-0.057}$	522^{+16}_{-18}	$1.13^{+0.21}_{-0.18}\ /\ 1.25^{+0.20}_{-0.15}$

Table 7.2: Gaussian, flat prior and best fit for the parameters used in the estimation of the matter density.

Chapter 8

Conclusions

The power spectrum statistic is important to extract cosmological information from the three dimensional distribution of galaxies. A lot of cosmological quantities can be extracted from the power spectrum such as the total amount of matter and baryons in the Universe which strongly affect its shape. The BAO signal associated with oscillating features in the monopole power spectrum and a detailed study of the anisotropic clustering can lead respectively to constraints on dark energy and theory of gravity. The modelling of the power spectrum is not simple being affected by non-linear evolution, bias and redshift-space distortions. These effects complicate the relation between the observations and the predictions of linear perturbation theory, making the interpretation of these measurements in terms of constraints on cosmological parameters more difficult.

In the last 10 years, numerous surveys have been constructed to sample larger and larger volume of the Universe in order to reduce statistical errors related to the estimation of cosmological quantities and to improve our modelling of the power spectrum.

Even the measurement of the galaxy power spectrum is complicated because it has to account for all the selection effects of the survey.

In this thesis, we apply the power spectrum analysis to the VIPERS sample, a spectroscopic survey that has been designed to analyse the clustering of the z ~ 1 Universe. In the estimation of the galaxy power spectrum, the shape and volume of a survey have a dominant role. Specifically, VIPERS has a small angular aperture along the declination, even more at the current stage. We fully characterised and corrected the effect of the window function using the suite of 26 and 31 (respectively for the W1 and W4 field) VIPERS mock samples built from the MultiDark simulation and described in de la Torre et al. (2013).

A number of other tests have been performed to include the VIPERS sampling strategy and other subtleties, with excellent results. This makes us confident that we are able to correctly model the effect of the survey geometry and sampling and thus extract cosmological information from the measured power spectrum.

Another important question to deal with, is the inclusion of redshift-space distortions in our modelling. The theoretical model has been corrected for RSD effects before performing the convolution with VIPERS window function. This approach is usually avoided in literature because of the 3D-convolution which is time-consuming.

To test the robustness of our approach, we fit the measured power spectrum from each of the VIPERS mock catalogue with a model using the standard chi-square technique. We measure the averaged power spectrum for both W1 and W4 field, in redshift space. We vary the matter density parameter and the bias factor (assumed to be linear and scale independent), and fix all the other parameters to the known MultiDark values. The covariance matrix has been divided by the number of mocks for each field in order to evidence systematic biases in our analysis. In all cases we were able to recover the true value Ω_{M} of the MultiDark simulation within statistical errors.

The power spectrum analysis has been applied to four subsamples of the current VIPERS data set, which has been divided into four independent volumes, two redshift bins $0.6 < \mathrm{z}_1 < 0.9$ and $0.9 < \mathrm{z}_2 < 1.1$ for each field W1 and W4. The four measured power spectra are very similar since the different effect of each window function is negligible compared to statistical fluctuations introduce by cosmic variance. Even the normalisation is of the same order between the power spectra at each redshift bin due to a compensation between the growth of perturbations (lower at z_2) and the bias factor (higher at z_2).

With a chi-square technique, we decide to follow the literature constraining the cosmological parameters that mainly affect the power spectrum shape, as the matter density Ω_{M} and the baryonic fraction f_{B}. We impose flat priors on the baryonic fraction, the matter density and the bias, Gaussian prior to the pairwise dispersion velocity of galaxies, while fixing the other cosmological parameters to Planck values (Planck Collaboration et al., 2013). In order to estimate the marginalised (over the bias) posterior likelihood distribution, we run an MCMC on the combined W1 and W4 measurements (accounting for the different window function), but treating separately the two redshift bins, given the different bias parameters. Applying a chi-square statistics, with the covariance matrix measured using a combination of 200 Pinocchio mock catalogues for each field and 26 (31) MultiDark mocks for W1 (W4), we compare the four measured power spectra with theoretical models generated with the CAMB algorithm and corrected for redshift-space distortions and window function effects. Fixing the cosmology to the Planck best fit except for the matter density and the baryonic fraction and varying the bias factor, assumed linear and scale independent, we obtain consistent results between the two redshift bins for the chi-square contours in the $\Omega_{\mathrm{M}} - \mathrm{f}_{\mathrm{B}}$ plane once we marginalise over the bias factor. As expected, we find a slightly higher value for the bias in the higher redshift samples once the growth of structures is fixed by the cosmology. We observe a well-known degeneracy (2dFGRS, WiggleZ) between the matter density and the baryonic fraction with the same probability for a high Ω_{M} and f_{B} Universe as for low Ω_{M} and f_{B} values. Combining with a joint likelihood the two redshift bins, we are partially able to break the degeneracy and drastically reduce statistical errors. Our best fit is consistent with the Planck results, even assuming an Hubble constant of $\mathrm{h} = 0.72$, different from the Planck best fit of $\mathrm{h} = 0.67$. We checked also how much our estimate depends on the choice of the maximum k-mode used for the fit. We found that the

particular choice of the maximum mode slightly changes the chi-square contours in the $\Omega_{\rm M} - {\rm f_B}$ plane in the lower redshift bin because more affected by non-linearities, while in the higher redshift bin the chi-squares, function of $\rm k_{max}$, are more consistent. Once we combine the two redshift bins, we observe that varying the maximum mode fitted does not change significantly the best fit.
Moreover we estimate the matter density alone fixing the other cosmological parameters to the best available known Gaussian prior: the spectral index and the scalar amplitude to Planck, the baryons density to the Big Bang Nucleosynthesis and the Hubble parameter to the best Hubble Space Telescope prior. The $\Omega_{\rm M}$ likelihood obtained marginalising over the set of parameters and combining $\rm z_1$ and $\rm z_2$ gives a best fit with 1σ errors of $\Omega_{\rm M} = 0.272^{+0.027}_{-0.031}$ in perfect agreement with the value measured from the VIPERS clustering ratio, $\Omega_{\rm M} = 0.270^{+0.029}_{-0.025}$, and with the Planck result once we rescale the Hubble factor of HST to the Planck best fit.

8.1 Future works

In the near future the VIPERS catalogue will be complete with the full data sample available for a new measurement of the galaxy power spectrum. The increase of the volume will lead to a lower cosmic variance reducing statistical errors in the estimation of the matter density and the baryonic fraction and a smaller window function effect. The increase of the number of galaxies observed will also limit the noise in the observed redshift distribution and reduce the shot-noise term which, for the moment, is not completely negligible for the higher redshift bin.
New mock catalogues (~ 400 for each field) generated by Sylvain De la Torre will also be available very soon. This will allow us to estimate an unbiased covariance matrix without using two different sets of mock catalogues, based on the Pinocchio code and on the MultiDark simulation, to estimate first the correlation matrix and, in a second moment, rescale it to the power spectrum variance measured with the MultiDark catalogues.

Future possibilities implying the use of the galaxy power spectrum could include the analysis of clustering for different class of galaxies. The power spectrum estimator could, in fact, be applied to galaxies with different magnitude cut and different colours in order to study, fixing the cosmology to break the degeneracy between b and σ_8, the bias associated to each class of objects (a similar work, in configuration space, has been done in Marulli et al. (2013) with the correlation function approach).

A parallel work to de la Torre et al. (2013) could be carried out exploring the anisotropy of the clustering due to RSD in Fourier space. A measurement of the power spectrum multipoles will allow us to set a constraint on the growth rate of structure at a mean redshift of $\rm z \sim 0.8$, never reached by other surveys, and confirm the main result, for which VIPERS has been designed, found by de la Torre et al. (2013).

Since it has been shown that neutrinos are massive (Lesgourgues and Pastor, 2012) and given the improvement of simulation with the inclusion of massive neutrinos, an accurate study of the shape of the VIPERS measured power spectrum could also give some additional constrains on their mass.

Appendix A

Convolution

The observed density contrast $\delta_{\rm V}(\mathbf{x})$ at the position $\mathbf{x}$ in a survey with volume V is equal to

$$\delta_{\rm V}(\mathbf{x}) \,=\, \delta_{\rm T}(\mathbf{x}) \times {\rm w}(\mathbf{x}) \;, \tag{A.1}$$

where ${\rm w}(\mathbf{x})$ is the window function of the survey meaning that the true density contrast $\delta_{\rm T}(\mathbf{x})$ could be measured inside the volume of the survey where ${\rm w}(\mathbf{x}) = 1$ and not outside where ${\rm w}(\mathbf{x}) = 0$. This simple multiplication in configuration space becomes a convolution in Fourier space. The power spectrum measured in a volume V is

$$\begin{aligned}\langle \delta_{\rm V}(\mathbf{k}_1)\delta^*_{\rm V}(\mathbf{k}_2)\rangle = & \int\!\!\int {\rm w}(\mathbf{x}_1)\,{\rm w}^*(\mathbf{x}_2)\,{\rm e}^{{\rm i}\mathbf{k}_1\cdot\mathbf{x}_1}\,{\rm e}^{-{\rm i}\mathbf{k}_2\cdot\mathbf{x}_2}\,\langle\delta_{\rm T}(\mathbf{x}_1)\,\delta^*_{\rm T}(\mathbf{x}_2)\rangle\,{\rm d}^3{\rm x}_1\,{\rm d}^3{\rm x}_2 \xrightarrow{\mathbf{x}_2\equiv\mathbf{x}_1+\mathbf{r}} \\ & \int\!\!\int {\rm w}(\mathbf{x}_1){\rm w}^*(\mathbf{x}_1+\mathbf{r}){\rm e}^{{\rm i}(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{x}_1}{\rm e}^{-{\rm i}\mathbf{k}_2\cdot\mathbf{r}}\langle\delta_{\rm T}(\mathbf{x}_1)\delta^*_{\rm T}(\mathbf{x}_1+\mathbf{r})\rangle\,{\rm d}^3{\rm x}_1\,{\rm d}^3{\rm r}\;,\end{aligned} \tag{A.2}$$

where we substitute the variable $\mathbf{x}_2$ with $\mathbf{x}_1 + \mathbf{r}$.

Given the homogeneity of the Universe the term $\langle\delta_{\rm T}(\mathbf{x}_1)\delta^*_{\rm T}(\mathbf{x}_1+\mathbf{r})\rangle$ corresponds to the definition of the correlation function $\xi(\mathbf{r})$, which is the Fourier Transform of the power spectrum, $\xi(\mathbf{r}) \equiv \int {\rm P}(\mathbf{k}'){\rm e}^{-{\rm i}\mathbf{k}'\cdot\mathbf{r}}\,\frac{{\rm d}^3{\rm k}'}{(2\pi)^3}$.

Fourier transforming the window function as ${\rm w}^*(\mathbf{x}_1+\mathbf{r}) = \int {\rm w}^*(\mathbf{k}''){\rm e}^{{\rm i}\mathbf{k}''\cdot(\mathbf{x}_1+\mathbf{r})}\frac{{\rm d}^3{\rm k}''}{(2\pi)^3}$, equation A.2 becomes

$$\begin{aligned}&\int\!\!\int\!\!\int\!\!\int {\rm w}(\mathbf{x}_1)\,{\rm e}^{{\rm i}(\mathbf{k}_1-\mathbf{k}_2+\mathbf{k}'')\cdot\mathbf{x}_1}\,{\rm e}^{-{\rm i}(\mathbf{k}_2-\mathbf{k}'-\mathbf{k}'')\cdot\mathbf{r}}\,{\rm w}^*(\mathbf{k}'')\,{\rm P}(\mathbf{k}')\,{\rm d}^3{\rm x}_1{\rm d}^3{\rm k}'\,{\rm d}^3{\rm k}''\,{\rm d}^3{\rm r} = \\ &\qquad\int\!\!\int {\rm w}(\mathbf{x}_1)\,{\rm e}^{{\rm i}(\mathbf{k}_1+\mathbf{k}')\cdot\mathbf{x}_1}\,{\rm P}(\mathbf{k}')\,{\rm w}^*(\mathbf{k}'+\mathbf{k}_2)\,{\rm d}^3{\rm x}_1\,{\rm d}^3{\rm k}'\;.\end{aligned} \tag{A.3}$$

where the last equality is due to the fact that $\int {\rm e}^{{\rm i}\mathbf{k}\cdot\mathbf{x}}{\rm d}^3{\rm x} \,=\, \delta^{\rm D}(\mathbf{k})\,(2\pi)^3$.

Fourier transforming again the window function as ${\rm w}(\mathbf{x}_1) = \int {\rm w}(\mathbf{k}''){\rm e}^{{\rm i}\mathbf{k}''\cdot\mathbf{x}}\frac{{\rm d}^3{\rm k}''}{(2\pi)^3}$, we finally obtain

$$\langle\delta_{\rm V}(\mathbf{k}_1)\delta^*_{\rm V}(\mathbf{k}_2)\rangle = \int {\rm W}(\mathbf{k}_1+\mathbf{k}'){\rm W}^*(\mathbf{k}_2+\mathbf{k}'){\rm P}(\mathbf{k}')\frac{{\rm d}^3{\rm k}'}{(2\pi)^3}\;. \tag{A.4}$$

Since in Fourier space modes associated to the density contrast are uncorrelated, equation A.4 becomes

$$\mathrm{P_V}(\mathbf{k}) = \int |\mathrm{W}(\mathbf{k}+\mathbf{k}')|^2 \mathrm{P}(\mathbf{k}') \frac{\mathrm{d}^3\mathrm{k}'}{(2\pi)^3} \ . \tag{A.5}$$

The observed power spectrum $\mathrm{P_V}(\mathbf{k})$ is therefore a convolution between the true one $\mathrm{P}(\mathbf{k})$ with the Fourier Transform of the window function $\mathrm{w}(\mathbf{x})$.

Appendix B

Γ-simulation

To construct a distribution of particles in a cubical box with a well known power spectrum in order to evaluate how good our estimator is, we want to match the one-point distribution probability of δ fluctuations, which is not Gaussian.

However, generating a Gaussian field with a known power spectrum is a simple task. We divide the whole process in two steps.

In the first step, we generate a stochastic field Φ_{k} onto a grid in Fourier space. Assuming the density field $\delta(\mathbf{x})$ to be Gaussian, the probability distribution of $\Phi_{\mathrm{k}}(\mathbf{k})$ is still Gaussian. We generate a general field at each point $\mathbf{k}$ of the grid, $\Phi_{\mathrm{k}} = \alpha_{\mathrm{k}} + \mathrm{i}\beta_{\mathrm{k}}$ with α_{k} and β_{k} two random variables Gaussian distributed with mean and variance respectively $\langle\alpha\rangle = \langle\beta\rangle = 0$ and $\sigma^2_\alpha = \sigma^2_\beta = \mathrm{P}(\mathbf{k})$. As a result, the stochastic field Φ_{k} is Gaussian with mean equal to zero and variance equal to the power spectrum which depends on the fiducial cosmology we assume for our simulations. Then we Fourier Transform back Φ_{k} to recover the Gaussian density field Φ in configuration space with variance given by $\sigma^2_\Phi = \int \mathrm{P}_\Phi(\mathbf{k})\mathrm{d}^3\mathrm{k}$.

In the second step, we locally transform the field Φ to match the one-point distribution of VIPERS galaxies. We define another field $\hat{\delta}$ related to Φ with a local transformation $\hat{\delta} = \mathrm{F}(\Phi)$. The function F should satisfies this condition

$$\mathrm{P}(\hat{\delta})\,\mathrm{d}\hat{\delta} = \mathrm{G}(\Phi)\,\mathrm{d}\Phi\ . \tag{B.1}$$

It is therefore defined as

$$\int_{-1}^{+\hat{\delta}} \mathrm{P}(\hat{\delta}')\,\mathrm{d}\hat{\delta}' = \int_{-\infty}^{+\Phi} \mathrm{G}(\Phi')\,\mathrm{d}\Phi'\ , \tag{B.2}$$

where $\mathrm{P}(\hat{\delta})$ and $\mathrm{G}(\Phi)$ are respectively the one-point probability distribution for the field $\hat{\delta}$ and the Gaussian distribution for Φ. With the expression B.2, while Φ is defined between $[-\infty; +\infty]$, the values for the stochastic field $\hat{\delta}$ ranges between $[-1; +\infty]$.

Usually simple simulations are generated with log-normal transform meaning that the relation between the two field is $\hat{\delta} = \mathrm{e}^\Phi - 1$ or inversely, to be consistent with the name, $\Phi = \ln(1+\hat{\delta})$. In our Γ-distribution, we simply assume the probability distribution of the new field $\hat{\delta}$ to be a Γ-distribution. This choice is motivated by the fact that in VIPERS the probability distribution of galaxies is well fitted by a Γ-function (Bel et al., in preparation). Since we apply a local transform to the initial field Φ to obtain the field $\hat{\delta}$, the covariance in Fourier space associated to the two fields is slightly different meaning that the power spectrum associated to $\hat{\delta}$ is no more equal to the input power spectrum we wanted to apply to our simulation. Therefore we should associate the true power spectrum to the new field $\hat{\delta}$. To account for this, we write down the covariance in configuration space for $\hat{\delta}$, which is the correlation function

$$\langle \hat{\delta}_1 \hat{\delta}_2 \rangle = \int \hat{\delta}_1 \hat{\delta}_2 \; \mathrm{P}(\hat{\delta}_1; \hat{\delta}_2) \; \mathrm{d}\hat{\delta}_1 \mathrm{d}\hat{\delta}_2 \; , \tag{B.3}$$

where we know the probability distribution $\mathrm{P}(\hat{\delta})$ but we do not know the joint probability $\mathrm{P}(\hat{\delta}_1; \hat{\delta}_2)$. To overcome this last part, we substitute in equation B.3 the joint probability $\mathrm{P}(\hat{\delta}_1; \hat{\delta}_2) \; \mathrm{d}\hat{\delta}_1 \mathrm{d}\hat{\delta}_2$ with $\mathrm{G}(\Phi_1; \Phi_2) \; \mathrm{d}\Phi_1 \mathrm{d}\Phi_2$ which is a bivariate Gaussian. Equation B.3 becomes

$$\langle \hat{\delta}_1 \hat{\delta}_2 \rangle = \int \mathrm{F}(\Phi_1)\mathrm{F}(\Phi_2)\mathrm{G}(\Phi_1; \Phi_2) \; \mathrm{d}\Phi_1 \mathrm{d}\Phi_2 \; , \tag{B.4}$$

where $\mathrm{G}(\Phi_1; \Phi_2)$ depends on the mean $\langle \Phi \rangle$ which is equal to 0 and the covariance

$$\mathrm{C} = \begin{pmatrix} \sigma_1^2 & \xi_{1,2} \\ \xi_{2,1} & \sigma_2^2 \; . \end{pmatrix} \tag{B.5}$$

For the assumption of homogeneity, the variance of the field Φ is the same at each point of the Universe $\sigma_1^2 = \sigma_2^2$ and for simplicity we choose $\sigma_\Phi^2 = 1$. Moreover the non-diagonal terms of the covariance are the correlation function which again are equals $\xi_{1,2} = \xi_{2,1}$. Equation B.4 gives us a simple relation between the correlation function associated to the field $\hat{\delta}$ and to the field Φ or in Fourier space between the two power spectra. For instance, in log-normal simulation the expression B.3 becomes simply $\xi_{\hat{\delta}} = \mathrm{e}^{\xi_\Phi} - 1$ while in Γ-simulation the relation is evaluated numerically.

The idea therefore is to assume a fiducial cosmology with an input power spectrum, Fourier Transform the power to obtain the correlation function $\xi_{\hat{\delta}}$ and with relation B.4 calculate the correlation associated to the field Φ. The next step will be to Fourier Transform ξ_Φ to obtain the unphysical power spectrum $\mathrm{P}_\Phi(\mathrm{k})$ and generate the initial Gaussian field Φ_k onto a k-space grid with variance $\mathrm{P}_\Phi(\mathbf{k})$. Then Fourier back Φ_k in configuration space to deal with the unphysical density field Φ and transform the obtained scalar field with F in order to have the field $\hat{\delta} = \mathrm{F}(\Phi)$. The field $\hat{\delta}$, defined onto a grid and with probability distribution given by a Γ-distribution, is limited in a range $[-1; +\infty]$ as the true density field δ and its covariance in Fourier space is equal to the true power spectrum of the cosmology we adopt.

Fixing a required mean density of particles $\bar{\mathrm{N}}$, we can transform the density contrast at each position of the grid into an expected number of galaxies $\Lambda = \bar{\mathrm{N}}(1 + \hat{\delta})$. However, this quantity is not an integer yet. We assign to each cell a number N of particles extracted from a poisson distribution with expectation value Λ at the same position. We populate randomly each cell with the number of particles N. Due to this last step,

we do not expect clustering on scales smaller than the size of the cell.
In our work using this technique, we simulates 30 boxes with size $\mathrm{L} = 486.4\,\mathrm{h}^{-1}\mathrm{Mpc}$ and cell size $\mathrm{H} = 0.95\,\mathrm{h}^{-1}\mathrm{Mpc}$. The mean number of particles is fixed to the highest mean density in the redshift distribution of VIPERS mock catalogues. The chosen input cosmology is the MultiDark one. Finally, to obtain VIPERS-like distribution of particles, we cut the 30 boxes in order to have a cone-like W1 geometry.

Bibliography

Alcock C., Paczynski B., 1979, Nature, 281, 358

Anderson L., et al., 2012, MNRAS, 427, 3435

Bardeen J.M., Bond J.R., Kaiser N., Szalay A.S., 1986, ApJ, 304, 15

Bartelmann M., Schneider P., 2001, Phys. Rept., 340, 291

Bel J., et al., 2014, A&A, 563, A37

Bernardeau F., Colombi S., Gaztañaga E., Scoccimarro R., 2002, Phys. Rept., 367, 1

Blake C., et al., 2008, Astronomy and Geophysics, 49, 19

Blake C., et al., 2010, MNRAS, 406, 803

Blake C., et al., 2011a, MNRAS, 415, 2892

Blake C., et al., 2011b, MNRAS, 415, 2876

Bottini D., et al., 2005, PASP, 117, 996

Carlson J., White M., 2010, ApJS, 190, 311

Coil A.L., et al., 2008, ApJ, 672, 153

Cole S., 2011, MNRAS, 416, 739

Cole S., et al., 2005, MNRAS, 362, 505

Colless M., et al., 2003, ArXiv Astrophysics e-prints

Cuillandre J.C.J., et al., 2012, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 8448 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*

Davidzon I., et al., 2013, A&A, 558, A23

Dawson K.S., et al., 2013, AJ, 145, 10

de la Torre S., et al., 2013, A&A, 557, A54

Di Porto C., et al., 2014, ArXiv e-prints

Dodelson S., Schneider M.D., 2013, Phys.Rev.D, 88, 063537

Doran M., Müller C.M., 2004, Journal of Cosmology and Astroparticle Physics, 9, 003

Eisenstein D.J., Hu W., 1998, ApJ, 496, 605

Eisenstein D.J., et al., 2005, ApJ, 633, 560

Feldman H.A., Kaiser N., Peacock J.A., 1994, ApJ, 426, 23

Frigo M., Johnson S.G., 2012, FFTW: Fastest Fourier Transform in the West, Astrophysics Source Code Library

Fritz A., et al., 2014, A&A, 563, A92

Fry J.N., Gaztanaga E., 1993, ApJ, 413, 447

Garilli B., et al., 2008, A&A, 486, 683

Garilli B., et al., 2014, A&A, 562, A23

Granett B.R., et al., 2012, MNRAS, 421, 251

Guth A.H., 1981, Phys.Rev.D, 23, 347

Guzzo L., et al., 2008, Nature, 451, 541

Guzzo L., et al., 2014, A&A, 566, A108

Hamilton A.J.S., 1993, ApJ, 417, 19

HASTINGS W.K., 1970, Biometrika, 57, 97

Hinshaw G., et al., 2013, ApJS, 208, 19

Jing Y.P., 2005, ApJ, 620, 559

Kaiser N., 1987, MNRAS, 227, 1

Larson D., et al., 2011, ApJS, 192, 16

Le Fèvre O., et al., 2005, A&A, 439, 845

Lesgourgues J., Pastor S., 2012, ArXiv e-prints

Lewis A., Challinor A., 2011, CAMB: Code for Anisotropies in the Microwave Background, Astrophysics Source Code Library

Lilly S.J., et al., 2009, ApJS, 184, 218

Marulli F., et al., 2013, A&A, 557, A17

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., 1953, , 21, 1087

Monaco P., Theuns T., Taffoni G., 2002, MNRAS, 331, 587

Parkinson D., et al., 2012, Phys.Rev.D, 86, 103518

Peacock J.A., Dodds S.J., 1994, MNRAS, 267, 1020

Percival W.J., et al., 2001, MNRAS, 327, 1297

Percival W.J., et al., 2007, ApJ, 657, 645

Percival W.J., et al., 2014, MNRAS, 439, 2531

Perlmutter S., et al., 1999, ApJ, 517, 565

Pettini M., Zych B.J., Murphy M.T., Lewis A., Steidel C.C., 2008, MNRAS, 391, 1499

Planck Collaboration, et al., 2013, ArXiv e-prints

Pollo A., et al., 2005, A&A, 439, 887

Prada F., Klypin A.A., Cuesta A.J., Betancort-Rijo J.E., Primack J., 2012, MNRAS, 423, 3018

Riess A.G., et al., 1998, AJ, 116, 1009

Riess A.G., et al., 2011, ApJ, 730, 119

Rubin V.C., Ford Jr. W.K., 1970, ApJ, 159, 379

Sato T., Hütsi G., Yamamoto K., 2011, Progress of Theoretical Physics, 125, 187

Smith R.E., et al., 2003, MNRAS, 341, 1311

Steigerwald H., Bel J., Marinoni C., 2014, Journal of Cosmology and Astroparticle Physics, 5, 042

Strauss M.A., et al., 2002, AJ, 124, 1810

Taylor A., Joachimi B., Kitching T., 2013, MNRAS, 432, 1928

Zwicky F., 1937, ApJ, 86, 217