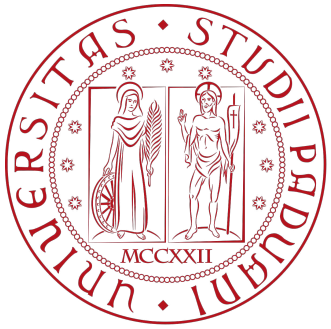


UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master degree in Astrophysics and Cosmology

Final dissertation

ANALYSIS OF SIMULATED GALAXY CATALOGUES AT VERY LARGE SCALES WITH RELATIVISTIC EFFECTS

Supervisor:
Dr. Daniele Bertacca
Co-supervisors:
Prof. Cristiano Porciani, Mr. Mohamed Yousry Elkhashab

Candidate:
Bartolomeo Bottazzi Baldi
Student ID: **2058926**

Academic Year 2022/2023

Introduction

Redshift space distortions distort the observed distribution of galaxies. These effects are caused by the inhomogeneous distribution of matter in the Universe that sources the galaxies' peculiar velocity and the underlying gravitational potential field. Both of these phenomena stretch or compress the wavelength of the upcoming observed radiation, and add up to the cosmological redshift sourced by the Hubble flow, changing the measured redshift [1]. As a consequence, since we use a galaxy's redshift to infer its distance by assuming a non-perturbed FLRW universe, the mapping from redshift space to real space distorts the galaxy maps we build, shifting their positions.

These distortions can be observed in the galaxy power spectrum and the higher-order statistics. The first work addressing this topic was done by Kaiser in [2]. In that work he shows that assuming the distant observer approximation, at linear order in density and velocity perturbations, the galaxies' peculiar velocities boost the monopole of the power spectrum. Kaiser's approximation breaks down at the largest scales, where we predict different contributions acting on the galaxy distribution [3,4]. This happens for multiple reasons.

First, the distant observer approximation is only valid in the regime in which the subtended angle from the observer to the region of the sky considered is small, which is not true when considering the largest scales. We refer to these large scale velocity effects as *wide angle* and *Doppler effects* [3], and they provide additional distortions to the galaxy distribution and source magnification effects onto the galaxies, altering the counts near the flux cut of a flux-limited sample [5].

Second, at those scales general relativistic effects become non negligible, and they generate further redshift distortions, magnification effects, and displacements in the angular positions of the galaxies. These effects depend also on integrals of the large scale potential along the line of sight, and we predict that their contribution will increase when studying large scales structures in the deep universe ($z > 1$) [6–11].

General relativity has withstood numerous observational tests on smaller scales, but its implications on cosmological scales remain an area of ongoing research and debate [12], thus to observe its effects in the field of redshift space distortions would be an important test of Einstein's theory of gravity in a cosmological framework.

Regarding the detection of GR effects, future redshift surveys will probe galaxy clustering in the distant Universe on unprecedentedly large scales. For example the Euclid mission [13], will measure in the optical and infrared the redshift of H-alpha emitters in a range of $0.9 < z < 1.9$, over a $15\,000\,\mathrm{deg}^2$ portion of the sky. DESI (Dark Energy Spectroscopic Instrument) [14] will measure galaxy redshifts from different

emission lines, in a range of $0.1 < z < 3.5$. SKAO (Square Kilometre Array) [15] in its final configuration, will probe radio emissions of galaxies up to $z = 3$.

It is still unclear whether GR effects will be observable in these next-generation surveys, and so it is useful to develop analysis tools that will allow to fully exploit the potential of the upcoming data. This translates in the necessity of creating mock galaxy catalogues of what a deep-Universe survey will observe.

With this purpose, the LIGER code (Light Cones using GEneral Relativity) was developed [16,17]. It is a post-processing tool that allows to imprint *a posteriori* general relativistic effects on a Newtonian N-body simulation, producing a light-cone that accounts for redshift space distortions corrections and magnification effects at first order in the perturbations. This is a valid approach to follow because at first order in the perturbation, the mathematical description of a pressureless fluid can be formulated so that its general relativistic approximation is in agreement with the Newtonian counterpart [18].

The corrections that LIGER implements are obtained by studying the linear perturbations of the photon geodesics that reach us from distant sources [8, 19, 20], which include both local and integrated contributions along the line of sight.

The code has been previously used to forecast the detectability of large-scales effects on deep surveys, for redshift bins $z > 1$. Since the transverse size of a lightcone rapidly increases with redshift, a very large simulation box (of the order of $10\,\mathrm{Gpc}/h$) is required to cover wide opening angles. However, running simulations with sufficient spatial and mass resolution to follow galaxy formation would be computationally prohibitive. For this reason, up until now LIGER has been used onto large, low-resolution dark matter only simulations, shifting the matter particles and creating *at posteriori* the galaxy distribution. This was done by implementing a biasing relation which involved the assumption of a set of survey functions (e.g. magnification and evolution biases [21]) that describe the large-scale tracer considered.

While this method allows for a fast production of many galaxy catalogues, there are some limits in this approach. First, the biasing procedure builds the tracer catalog on a cubic grid, losing any information regarding clustering on scales smaller than the cell size. This biasing procedure forces us to increase the cell size as the linear bias grows, decreasing the number of cells per side of the grid and reducing the resolution. This limit can be a issue when computing Fourier-based statistics of highly biased tracers, as aliasing effects might arise at larger scales due to a low number of cells per side. Second, this procedure makes it difficult to model selection criteria such as shape selection or flux selections that are more complex than a flux cut. Last, Poisson shot-noise is artificially implemented by sampling the number of tracers in each cell of the produced grid, and this procedure gives rise to spurious effects as we probe smaller scales.

In my master thesis, we update the LIGER code to build lightcones of tracers by mapping them onto redshift space directly, without recurring to the biasing procedure. This approach solves the issues mentioned above, allowing to produce a more realistic galaxy catalogue, which does not suffer from the limitations mentioned above.

This thesis is organized as follows: in the first chapter we review the theory

of structure formation and of redshift space distortions, considering also relativistic effects. In the second chapter we describe LIGER's functioning and the details related to our new implementation. We will also give a quick overview of the FKP estimator for the galaxy power spectrum, which we will use in our tests of the code. In the third and fourth chapters we will show the results of our tests done to validate the new code and our conclusions.

Contents

List of Figures

Chapter 1

Theory

The Standard Model of Cosmology is the currently accepted cosmological model that explains the structure, components, and evolution of the Universe It describes a universe constituted of three main components: radiation, matter and dark energy.

We refer as "radiation" to all the particles that approach relativistic speeds, i.e. for which special relativistic effects are not negligible, such as photons or neutrinos in the early phases of the Universe [31].

The matter component is divided in two categories: baryonic matter and dark matter. Baryonic matter can interact with the radiation field through different processes, and ultimately constitutes the "light-interacting" part of galaxies, such as stars or dust. Dark matter, that is yet to be directly observed [32], interacts weakly or not at all with the radiation field, and was introduced independently in the description of different astrophysical and cosmological processes, to match observations with the currently accepted theory of gravity [33].

Lastly, dark energy was theorized to explain the late accelerating expansion of the Universe, which is not compatible with the behaviour of any of the previously listed components. For this reason the Standard Model of Cosmology is also called the ΛCDM model ([34, 35]), where Λ is the *cosmological constant* and is related to the density of this dark energy component.

The ΛCDM model describes a universe that in its earliest stages was subjected to a rapid expansion called *Inflation*, in which any region of space would grow in size by many orders of magnitudes during a small period of time [36]. This expansion is driven by the *inflaton* field. This quantum field after inflation starts to decay, giving ultimately origin to the particles that make up our current Universe [37], in the so-called *reheating* phase. During inflation, the quantum fluctuations of the field are stretched on scales comparable with the regime of classical physics, thus going from a quantum stage to a classical one [38]. These perturbations leave an imprint on the matter and radiation contents of the Universe, generating density perturbations.

After the reheating phase, the Universe is dominated by radiation, which suppresses the growth of perturbations due to its pressure [39], until the matter component starts to dominate. We refer to the epoch in which matter starts dominating over radiation as matter-radiation equality. After the Universe becomes matter dominated, the radiation pressure becomes negligible, and the density perturbations in

the matter component can start to grow under the influence of gravity, generating the LSS of the Universe.

In the most overdense regions of the Universe, dark matter haloes can form: these are gravitationally bound regions of the universe which overcame the cosmological expansion and collapsed into a virialized structures [40]. A dark matter halo can be host of substructures, called sub-haloes, which consist in smaller haloes that orbit the main halo's gravitational well. Baryonic matter clusters at the center of each halo or sub-halo, following the gravitational well that they generate. In time, these structures grow, attracting matter from the surrounding underdense regions [41]. Under certain conditions, the radiative processes that allow for the formation of stars can trigger, and galaxies form at the center of haloes and sub-haloes, forming a **galaxy cluster**.

The result of this description is a universe populated by dark matter haloes, which host galaxies that *cluster* together due to their mutual gravitational interactions [42].

In this framework, we see how studying the clustering of matter in the large scales of the Universe could give us insights about the primordial fluctuations of the inflaton field.

There is one main limitation to this approach: the main source of information about the large scale structure of the Universe is the light emitted from the galaxies. However, within our current theories, baryonic matter constitutes a small fraction of the matter content, leaving most of it undetectable. Still, due to its link with the underlying density field, the study of galaxy clustering is one of the most promising ways to infer information about the late Universe matter distribution.

1.1 Homogeneous and isotropic Universe

In order to study the Universe as a whole, we need to make some assumptions about its largest scales structure: we introduce the **cosmological principle**, which states that the Universe is homogeneous and isotropic around any point. This is does not hold at any scale: for example, the Local Group of galaxies spans $\approx 3\,\mathrm{Mpc}$ in diameter, while the typical size of a supercluster is $\sim 100\,\mathrm{Mpc}$, so at least for scales smaller than $100\,\mathrm{Mpc}$ we know that the Universe presents inhomogeneities and anysotropies. Since surveys of the Universe haven't found structures larger than superclusters, we can say that the cosmological principle holds for scales larger than $100\,\mathrm{Mpc}$ (see e.g. [43]).

This principle admits an expanding Universe, which is consistent with our current observations [22, 44]. In order to account for this expansion we define **fundamental observers**, who are comoving with the cosmic fluid. In their frame of reference we use the **comoving position x**, which is related to the physical position **r** through [30]

$$\mathbf{r}(t) = a(t)\,\mathbf{x}\,, \tag{1.1}$$

where $a(t)$ is the scale factor, that describes how the Universe expands in time. We set $a_0 \equiv a(t_0) = 1$, where we define t_0 as our instant of time in the Universe's evolution, such that $\mathbf{r}(t_0) = \mathbf{x}$.

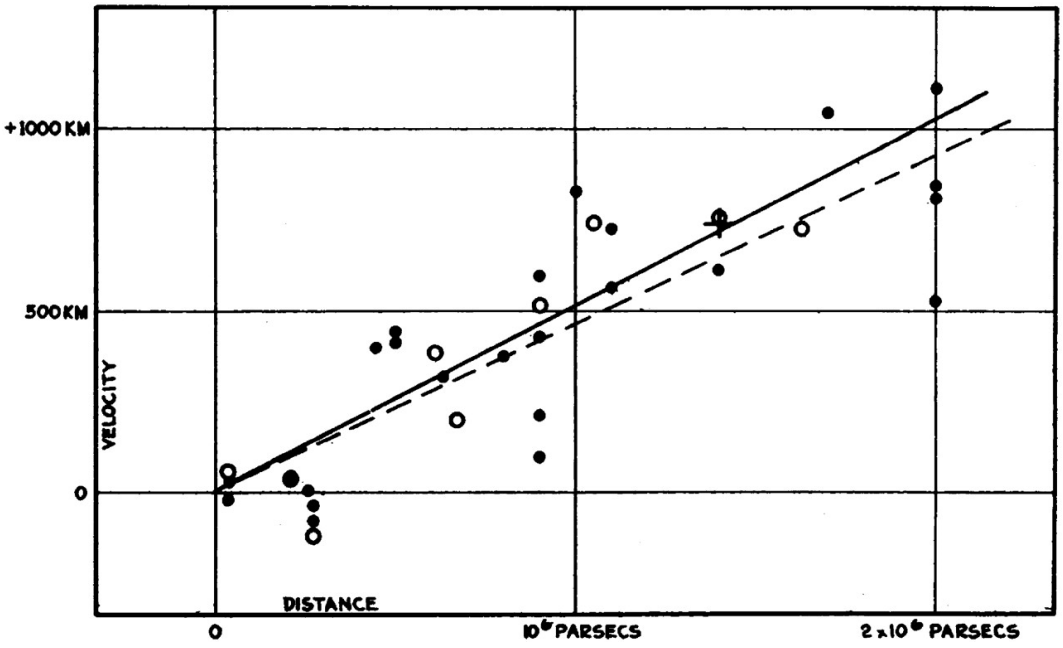


Figure 1.1: Velocity-Distance Relation among Extra-Galactic Nebulae. Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually. Reference for the data: [22].

If we consider a steady point in comoving space ($\mathbf{x}$ = constant), and take the derivative of equation (1.1) we retrieve its physical velocity

$$\mathbf{v}_{\text{phys}} \equiv \dot{\mathbf{r}} = \dot{a}\,\mathbf{x} = H(t)\,\mathbf{r}\,, \tag{1.2}$$

where we defined the **Hubble parameter** $H(t) \equiv (\dot{a}/a)$. This equation relates the position of a steady object located at $\mathbf{x}$ to its velocity sourced by the Universe expansion. We refer to the motion due to the cosmological expansion as the **Hubble flow**. The value of H at $t = t_0$ is known as **Hubble's constant** H_0, from which we define **Hubble's law** [22]

$$v_{\text{phys}} = H_0\, r\,, \tag{1.3}$$

where r is the physical distance of a steady object from us and v_{phys} is its velocity. This equation is an approximation of equation (1.2), valid only in the local Universe, where we observe objects at a time $t \approx t_0$. In figure (1.1) we show Hubble's diagram, which prompted the discovery of an expanding Universe.

In the context of general relativity, we can describe the most general metric that satisfies the cosmological principle, the Friedmann-Lemaître-Robertson-Walker

(FLRW) metric, which allows for an expanding, curved spacetime, with a line element

$$ds^2 = -c^2\, a(t)^2\, dt^2 + a(t)^2 \left\{ d\chi^2 + f_k(\chi) \left[d\theta^2 + \sin^2(\theta)\, d\phi^2 \right] \right\}, \tag{1.4}$$

where c is the speed of light, (χ, θ, ϕ) are the spherical coordinates in the comoving frame of reference, and k is the curvature parameter, which states if the Universe is cosmologically open ($k > 0$), flat ($k = 0$) or closed ($k < 0$) [45]. The function f_k gets a different form depending on the curvature:

$$f_k^2(\chi) = \begin{cases} \frac{1}{\sqrt{k}} \sin\left(\chi \sqrt{k}\right) & (k > 0) \\ \chi & (k = 0) \\ \frac{1}{\sqrt{-k}} \sinh\left(\chi \sqrt{-k}\right) & (k < 0) \end{cases} . \tag{1.5}$$

This function is closely related to the distance of two objects in the sky, and we will return to it in the following section.

1.1.1 Distances in cosmology

In a FLRW spacetime, there are different ways in which we can define the distance of an object, depending on the method of measurement used. In this subsection we will breafly list all the different distance definitions, and we will be referring to [46].

Redshift

Analogously to the propagation of sound waves, the velocity of a source can shift the wavelength of the emitted light. We define the redshift of a source as the relative Doppler shift of its emitted light resulting from radial motion

$$z \equiv \frac{\lambda_0 - \lambda_e}{\lambda_e}, \tag{1.6}$$

where λ_e is the emitted wavelength and λ_0 is the one observed.

If we consider only the radial velocities sourced by the Hubble flow we obtain

$$1 + z = \frac{1}{a}, \tag{1.7}$$

where the scale factor a is evaluated at the time of emission. Due to this relation, it is common practise to refer to epochs of the Universe with their redshift z, evaluated from the scale factor of that epoch. For this reason, the redshift sourced only by the Hubble flow is often referred to as **cosmological redshift**.

Comoving distance (radial)

Since light travels only null geodesics, i.e. $\mathrm{d}s = 0$, looking at an object at a given distance means observing it as it was in the past. If we consider the radial path of a

photon that starts from the source at a time t and reaches us at a time t_0, knowing that it travels on null geodesics we obtain the **comoving distance**

$$\chi(t) = \int_t^{t_0} \frac{c}{a(t')}\, dt' \,, \tag{1.8}$$

which is the distance, in the comoving frame of reference, from the emitter to us (the observer).

Comoving distance (transverse)

If we consider instead two objects located at the same redshift from us, but separated on the sky, their distance in the comoving frame is given by the **comoving transverse distance**, given by the function $f_k(\chi)$ in equation (1.5). As we can see from its form, this distance differs from the radial one due to the curvature of the Universe: for a flat Universe ($k = 0$), the two forms coincide.

Luminosity distance

If we measure the flux F received from a source and we know its luminosity L, we can infer the **luminosity distance**

$$D_{\rm L}(z) \equiv \frac{L}{4\pi F} \,, \tag{1.9}$$

which scales with the flux in the way we would expect χ to scale in the nearby Universe, where we would have $a \approx a_0 = 1$ and $\chi \to 0$. Due to the curvature and cosmological expansion of the Universe, the comoving and luminosity distances differ, as they are related by

$$D_{\rm L} = \frac{f_k(\chi)}{a} \,, \tag{1.10}$$

which reduces to $D_{\rm L} = \chi/a$ in a flat Universe. Objects with a known luminosity L can be used to measure the distance of their portion of the Universe, and for this reason they are called *standard candles*.

Angular distance

Similarly as in the luminosity case, measuring the angular extension θ of an object in the sky for which we know its intrinsic size R, we can infer the **angular distance**

$$D_{\rm A} = \frac{R}{\theta} \,, \tag{1.11}$$

which scales with the size in the way we would expect χ to scale in the nearby Universe. The comoving and angular distances are related as

$$D_{\rm A} = a\, f_k(\chi) \,, \tag{1.12}$$

which, again, reduces to $D_{\rm A} = a\chi$ in a flat Universe. Objects with known intrinsic size are called *standard rulers*, as they can be used to infer the distance of their portion of the Universe.

1.1.2 Energy content of the Universe

We want to study how the matter and energy content of the Universe interact and shape the metric. To do so, we make use of the Einstein equations (EE), which relate the Einstein tensor $G_{\mu\nu} = R_{\mu\nu} - 1/2 g_{\mu\nu} R$, that describes the spacetime geometry, and the energy-momentum tensor $T_{\mu\nu}$ that describes the Universe energy content:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G \, T_{\mu\nu} \,, \tag{1.13}$$

where we see that $G_{\mu\nu}$ is linked to the Ricci curvature tensor $R_{\mu\nu}$ and scalar R. In order to solve these equations we need to plug in into the left-hand side the FLRW metric and in the right-hand side the energy-momentum tensor, for which we need to make some assumptions. We consider a perfect isotropic fluid, comoving with the Universe [30], for which

$$T^{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \qquad , \tag{1.14}$$

where ρ is the density and p is the pressure, which is isotropic due to the perfect fluid assumption.

From the different components of the equations (1.13), we get the **Friedmann equations**

$$H^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2} \,, \tag{1.15}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3} \left(\rho + 3 \frac{P}{c^2} \right) , \tag{1.16}$$

$$\dot{\rho} = -3H \left(\rho + \frac{P}{c^2} \right) , \tag{1.17}$$

where G is the gravitational constant. These equations are not independent from each other, given two of them, the third one can always be derived.

Since only two of them are independent, in order to fully solve the Friedmann equations we need to specify the equation of state of the cosmological fluid, which in most cases [30], can be written as

$$p = \omega \rho c^2 \,, \tag{1.18}$$

where ω can be constant for simpler fluid models, or vary in time. Solving equation (1.17) with this equation of state we get

$$\rho(a) = \rho(a_0) \left(\frac{a}{a_0} \right)^{-3\,(1+\omega)} \qquad , \tag{1.19}$$

that tells us how the cosmic fluid density changes with the cosmological expansion.

As we said, in the ΛCDM model we consider three main components that make up the Universe. It can be shown that their equation of states are [30]

- $\omega_m = 0 \to \rho_m \propto a^{-3}$ in the case of matter;
- $\omega_\gamma = \frac{1}{3} \to \rho_\gamma \propto a^{-4}$ in the case of radiation;
- $\omega_\Lambda = -1 \to \rho_\Lambda = \text{const.}$ in the case of dark energy.

The total energy of the Universe will be given by

$$\rho = \rho_m + \rho_r + \rho_\Lambda \,, \tag{1.20}$$

and for each component, we define the density parameters

$$\Omega_i = \frac{\rho_{i,0}}{\rho_c} \,, \tag{1.21}$$

where $\rho_c \equiv 3H_0^2/(8\pi G)$ is called critical density, and it is defined as the energy density that would source a flat Universe. We can then write equation (1.15) as

$$H^2 = H_0^2 \left[\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_\Lambda + (1-\Omega_k) a^{-2}\right] , \tag{1.22}$$

where we defined the curvature density parameter as

$$\Omega_k = 1 - \frac{c^2 k}{H_0^2} \,. \tag{1.23}$$

1.2 The formation of large-scales structures

Starting from this section, we will set the speed of light to $c = 1$ and we will measure distances in Mpc/h, where h is an adimensional that accounts for the uncertainties regarding the Hubble's constant measurement. We write Hubble's constant as

$$H_0 = 100\, h\, \mathrm{km\, s^{-1}\, Mpc^{-1}}, \tag{1.24}$$

where given the current measurements, the value of h is subjected to tensions from different experiments, varying in the range $[0.65, 0.75]$ (see e.g. [47]).

Until now we described a smooth universe, in the sense that all of its components are homogeneous perfect fluids with the same density at any point of space, consistent with the Cosmological principle. However, this principle does not hold for scales smaller than $100\,\mathrm{Mpc}\,h^{-1}$, at which the Universe presents structures such as galaxy clusters, filaments and voids [48, 49], as it can be seen in figure (1.2), where the distribution of galaxies observed by the Sloan Digital Sky Survey [23] is shown. To describe the Universe at these scales, we thus have to generalize from a smooth description, allowing the local density of a fluid $\rho(\mathbf{x})$ to deviate from its mean value $\bar{\rho}$. Since we expect that at the largest scales the cosmological principle will still hold, we will describe these density deviations with perturbation theory, perturbing over the smooth background we described in the previous section.

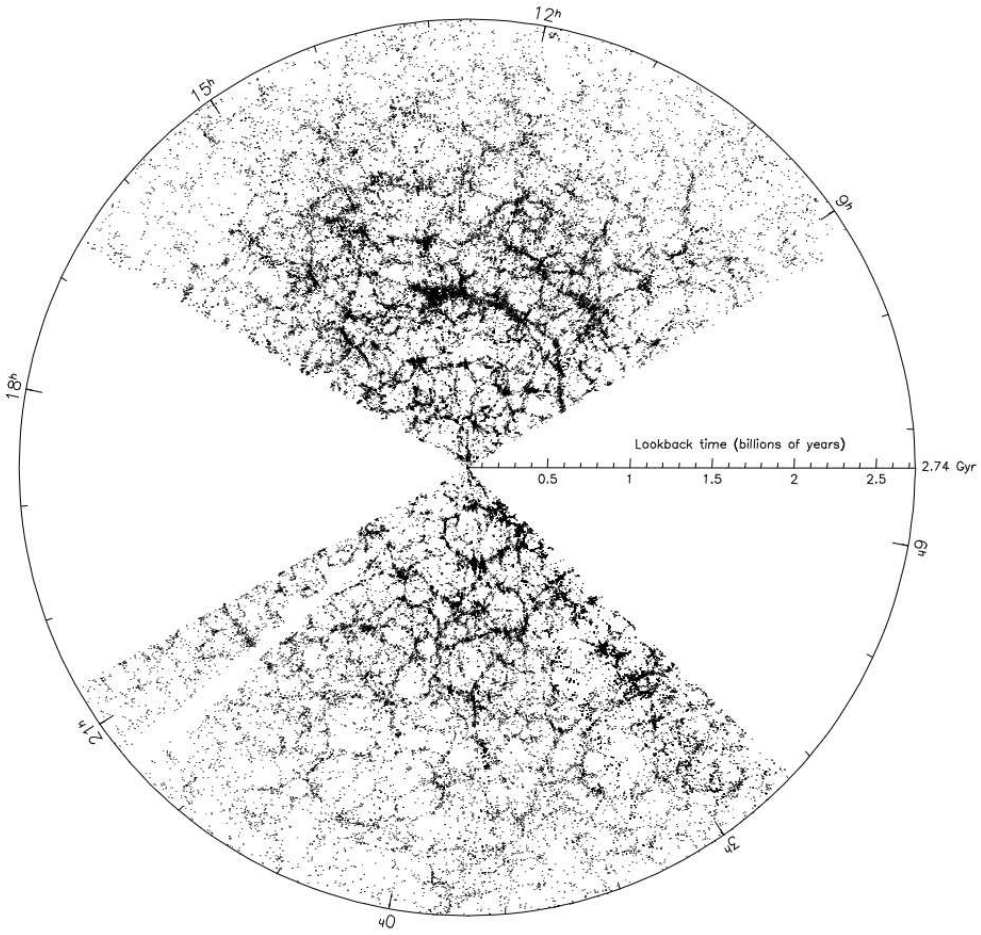


Figure 1.2: Distribution of galaxies according to the Sloan Digital Sky Survey (SDSS). This figure shows galaxies that are within 2° of the equator and closer than 858 Mpc (assuming $H_0 = 71\,\mathrm{km\,s^{-1}Mpc^{-1}}$). Picture taken from [23].

In particular, since in this work we use dark matter only simulations to describe large scales structures formation, we will focus on the behaviour of non-relativistic, collisionless particles under the influence of gravity and of the Hubble expansion. Furthermore, since we are interested in regions of the Universe which will be probed by next-gen galaxy surveys ($z \lesssim 2$) we will consider a matter and Λ dominated Universe.

1.2.1 Linear perturbation theory

In this work we are interested in the clustering of matter on large scales, where the density deviations from the background value are small [30, 48]. More specifically we will study scales where

$$\delta \equiv \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}} \ll 1\,, \tag{1.25}$$

where $\rho(\mathbf{x})$ is the matter density evaluated at the comoving position $\mathbf{x}$, $\bar{\rho}$ is its mean value, and we defined δ as the **density contrast**, or **overdensity**. Since we treat small density fluctuations, we will work with linear perturbation theory.

It can be shown that at large enough scales, where the motion of matter particles can be described by a bulk velocity and we can neglect multi-stream effects, the behaviour of non relativistic, collisionless matter can be modelled after a pressureless fluid [48–50]. This approximation breaks down at smaller scales, where multi-stream regions form and we can not define a unique velocity field (i.e. in virialized regions such as dark matter haloes).

For this section we will work in the Newtonian regime for gravity, and thus we will be limited in considering scales smaller than the **Hubble radius** $d_{\rm H} \equiv 1/H(t)$, since at those scales the gravitational potential would generate non-negligible curvature effects [9, 11, 51].

The equations for a pressureless fluid in a FLRW metric under Newtonian gravity are [30, 48, 50]

$$\frac{\partial \rho}{\partial t} + \nabla_{\mathbf{r}} \cdot (\rho \mathbf{u}) = 0 \qquad \text{Continuity equation,} \tag{1.26}$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla_{\mathbf{r}})\mathbf{u} = -\nabla_{\mathbf{r}}\Phi \qquad \text{Euler equation,} \tag{1.27}$$

$$\nabla^2_{\mathbf{r}}\Phi = 4\pi G\rho - \Lambda \qquad \text{Poisson equation,} \tag{1.28}$$

where ρ is the density of the pressureless fluid, $\mathbf{u}$ is its bulk velocity, and Φ is the gravitational potential, which as we can see from equation (1.28), is both sourced by the fluid density and the cosmological constant.

In order to proceed with perturbation theory we will separate the quantities ρ, $\mathbf{u}$ and Φ from their background value, moving also from proper to comoving

coordinates, writing:

$$\begin{aligned}\rho(\mathbf{r},t) &= \bar{\rho}(t)\,(1+\delta(\mathbf{x},t))\,,\\ \mathbf{u}(\mathbf{r},t) &= \dot{a}\,\mathbf{x}+\mathbf{v}(\mathbf{x},t)\,,\\ \Phi(\mathbf{r},t) &= \bar{\Phi}(t)+\phi(\mathbf{x},t)\,,\end{aligned} \tag{1.29}$$

where, solving for the background values, it can be shown that [48, 49]

$$\bar{\Phi}(t) = \frac{\ddot{a}\dot{a}}{2}|\mathbf{x}|^2\,. \tag{1.30}$$

Assuming small peculiar velocity perturbations $\mathbf{v}$ and a small density contrast δ, it can be shown that the linearized fluid equations are [48, 50]

$$\frac{\partial\delta}{\partial t}+\frac{1}{a}\nabla_{\mathbf{x}}\cdot\mathbf{v}=0\,, \tag{1.31}$$

$$\frac{\partial\mathbf{v}}{\partial t}+H\,\mathbf{v}=-\frac{1}{a}\nabla_{\mathbf{x}}\phi\,, \tag{1.32}$$

$$\nabla^2_{\mathbf{x}}\phi = 4\pi G a^2\bar{\rho}\,\delta\,. \tag{1.33}$$

We can combine this set of equations to retrieve a second order differential equation for δ

$$\frac{\partial^2\delta}{\partial t^2}+2H\frac{\partial\delta}{\partial t}=4\pi G a\bar{\rho}\,\delta\,. \tag{1.34}$$

Since equation (1.34) presents only temporal derivatives, we can can separate time and space dependence in $\delta(\mathbf{x},t)$: it admits a superimposition of two solutions, a growing one $D^{(+)}(t)\,A(\mathbf{x})$ and a decaying one $D^{(-)}(t)\,B(\mathbf{x})$, where $A(\mathbf{x})$ and $B(\mathbf{x})$ are two functions describing the initial overdensity configuration:

$$\delta(\mathbf{x},t) = D^{(+)}(t)\,A(\mathbf{x})+D^{(-)}(t)\,B(\mathbf{x})\,. \tag{1.35}$$

$D^{(+)}(t)$, $D^{(-)}(t)$ are called growing and decaying modes, and in general we have [48, 50]:

$$D^{(+)}(t)\propto H(t)\int_0^t\frac{dt'}{a^2(t')H^2(t')}\,, \tag{1.36}$$

$$D^{(-)}(t)\propto H(t)\,. \tag{1.37}$$

It's not necessarily true that $D^{(-)}$ decays with time, while $D^{(+)}$ grows: for example, as we will see next, in a Λ dominated universe both $D^{(-)}$ and $D^{(+)}$ freeze over time, approaching a constant value [50]. We will still assume that $D^{(+)}$ dominates over

$D^{(-)}$ unless stated otherwise, since it is true for a matter-dominated universe, as we will see next. In conclusion, when talking about the growth of structures, we will refer only to the growing mode, also called **growth factor**.

Using this result and equation (1.33), we can see that the gravitational potential ϕ will evolve in time as

$$\phi \propto \frac{D^{(+)}}{a}\,. \tag{1.38}$$

Lastly, we define the linear growth rate as the logarithmic derivative of the growth factor over the scale factor

$$f \equiv \frac{d\ln D^{(+)}}{d\ln a}, \tag{1.39}$$

which as we will see in the next section, appears when linking the density perturbations to the velocity streams it sources.

Structures growth in an EdS universe

We will first solve equation (1.34) for a simplified situation: a flat, matter dominated universe, also called Einstein-de Sitter universe (Eds). In this universe we have $\Omega_m = 1$, while all the other densities are set to zero. Using this condition we find $a(t) \propto t^{2/3}$, $H = 2/(3t)$ and $\bar{\rho} = \left(6\pi G t^2\right)$. From equations (1.36) and (1.37) we then get:

$$D^{(+)} \propto t^{2/3}, \quad D^{(-)} \propto t^{-1}\,, \tag{1.40}$$

which leads to:

$$\mathbf{v}^{(+)} \propto t^{1/3}, \quad \mathbf{v}^{(-)} \propto t^{-4/3}\,, \tag{1.41}$$

$$\phi^{(+)} \propto t^{0}, \quad \phi^{(-)} \propto t^{-5/3}\,. \tag{1.42}$$

As anticipated, we see that in this case the growing mode dominates over the decaying mode, and also that the potential field stays constant in time.

Structures growth in an Universe with matter and dark energy

Now we will consider a more realistic setting for the late Universe, in which both matter and dark energy are present. Current observations state that the Λ component started dominating over the matter one in the recent Universe, around $z_* \approx 0.55$ [52], so it is useful to check how the growth factor behaves in this case. In [50] it is shown that in this case the solution for $D^{(+)}$ and $D^{(-)}$ is

$$D^{(+)} \propto B_x\left(\frac{5}{6}, \frac{2}{3}\right)\left(\frac{\Omega_m}{\Omega_\Lambda}\right)^{1/3}\left[1 + \frac{\Omega_m}{a^3\Omega_\Lambda}\right], \tag{1.43}$$

$$D^{(-)} \propto a^{-3/2} \left(1 + \frac{\Omega_\Lambda}{\Omega_m} a^3\right)^{3/2}, \tag{1.44}$$

where

$$x = \frac{\Omega_\Lambda}{\Omega_m + a^3 \Omega_\Lambda}, \tag{1.45}$$

and $B_x(\alpha, \beta)$ is the incomplete Beta function. We see that as Ω_m/Ω_Λ becomes smaller the growing mode $D^{(+)}$ slows down, approaching a constant value. In the later times, when $z < z_* = (\Omega_\Lambda/\Omega_m)^{1/3} - 1$, the growth of structures slows down in respect to $a(t)$, due to the expansion of the Universe.

1.2.2 Peculiar velocity field

We will now see how the peculiar velocity field $\mathbf{v}$ is linked to the density perturbation δ, since it will important for the rest of the work. To do so, we start by showing that at large scales, the bulk motion of the pressureless fluid is purely irrotational. If we write $\mathbf{v} = \mathbf{v}_{\rm curl} + \mathbf{v}_{\rm irr}$, the sum of a divergenceless ($\mathbf{v}_{\rm curl}$) and irrotational ($\mathbf{v}_{\rm irr}$) component, from equation (1.32) we can see that:

$$\dot{\mathbf{v}}_{\rm curl} + H\, \mathbf{v}_{\rm curl} = 0\,, \tag{1.46}$$

that implies $\mathbf{v}_{\rm curl} \propto 1/a$: the divergenceless component of the velocity will be smoothed out with the expansion of the universe. We remind that this hold as long as the pressureless fluid model holds: when probing smaller scales this will not valid anymore.

Now that we know that δ evolves in time as $D^{(+)}(t)$ we plug it in equation (1.31), which can now be written as [30, 48, 50]

$$\nabla \cdot \mathbf{v} = -aHf\,\delta\,. \tag{1.47}$$

Since we are working with a purely irrotational velocity field we can write it in terms of a potential field:

$$\mathbf{v} = \nabla \psi_v\,, \tag{1.48}$$

which means

$$\nabla^2 \psi_v = -aHf\,\delta\,. \tag{1.49}$$

Solving for ψ_v and deriving we get:

$$\mathbf{v}(\mathbf{x}) = -aHf\,\nabla\nabla^{-2}\delta = \frac{aHf}{4\pi} \int d^3x'\,\delta(\mathbf{x}', a) \frac{\mathbf{x}' - \mathbf{x}}{|\mathbf{x}' - \mathbf{x}|^3}\,. \tag{1.50}$$

This equation relates the velocity field to the surrounding matter distribution and it can be applied to use a galaxy peculiar velocity (since it is expected to follow the matter flow) to probe the underlying matter density field.

1.3 Spatial statistics of the density field

Following the definition of δ in equation (1.25) we see that by design its average value is expected to be $\bar{\delta} = 0$. To study the spatial properties of the density field it is then useful to recur to 2-point statistics. In this section we will study the auto correlation of the density contrast δ (either of matter or any of its tracers) in a equal-time hyper-surface: the quantity defined in this way does not represent a true observable because what we measure in a survey lies on the lightcone hypersurface instead, but we will see in section (1.4) an analog treatment for that case.

Before starting, we will have to introduce the density contrast for a tracer, such as galaxies or haloes. Calling the comoving number density of a tracer $n_g(\mathbf{x})$, we define its density contrast as

$$\delta_g(\mathbf{x}) = \frac{n_g(\mathbf{x}) - \bar{n}_g}{\bar{n}_g}\,, \tag{1.51}$$

in which $\bar{n}_g$ is its mean, background value.

1.3.1 Discrete fields

Before studying the properties of δ we want to be more detailed about the definition of the tracer density field $n_g(\mathbf{x})$. Looking back at equation (1.51) we see that we treated n_g as a continuous field, in a similar way as the matter one. Tracers are discrete objects, and thus a proper density description would be

$$n_g^d(\mathbf{x}) = \sum_{\mathbf{x}_i} \delta_D(\mathbf{x}_i - \mathbf{x})\,, \tag{1.52}$$

where the sum is over the tracers positions and the label "^{d}" specifies that the density is computed from a discrete set of objects.

The way we treat this issue is by considering the discrete tracer distribution as one realization of the sampling of the underlying continuous field (see e.g. [49]). The most common way to do so is by **Poisson sampling**: suppose to divide the space in cells of volume ΔV, then the expected number of particles in each cell is

$$\mathcal{N}_i = \int_{V_i} dx^3\, n_g(\mathbf{x}), \tag{1.53}$$

where we integrated in the i-th cell's volume. At each cell then, the number of tracers present N_i will follow the distribution

$$\Pr\{N_i = M\} = \frac{\mathcal{N}_i^M}{M!}\, e^{-\mathcal{N}_i}. \tag{1.54}$$

The Poisson sampling we considered is local, in the sense that the probability of finding N tracers in a given cell is independent on the configuration of the field in the surrounding space. In principle this is not a given truth: the probability of finding N particles in a region of space could depend on the configuration of the density field in the surroundings, or also in the sampling of particles in the same of surrounding regions, calling for more complex models.

1.3.2 Correlation function

For simplicity in the rest of this section we will refer to both matter and tracer density perturbations with a general density field $\delta(\mathbf{x})$, unless specified otherwise. We can use the overdensity δ to construct the **correlation function**, defined as its covariance. Considering two comoving positions $\mathbf{x}_1$ and $\mathbf{x}_2$:

$$\xi(x_{12}) = \langle \delta(\mathbf{x}_1)\delta(\mathbf{x}_2)\rangle, \tag{1.55}$$

where $x_{12} \equiv |\mathbf{x}_1 - \mathbf{x}_2|$. The dependence on only the distance between the two points is granted by the assumption that the Universe is isotropic and homogeneous. In the next section we will see that this is not true when dealing with the *observed* density field, which is projected upon the lightcone [53] and subjected to various distortions due to the redshift-distance mapping.

By our definition in equation (1.25) the density contrast has average $\langle\delta\rangle = 0$, then its variance is then given by

$$\sigma^2 = \langle \delta(\mathbf{x})^2\rangle = \xi(0). \tag{1.56}$$

When instead of using a continuous field we need to study the correlation of a discrete set of particles, we need to account for the noise generated by the sampling process, known as **Poisson shot noise**. This results in an extra signal generated by the spurious self-correlation the particles. Computing the covariance of the overdensities measured from a discrete set of points δ^d one can show (see e.g. [49])

$$\langle \delta^d(\mathbf{x}_1)\delta^d(\mathbf{x}_2)\rangle = \xi(x_{12}) + \frac{1}{\bar{n}}\delta_D(\mathbf{x}_1 - \mathbf{x}_2), \tag{1.57}$$

where $\bar{n}$ is the mean number density of the particles. We can see that the term that adds up to the correlation function is nonzero and singular only if $\mathbf{x}_1 = \mathbf{x}_2$, so for $\mathbf{x}_1 \neq \mathbf{x}_2$ the covariance of δ^d provides an estimate for the correlation function.

Meaning of the correlation function

Considering the number density of a given tracer n we have:

$$n(\mathbf{x}) = \bar{n}\,[1 + \delta(\mathbf{x})]. \tag{1.58}$$

If we then compute the correlation of the number density at two different locations we get:

$$\langle n(\mathbf{x}_1)n(\mathbf{x}_2)\rangle = \bar{n}^2\,[1 + \xi(x_{12})], \tag{1.59}$$

where we have made use of (1.55) and of the fact that $\langle\delta(\mathbf{r})\rangle = 0$. Considering an uniformly distributed tracer then $n(\mathbf{r}) = \bar{n}$ and we describe the actual distribution of discrete tracers as a result of a Poisson-like process with probability $\bar{n}$. In general the tracer won't be distributed uniformly and equation (1.59) shows how the correlation function describes the excess probability of finding two tracers at a distance r_{12}

with respect to a Poisson process. For small separations r_{12} we expect galaxies to be strongly clustered due to gravity, so ξ is positive. As the distances get bigger we expect galaxies to be less correlated and so ξ dies off to zero, eventually going negative, due to the presence of regions devoid of galaxies between clusters. At even larger scales where we expect the cosmological principle to hold, ξ will grow back to 0, since the galaxy distribution will appear uniform, following a properly random process.

In figure (1.3) we show the correlation function measured from the 2dF Galaxy Redshift Survey [24], with its best fit with a power law function, which is often used to model the correlation of galaxies at scales smaller than $10\,\mathrm{Mpc}/h$:

$$\xi_{\mathrm{fit}}(x) = \left(\frac{x}{r_0}\right)^{-\gamma}, \tag{1.60}$$

where γ and r_0 are the fit parameters.

1.3.3 Power spectrum

We define the Fourier modes of the density field $\tilde{\delta}(\mathbf{k})$ as the Fourier transform of $\delta(\mathbf{r})$:

$$\tilde{\delta}(\mathbf{k}) \equiv \int e^{i\mathbf{kx}}\,\delta(\mathbf{x})\,d^3x\,, \quad \delta(\mathbf{x}) \equiv \int e^{-i\mathbf{kx}}\,\tilde{\delta}(\mathbf{k})\,d^3k/(2\pi)^3\,, \tag{1.61}$$

We then define the **power spectrum** $P(k)$ as the covariance of the Fourier modes:

$$(2\pi)^3\,\delta_D(\mathbf{k}_1+\mathbf{k}_2)\,P(k_1) \equiv \langle\tilde{\delta}(\mathbf{k}_1)\tilde{\delta}(\mathbf{k}_2)\rangle\,, \tag{1.62}$$

where δ_D is the Dirac Delta, which appears due to the homogeneity assumption that grants translational invariance, while the fact that $P(k)$ depends only of the module of the wave vector is granted by the isotropy assumption.

It can be shown that the power spectrum is the Fourier transform of the correlation function [30, 48, 50]:

$$P(k) = \int e^{i\mathbf{kx}}\,\xi(x)\,d^3x = 2\pi\int_0^\infty dx\;x^2\mathrm{sinc}(kx)\xi(x)\,. \tag{1.63}$$

Similarly to the case of the correlation function, computing the power spectrum from a discrete set of points also generates a spurious noise due to the sampling. Fourier transforming equation (1.57) we obtain that the shot noise contribution to the power spectrum consists in a constant additive term (see e.g. [49])

$$\langle\tilde{\delta}^d(\mathbf{k}_1)\tilde{\delta}^d(\mathbf{k}_2)\rangle = P(k_1) + \frac{1}{\bar{n}}\,, \tag{1.64}$$

which needs to be subtracted to correctly estimate the power spectrum.

In figure (1.4) we show the power spectrum measured from the Sloan Digital Sky Survey III [54], with its best fit from the ΛCDM model.

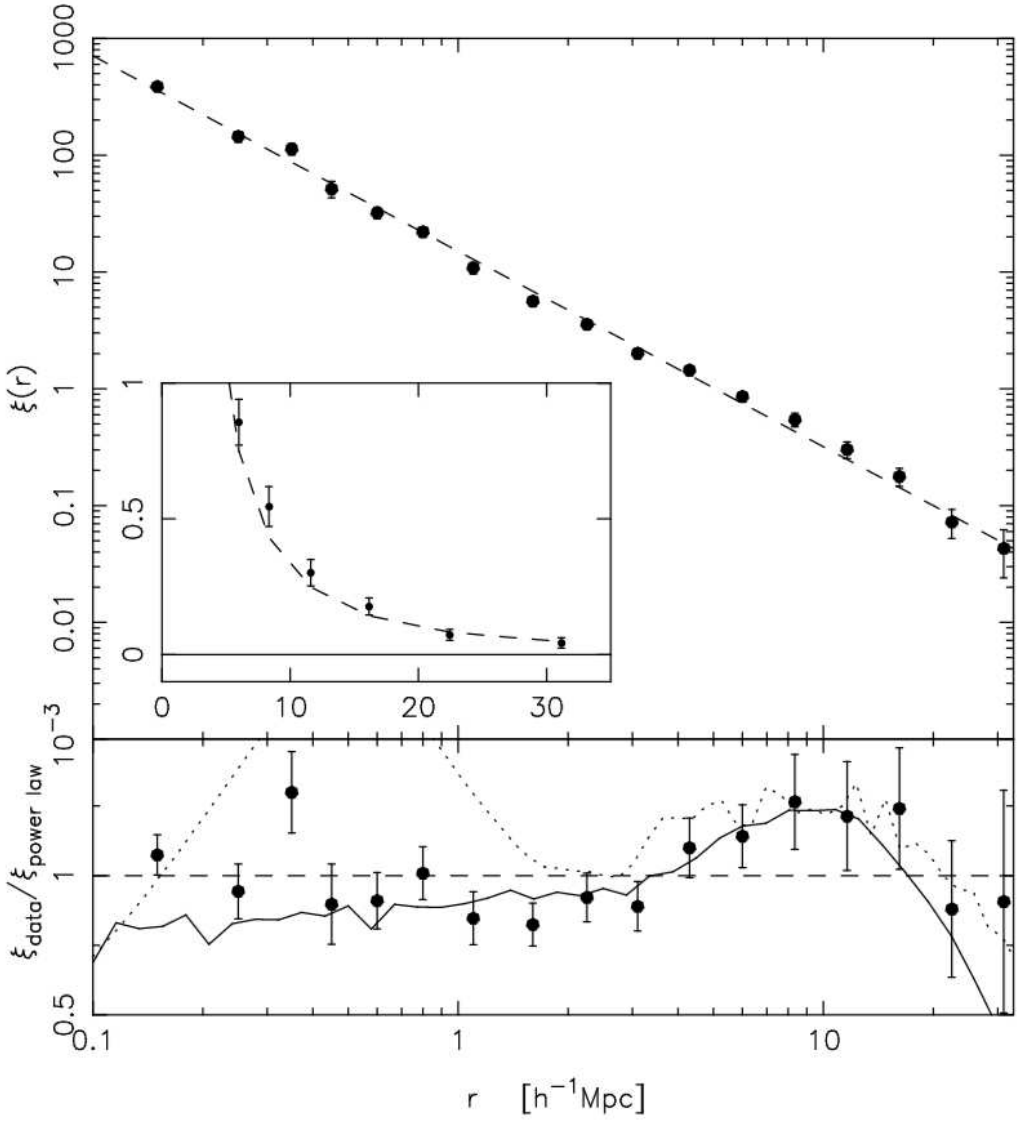


Figure 1.3: Real-space correlation function of the 2dF Galaxy Redshift Survey [24], with error bars from the rms spread between mock catalogues. The data is compared with the best-fit power law described in equation (1.60). Figure taken from [25].

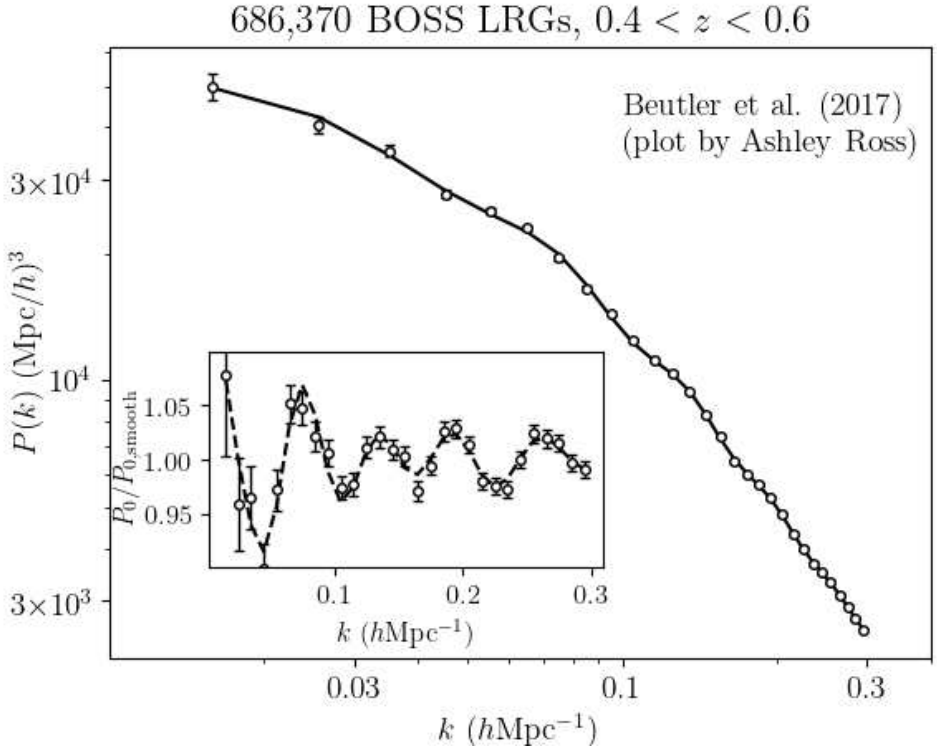


Figure 1.4: Galaxy power spectrum measured from the Baryon Oscillation Spectroscopic Survey (BOSS), part of the Sloan Digital Sky Survey III. The galaxies considered are in the redshift bin $[0.4, 0.6]$ and follow a specific target selection criterion. Reference for the data: [26].

Position-dependent power spectrum

Projection effects and, as we will see, redshift space distortions break the translational symmetry of the observed galaxy distribution. This complicates the treating of Fourier-based statistics such as the power spectrum. For instance, the covariance of two modes of δ_g is not diagonal anymore, i.e. [55, 56]

$$\langle \tilde{\delta}_g(\mathbf{k}_1)\tilde{\delta}_g(\mathbf{k}_2)\rangle = (2\pi)^3\, C(\mathbf{k}_1, \mathbf{k}_2)\,. \tag{1.65}$$

To treat this issue we can define a position-dependent power spectrum [57]:

$$\begin{aligned} P_{\rm loc}(\mathbf{x}, \mathbf{k}) &\equiv \int \langle \delta_g\left(\mathbf{x} - \frac{\mathbf{y}}{2}\right) \delta_g\left(\mathbf{x} + \frac{\mathbf{y}}{2}\right)\rangle\, e^{-i\mathbf{k}\cdot\mathbf{y}}\, d^3y \\ &= \int C\left(-\mathbf{k} + \frac{\mathbf{q}}{2}, \mathbf{k} + \frac{\mathbf{q}}{2}\right)\, e^{i\mathbf{q}\cdot\mathbf{x}}\, d^3y\,, \end{aligned} \tag{1.66}$$

where the second equality was obtained by expressing the overdensity in terms of its Fourier transform.

To extrapolate position-indipendent information from this quantity we can expand it in Legendre polynomials $\mathcal{L}_l$ over $\mathbf{x}$, writing

$$P_l(k) = (2l+1) \int_{V_\mathbf{k}} \frac{d\Omega_k}{4\pi} \int \frac{d^3x}{(2\pi)^3} P_{\rm loc}(\mathbf{x}, \mathbf{k}) \mathcal{L}_l(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}})\,, \tag{1.67}$$

where we computed the angular average over a thin shell in Fourier space centered in k of volume V_k.

Of particular interest is the $l = 0$ expansion of $P_{\rm loc}$, the **power spectrum monopole**, which has the same form as the traditional power spectrum when translational symmetry is preserved, as it can be hinted by integrating over space the position-depedent power spectrum

$$\begin{aligned} \int P_{\rm loc}(\mathbf{x}, \mathbf{k})\, dx^3 &= \int \left[\int C\left(-\mathbf{k} + \frac{\mathbf{q}}{2}, \mathbf{k} + \frac{\mathbf{q}}{2}\right)\, e^{i\mathbf{q}\cdot\mathbf{x}}\, d^3y\right] d^3x \\ &= (2\pi)^3\, C(\mathbf{k}, -\mathbf{k})\,. \end{aligned} \tag{1.68}$$

In conclusion, given an overdensity field $\delta(\mathbf{x})$, the operative definition of computing its power spectrum monopole is the same as the traditional power spectrum.

1.3.4 Gaussian random fields

The correlation function and power spectrum give an equivalent statistical description of the δ field. In general, they don't uniquely characterize the field in question, beside specific cases. One example are the **Gaussian random fields**, which are characterized by the fact that at each point the probability distribution of the density fluctuation $\delta(\mathbf{x})$ is Gaussian. These fields are particularly important in cosmology because it is often assumed that for very early epochs, the density field obeyed Gaussian statistics. This is predicted at large scales by many inflationary models which generate the primordial density fluctuations in the Universe [58]. For this reason, the observation of non-gaussianities in the primordial perturbations would allow

to discriminate from different candidate models for inflation. Furthermore, there is observational evidence of this in the study of the anisotropies of the cosmic wave background radiation [59], which strongly constrain this model.

1.3.5 Linear bias

The complexity of the physics of galaxy formation makes it difficult to assess the relation between galaxies' and the underlying dark matter's distributions. One possible solution is to parametrize the way in which dark matter and galaxy clustering are correlated, by introducing the linear galaxy bias. As we already mentioned, we define the galaxy overdensity as

$$\delta_g(\mathbf{x}) = \frac{n_g(\mathbf{x}) - \bar{n}_g}{\bar{n}_g}\,, \tag{1.69}$$

where n_g is the galaxies' number density and $\bar{n}_g$ its average value.

We will refer instead to δ for the dark matter overdensity. At a given epoch the galaxy linear bias b is introduced as:

$$\delta_b(\mathbf{x}) = b\,\delta(\mathbf{x})\,, \tag{1.70}$$

which also implies:

$$n_g(\mathbf{x}) = \bar{n}_g\,[1 + b\,\delta(\mathbf{x})]\,. \tag{1.71}$$

This definition assumes a non-local and linear relation between galaxies and the underlying matter field. Equation (1.70) will be applicable only on large scales, where $|\delta| \ll 1$, otherwise in underdense regions with $\delta = -1$ it could imply a negative galaxy density $n_g(\mathbf{x})$.

Under this assumption it is easily shown that the galaxy correlation function and power spectrum are related to the matter ones by:

$$\xi_g(r) = b^2\,\xi(r)\,, \qquad P_g(k) = b^2\,P(k)\,. \tag{1.72}$$

These relations are not accurate in scales in which equation (1.70) is not valid.

Justification for equation (1.70)

The assumptions of equation (1.70) can be explained with different approaches. In this subsection we will look into the **threshold bias** approach [60].

This approach starts from the assumption that gravitationally bound structures form in regions where the linear density field, when coarse-grained on a scale R, lays above some threshold.

Consider a zero-mean stationary Gaussian random field $\delta(\mathbf{x})$, with correlation function $\xi(r)$ and variance $\sigma^2 = \xi(0)$. We want to study the distribution of regions in which $\delta > \nu\sigma$, where ν parameterizes the said threshold. These regions have an overdensity $\delta_{>\nu} = \Theta[\delta(\mathbf{x}) - \nu\sigma]$, where Θ is the Heaviside step distribution. Then,

the correlation function of the distribution of these points can be written as [61,62]:

$$\begin{aligned} 1+\xi_{>\nu}(r) &= \frac{\langle\Theta[\delta(\mathbf{x})-\nu\sigma]\Theta[\delta(\mathbf{x}+\mathbf{r})-\nu\sigma]\rangle}{\langle\Theta[\delta(\mathbf{x})-\nu\sigma]\rangle^2} \\ &= \frac{\int_\nu^\infty\int_\nu^\infty G(y_1,y_2)\,dy_1\,dy_2}{\left[\frac{1}{2}\mathrm{erfc}\left(\frac{\nu}{2}\right)\right]^2}\,, \end{aligned} \tag{1.73}$$

where $G(y_1, y_2)$ is the joint probability function for the normalized field $y = \delta/\sigma$ at the locations $\mathbf{x}$ and $\mathbf{x}+\mathbf{r}$:

$$G(y_1,y_2) = \frac{\exp\left[-\frac{y1^2+y_2^2-2w(r)y_1y_2}{2[1-w^2(r)]}\right]}{2\pi[1-w^2(r)]^{\frac{1}{2}}}\,, \tag{1.74}$$

with $w(r) = \xi(r)/\sigma^2$.

If we expand equation (1.73) to first order in $w(r)$ and we take the limit $\nu \gg 1$ we find:

$$\xi_{>\nu}(r) \approx \frac{\nu^2}{\sigma^2}\,\xi(r)\,, \tag{1.75}$$

which is consistent with the result obtained by assuming a linear bias relation, assigning $b = \nu/\sigma$.

1.4 Projection effects on the galaxy distribution

The main source of information regarding the galaxy distribution in the Universe is given by **redshift surveys**, which measure the angular positions and redshift of galaxies in the sky.

A survey is mainly characterized by its shape and selection criteria. Regarding the shape, in general only a region of the sky is probed: this can be due to the physical positioning of the measuring instrument, such as in the case of ground based telescopes, or to avoid specific regions, such as the Galactic plane. Selection criteria to the sources are often applied due to practical reasons, such as choosing only targets bright enough to have a precise redshift measurement (i.e. applying a flux cut), or selecting based of the target size, to avoid including stars.

From the data provided from the survey we can construct the spatial distribution of galaxies by retrieving the distance of the source making use of the measured redshift and of its relation with the comoving distance, mapping $(z, \theta, \psi) \to (\chi, \theta, \psi)$.

Tainting this prescription is the fact that the galaxies' peculiar velocities add up to the Hubble flow, altering the emitted radiation and shifting the observed redshift with respect of the cosmological one, thus shifting the inferred distance with respect to the true, comoving one. This effect was first studied by Kaiser in [2], where he showed that at large scales the galaxies' peculiar velocities boost the power spectrum monopole by a factor that depends on the linear bias and on the linear growth rate. In a similar way, also the observer's peculiar velocity distorts the measured redshift,

creating a dipolar patter in the galaxy distribution [1,63,64]. Lastly, the path that a photon follows in reaching us from a galaxy is curved by the underlying gravitational potential, which shifts both the observed radial and angular position of the galaxy, causing also a time delay to the observed signal [8–11].

Beside affecting the spatial distribution of the galaxies, both the peculiar velocities and the gravitational potential also distort the apparent flux measured from a source, and this can change the galaxy distribution properties near the survey's flux cut. We will refer to this effect as **magnification** [8,65].

It is a known fact that the perturbations due to the peculiar velocities dominate over the terms generated by the gravitational potential [8–11], but the next generation of galaxy surveys (see e.g. Euclid [13], DESI [14], SKA [15]) will probe scales comparable with the Hubble radius, at which we expect to observe these contributions.

In this section we will review these effects, starting by considering only the ones sourced by the peculiar velocities, following the treatments presented in [1, 30, 45]. We will then incorporate the complete general-relativistic prescription, accounting in this way also for the gravitational potential contributions and for the light-cone projection effects we mentioned in section (1.3), following [8–11].

1.4.1 Redshift space distortions

In this subsection we will start by describing the first treatment of velocity distortions to the measured redshift, labelled as **redshift space distortions** (RSD).

As we have anticipated, the velocity of a light source stretches or compresses the emitted wavelength generating the **doppler redshift**, which for a peculiar velocity much smaller that the speed of light $|\mathbf{v}| \ll 1$ can be linearized to

$$z_{\text{doppler}} = \mathbf{v} \cdot \hat{\mathbf{x}} \,, \tag{1.76}$$

where $\mathbf{x}$ is the source's comoving position, and $\hat{\mathbf{x}}$ is the associated versor. Despite the nomenclature, this effect can either stretch or compress the emitted light, allowing also for blueshifting effects. In general, also the velocity of the observer will distort the observed redshift in a similar way: we will account for this effect later in this section, while for now we will consider the case of a stationary observer with respect of the CMB.

The doppler redshift then adds up to the cosmological one z_{cosmo}, leading to the relation [30, 49, 66]

$$z_s = z_{\text{cosmo}} + z_{\text{doppler}} = z_{\text{cosmo}} + \mathbf{v} \cdot \hat{\mathbf{x}}_r \,, \tag{1.77}$$

where the subscript "$_s$" denotes that a quantity is observed or inferred by the observations while "$_r$" denotes the real one. We will then define the space built from the redshift measurements as **redshift space**.

Around the time of the introduction of this field, most redshift surveys probed only the $z \lesssim 0.1$ region of the Universe (e.g. the CfA and Cfa2 surveys [67]), so the redshift-comoving distance relation was used in its linearized form in z, that consists

in a different formulation of Hubble's law [22]

$$\chi(z) \approx \frac{z}{H_0}\,, \tag{1.78}$$

which becomes progressively less accurate as z grows, and holds only if z is the cosmological redshift. Plugging equation (1.77) in equation (1.78) we obtain the mapping

$$|\mathbf{x}_s| = |\mathbf{x}_r| + \frac{\mathbf{v}\cdot\hat{\mathbf{x}}_r}{H_0}\,, \tag{1.79}$$

which in three dimensions becomes

$$\mathbf{x}_s = \mathbf{x}_r + \frac{\mathbf{v}\cdot\hat{\mathbf{x}}_r}{H_0}\,\hat{\mathbf{x}}_r\,. \tag{1.80}$$

From equation (1.80) we see how the velocity effect to the inferred position is purely radial, with a contribution that depends only on its component along the Line of Sight (LOS). More specifically, an object moving towards us will have $\mathbf{v}\cdot\hat{x}_r < 0$, which means it will appear closer to us, while in the opposite case $\mathbf{v}\cdot\hat{x}_r > 0$ it will appear farther away. Lastly, for a motion purely perpendicular to the LOS the contribution will be null, since $\mathbf{v}\cdot\hat{x}_r = 0$.

In figure (1.5) we show the cumulative effect of these distortions when considering the clustering of a spherical distribution of galaxies at different scales, by comparing their positions in real and redshift space, positioning us as an observer from below the image.

As we expect, the angular size of these structures projected on the sky is left unchanged, while for their LOS component the effect depends on the scale and on the phase of the collapse. In the first panel we consider large scales, where the $\mathbf{v}\cdot\hat{x}_r/H_0$ term is small compared to the size of the structure, and the cumulative effect consists in a squashing along the LOS. This is because galaxies from the farther side of the sphere will be moving towards us, appearing closer, while the opposite happens from the galaxies in the closer side of the sphere. In the second panel we consider the turnaround phase, in which the distortions from the two sides of the sphere cancel out with the radius of the sphere, generating a straight line configuration. In the last panel we are considering an already virialized structure, in which the velocity effects can be larger that the actual size of the sphere, turning "inside out" the shape and elongating it along the LOS. These effects are referred to as "Fingers of God".

1.4.2 Observed density field

As we already mentioned, in a redshift survey we only probe a region of the Universe and we only measure the redshift of a restricted sample of all the galaxies present in that region.

The **selection function** $\bar{n}_{g,s}(\mathbf{x}_s)$ describes the expected number of galaxies that a given survey will observe at a position $\mathbf{x}_s$. Since $\bar{n}_{g,s}$ is evaluated in redshift space it can be estimated from the observed data [1].

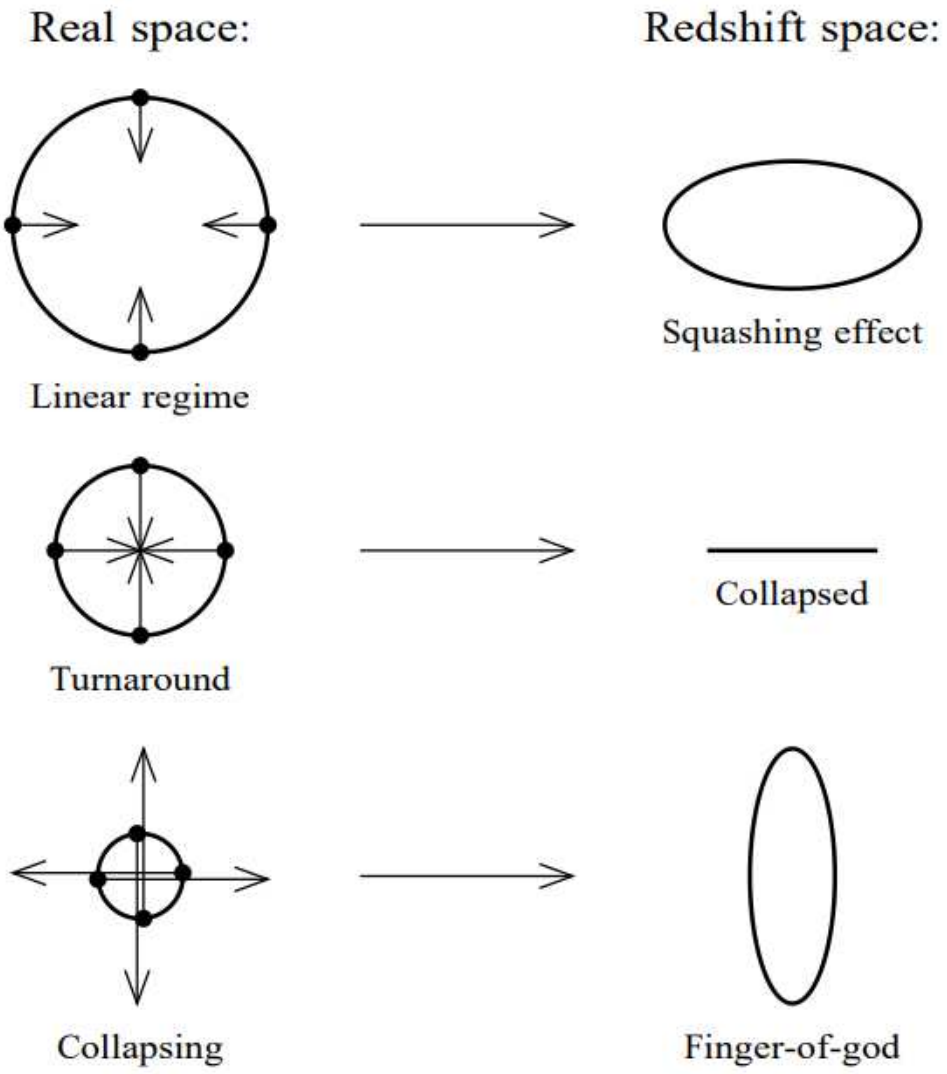


Figure 1.5: Detail of how peculiar velocities lead to the redshift distortions. The dots are 'galaxies' undergoing infall towards a spherical overdensity, and the arrows represent their peculiar velocities. At large scales, the peculiar velocity of an infalling shell is small compared to its radius, and the shell appears squashed. At smaller scales, not only is the radius of a shell smaller, but also its peculiar infall velocity tends to be larger. The shell that is just at turnaround, its peculiar velocity just cancelling the general Hubble expansion, appears collapsed to a single velocity in redshift space. At yet smaller scales, shells that are collapsing in proper coordinates appear inside out in redshift space. The combination of collapsing shells with previously collapsed, virialized shells, gives rise to Fingers-of-God. Picture and description taken from [1].

If at a position $\mathbf{x}_s$ we measure a number density $n_g(\mathbf{x}_s)$, we can define the observed galaxy overdensity as

$$\delta_{g,ss}(\mathbf{x}_s) \equiv \frac{n_{g,s}(\mathbf{x}_s) - \bar{n}_{g,s}(\mathbf{x}_s)}{\bar{n}_{g,s}(\mathbf{x}_s)}\,, \tag{1.81}$$

which is an observable, since it is evaluated in redshift space.

The situation is more complex when we define the overdensity in real space

$$\delta_{g,r}(\mathbf{x}_r) \equiv \frac{n_{g,r}(\mathbf{x}_r) - \bar{n}_{g,r}(\mathbf{x}_r)}{\bar{n}_{g,r}(\mathbf{x}_r)}\,. \tag{1.82}$$

In this case the selection function $\bar{n}_{g,r}$ is not only evaluated at a different position with respect to the redshift space counterpart, it is a completely different function, and in general can not be estimated from the data, since we would need to measure the real space galaxy distribution. Many works mistakenly stated that the two functions agree at linear order, but this is not true, as it was shown by Hamilton in [1].

In conclusion, the two selection functions are different, but since the standard approach to model the redshift distortions relied on their equivalence, we will proceed by re-defining the redshift space overdensity as

$$\delta_{g,s}(\mathbf{x}_s) \equiv \frac{n_{g,s}(\mathbf{x}_s) - \bar{n}_{g,r}(\mathbf{x}_s)}{\bar{n}_{g,r}(\mathbf{x}_s)}\,, \tag{1.83}$$

in which we used the real space selection function.

1.4.3 Two-point statistics

Starting from the redshift space overdensity $\delta_{g,s}$ we can compute its covariance, the **redshift space correlation function**

$$\xi^s(x_{s,12}, x_{s,1}, x_{s,2}) \equiv \langle \delta_{g,s}(\mathbf{x}_{s,1}) \delta_{g,s}(\mathbf{x}_{s,2}) \rangle\,, \tag{1.84}$$

where $x_{s,12} = |x_{s,2} - x_{s,1}|$.

One thing to notice here is that redshift-space distortions destroy the translational symmetry of the observed Universe, but they keep the rotational symmetry about the position of the observer (as long as the selection function is spherically symmetric) [1].

As a consequence, the correlation function depends also on the distance of the two points from the observer. This is because, in a general case without any symmetry, the correlation function would depend on both the points $\mathbf{x}_{1,s}$, $\mathbf{x}_{2,s}$. Introducing a spherical symmetry around the observer means that we only have to keep fixed their relative position and their distance from us, meaning that the only information needed are $x_{s,12}$, $x_{s,1}$ and $x_{s,2}$.

We can also define the **redshift space power spectrum**

$$\langle \tilde{\delta}_{g,s}(\mathbf{k}_1) \tilde{\delta}_{g,s}(\mathbf{k}_2) \rangle = \int e^{i\mathbf{k}_1\mathbf{s}_1 + i\mathbf{k}_2\mathbf{s}_2} \xi^s(x_{s,12}, x_{s,1}, x_{s,2})\, d^3x_{s,1}\, d^3x_{s,2}\,, \tag{1.85}$$

where $\tilde{\delta_{g,s}}(\mathbf{k})$ are the Fourier modes of the redshift space overdensity. Notice how we cannot write the power spectrum as a diagonal matrix since translational symmetry

is broken. We can still make use of the rotational symmetry around the observer to write it only in function of the modulus of the combination of the modes, and not in function of their orientation

$$\langle \tilde{\delta}_{g,s}(\mathbf{k}_1)\tilde{\delta}_{g,s}(\mathbf{k}_2)\rangle = \tilde{\xi}^s(|\mathbf{k}_1+\mathbf{k}_2|, k_1, k_2)\,. \tag{1.86}$$

1.4.4 Linear redshift distortion operator

In order to quantify how RSDs influence the measured summary statistics we first have to study how they affect the observed galaxy overdensity field δ_g. We will show that the the overdensity measured in redshift space $\delta_{g,s}$ is related to the one in real space $\delta_{g,r}$ via a linear operator $\mathbf{S}$, called the **linear redshift distortion operator**

$$\delta_{g,s} = \mathbf{S}\,\delta_{g,r}\,. \tag{1.87}$$

The starting point of the derivation of this relation is the conservation equation, which states that peculiar velocities displace galaxies along the line-of-sight, but they do not make them appear or disappear (we will see later that taking into account selection and projection effects will contradict this statement)

$$n_{g,s}(\mathbf{x}_s)\,d^3x_s = n_{g,r}(\mathbf{x}_r)\,d^3x_r\,. \tag{1.88}$$

Writing the equation in terms of the overdensities we get

$$\bar{n}_{g,r}(\mathbf{x}_s)[1+\delta_{g,s}(\mathbf{x}_s)]{x_s}^2\,dx_s = \bar{n}_{g,r}(\mathbf{x}_r)[1+\delta_{g,r}(\mathbf{x}_r)]{x_r}^2\,dx_r\,, \tag{1.89}$$

where we remind that we are using $\delta_{g,s}$ and not $\delta_{g,ss}$. Using equation (1.80) and defining $v_\parallel = \mathbf{v}\cdot\hat{\mathbf{x}}$ we get we can rewrite it as [1, 30]:

$$1+\delta_{g,s}(\mathbf{x}_s) = \frac{{x_r}^2\,\bar{n}_{g,r}(\mathbf{x}_r)}{(x_r+v_\parallel/H_0)^2\bar{n}_{g,r}(\mathbf{x}_r+v_\parallel/H_0\hat{\mathbf{x}}_r)}\left(1+\frac{1}{H_0}\frac{\partial v_\parallel}{\partial x_r}\right)^{-1}[1+\delta_{g,r}(\mathbf{x}_r)]\,, \tag{1.90}$$

which is an exact solution, valid both in linear and nonlinear regimes up to shell crossing, where multiple streams overlap and we cannot assign a unique velocity field to each point.

We now want to linearize the equation. The assumption $|\delta| \ll 1$ also implies $|\partial v_\parallel/\partial x_r|/H_0 \ll 1$ for a irrotational velocity field in linear theory (see equations (1.48) and (1.49)). We will also assume that the peculiar velocities in units of H_0 of the galaxies are small compared to their distances from the observer $|v/H_0| \ll r$. Then at linear order:

$$\delta_{g,s}(\mathbf{x}_r) = \delta_{g,r}(\mathbf{x}_r) - \frac{1}{H_0}\left(\frac{\partial}{\partial x_r}+\frac{\alpha(\mathbf{x}_r)}{x_r}\right)v_\parallel(\mathbf{x}_r)\,, \tag{1.91}$$

where

$$\alpha(\mathbf{x}_r) \equiv \frac{\partial \ln {x_r}^2\bar{n}(\mathbf{x}_r)}{\partial \ln x_r} \tag{1.92}$$

describes the change with distance of the selection function.

Inverting the linearized continuity equation (1.50) and assuming a bias relation $\delta_{g,r} = b\,\delta_{\rm matter}$ we obtain:

$$v_\parallel = -aH\beta\frac{\partial}{\partial r}\nabla^{-2}\delta_{g,r}\,, \tag{1.93}$$

where

$$\beta = \frac{f}{b} \tag{1.94}$$

is the **linear redshift distortion parameter**, given by the fraction of the linear growth rate and galaxy linear bias. Since we are working under the $z \ll 1$ assumption we evaluate all the parameters at $z = 0$, which means setting $a = 1$, $H = H_0$ and $\beta = \beta_0 = \beta(z = 0)$.

Plugging equation (1.93) into (1.91) we finally get the form of the operator

$$\mathbf{S} = 1 + \beta_0\left(\frac{\partial^2}{\partial {x_r}^2} + \frac{\alpha(\mathbf{x}_r)\partial}{x_r\partial x_r}\right)\nabla^{-2}\,, \tag{1.95}$$

which we remind is valid for a stationary observer and a selection function estimated in real space $\bar{n}(\mathbf{x_r})$.

1.4.5 Plane-parallel limit

We want to study $\mathbf{S}$ in the plane-parallel limit, which is obtained by asking that the galaxies observed are far enough that we can approximate their light rays reaching us to be parallel. More precisely, we are asking that the separation angle of two given galaxies is $\theta \ll 1$ (see e.g. figure (1.6)).

Within these assumptions it can be shown that the redshift space correlation functions depends only on the components of $\mathbf{x}_{s,12}$ parallel ($x_{s,\parallel}$) and perpendicular ($x_{s,\perp}$) to the LOS [1, 30]

$$\xi^s(x_{s,12}, x_{s,1}, x_{s,2}) \approx \xi^s(x_{s,\parallel}, x_{s,\perp})\,. \tag{1.96}$$

An example of a measured correlation function in redshift space is shown in figure (1.7). For the same simmetry reasons, in this limit the power spectrum becomes a diagonal matrix [1, 30],

$$\langle\hat{\delta}_{g,s}(\mathbf{k}_1)\hat{\delta}_{g,s}(\mathbf{k}_2)\rangle = (2\pi)^3\,\delta_D(\mathbf{k}_1 + \mathbf{k}_2)\,P^s(k_{1\parallel}, k_{1\perp})\,, \tag{1.97}$$

where $k_\parallel$ and $k_\perp$ are the $\mathbf{k}_1$ components parallel and perpendicular to the line-of-sight. Regarding the linear redshift distortion operator, this approximation allows us to write it as [30, 66]:

$$\mathbf{S}^p = 1 + \beta_0\,\frac{\partial^2}{\partial z^2_{\rm LOS}}\nabla^{-2}\,, \tag{1.98}$$

where $z_{\rm LOS}$ is the distance along the line of sight. Writing the operator in Fourier space we find out that it is diagonal

$$\mathbf{S}^p = 1 + \beta_0\,\mu_{\mathbf{k}}^2\,, \tag{1.99}$$

with $\mu_{\mathbf{k}} \equiv \hat{\mathbf{z}}_{\rm LOS}\cdot\hat{\mathbf{k}}$ is the cosine of the angle of the $\mathbf{k}$ wavevector with the line of sight. Then we can write the redshift space power spectrum in a diagonal form

$$P^s(\mathbf{k}) = (1+\beta_0\,\mu_{\mathbf{k}}^2)^2 P(k)\,. \tag{1.100}$$

The reason why we retrieved a diagonal power spectrum is to be found into the fact that in the plane-parallel approximation redshift space distortions preserve translation invariance: the translation operator $-i\nabla$ commutes with the plane-parallel operator and thus they share eigenmodes [1]. We can retrieve the monopole of the power spectrum simply by performing an angular average, since $P_s(\mathbf{k})$ is translationally invariant [1, 2]

$$P_{0,s}(k) \equiv \int \frac{d\Omega_k}{4\pi}(1+\beta_0\,\mu_{\mathbf{k}}^2)^2 P(k) = \mathcal{F}P(k)\,, \tag{1.101}$$

where

$$\mathcal{F} = 1 + \frac{2}{3}\beta_0 + \frac{1}{5}\beta_0^2 \tag{1.102}$$

is the **Kaiser factor**. In conclusion, in the plane-parallel approximation at large scales, the galaxies' peculiar velocities boost the power spectrum monopole by a factor that depends on the linear bias and on the linear growth factor.

1.4.6 Velocity of the observer

The mapping we defined in equation (1.80) is valid for a stationary observer, but in order to take into account a moving observer with velocity $\mathbf{v}^{\rm obs}$ we will generalize it to (see e.g. [1, 63, 64]):

$$\mathbf{x}_s = \mathbf{x}_r + \frac{\mathbf{v}\cdot\hat{\mathbf{x}}_r}{H_0}\,\hat{\mathbf{x}}_r - \frac{\mathbf{v}_{\rm obs}\cdot\hat{\mathbf{x}}_r}{H_0}\,\hat{\mathbf{x}}_r\,. \tag{1.103}$$

One can show that following a similar approach as in the previous case the overdensity maps as:

$$\delta_{g,s}^{\rm obs}(\mathbf{x}_r) = \delta^s(\mathbf{x}_r) + \alpha(\mathbf{x}_r)\frac{\hat{\mathbf{x}}_\mathbf{r}\cdot\mathbf{v}^{\rm obs}(\mathbf{x}_\mathbf{r})}{x_r}\,. \tag{1.104}$$

The result is a dipole contribution directed along the observer's motion direction. The dipole sourced by the peculiar motion of the observer is known as **rocket effect**, and it calls for a delicate approach: the Local Group motion induces a spurious overdensity in the direction of motion, which may then be wrongly identified as the cause of the motion if not correctly accounted for [63, 64]. In fact, this effect could affect the estimation of the observer's peculiar velocity via clustering measurements, as it acts on the observed overdensity in equation (1.50), and could also bias the estimation of cosmological parameters, due to its contribution on summary statistics (see e.g. [63]). We will see how this effect affects the galaxy power spectrum in subsection (1.4.13), where we will consider a full relativistic treatment.

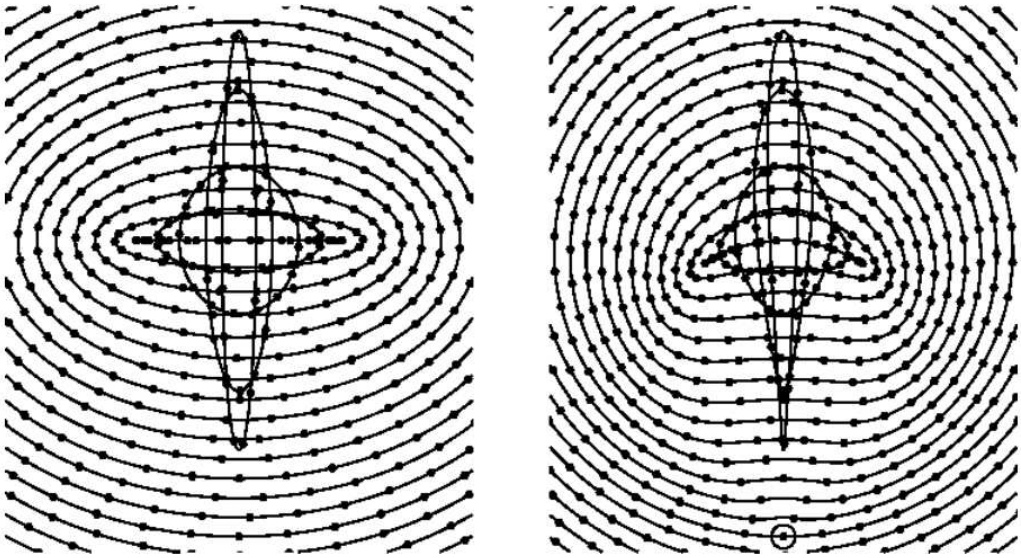


Figure 1.6: A spherical overdensity appears distorted by peculiar velocities when observed in redshift space. On the left plot, the overdensity is far from the observer (who is looking upward from somewhere way below the bottom of the diagram), and the distortions are effectively plane-parallel. On the right plot, the overdensity is near the observer (large dot), and the large scale distortions appear kidney-shaped. The observer shares the infall motion towards the overdensity. Picture taken from [1].

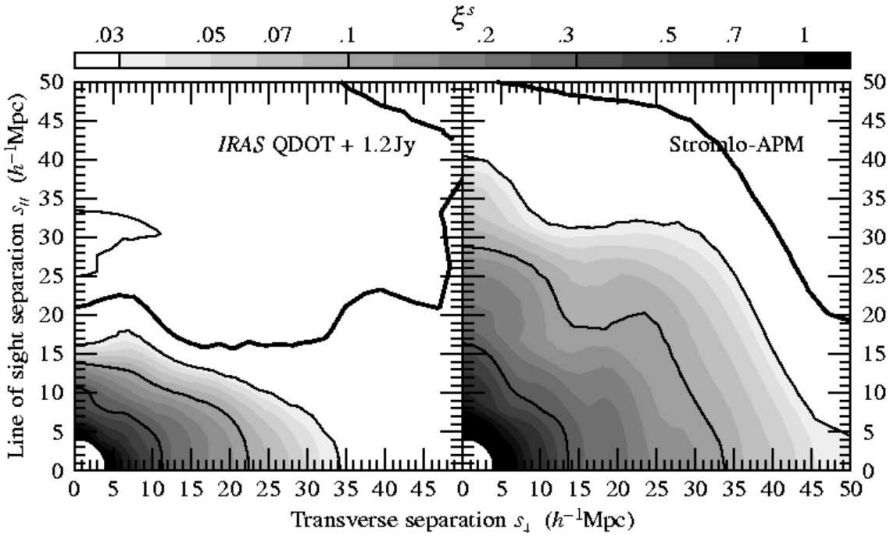


Figure 1.7: Contour plots of the redshift space two-point correlation function ξ_s as a function of separations $s_{\parallel}$ and $s_{\perp}$ parallel and perpendicular to the line of sight in: (left) the IRAS QDOT [27] and 1.2 Jy [28] redshift surveys, merged over the angular region of the sky common to both surveys; and (right) the optical Stromlo-APM [29] survey. Picture and description taken from [1]

1.4.7 Doppler effect

The term proportional to α that we neglected in the previous treatment is referred to as **Doppler term** [3,4] and it is proportional to $\alpha(x_r)\, v_\parallel / x_r$. This approximation has been shown to be accurate for past surveys, such as the SDSS-II [4], but it stops holding when considering wide angular separations. In figure (1.8) we see how the introduction of varying line of sights modifies the directions along we shift galaxies from real to redshift-space. Consider a spherical distribution in real-space: in the plane-parallel approximation (left plots) the Kaiser effect will flatten it at large scales, producing the *Pancakes of God*, while at small scales we see the Fingers of God. Without the plane parallel approximation (right plots) the different line-of-sights will deform the spherical distribution in *Croissants of God*. These contributions will be correctly implemented in the next subsection, where we consider also other relativistic corrections that should be taken into account at large angular separations.

1.4.8 Relativistic projection effects

Now we will include general relativistic effects in the study of redshift space distortions, which will allow us to consider scales comparable with the horizon scale. We can show why we expect these effects to become important at large scales by considering a flat Universe and combining the Poisson equation (1.33) and the first Friedmann equation (1.15)

$$\nabla^2_{\mathbf{x}}\phi = 4\pi G\bar{\rho}a^2\,\delta\,, \qquad H^2 = \frac{8\pi G}{3}\bar{\rho}\,. \tag{1.105}$$

We obtain

$$\nabla^2_{\mathbf{x}}\phi = \frac{3}{2}a^2H^2\,\delta\,, \tag{1.106}$$

which in Fourier space is

$$k^2\tilde{\phi}_{\mathbf{k}} = \frac{3}{2}a^2H^2\,\tilde{\delta}_{\mathbf{k}}\,, \tag{1.107}$$

Considering this equation in terms of physical scales we obtain

$$\tilde{\phi}_{\mathbf{k}} \approx \left(\frac{\lambda_{\rm phys}}{\lambda_H}\right)^2 \tilde{\delta}_{\mathbf{k}}\,, \tag{1.108}$$

where the scale related to the wave vector k is $\lambda_{\rm phys} = a/k$, while $\lambda_H = 1/H$ is the Hubble radius. We see that, at large scales, where $\delta < 1$, we have $\phi \sim 1$ only when $\lambda_{\rm phys} \approx \lambda_H$: at scales comparable with the Hubble radius we must account for effects sourced by the gravitational potential. Beside terms sourced by the potential, we will also account for the fact that we are observing galaxies on the lightcone hypersurface and we will use the true redshift-comoving distance relation $\chi(z)$, so that the results will be valid also in the $z \gtrsim 1$ domain.

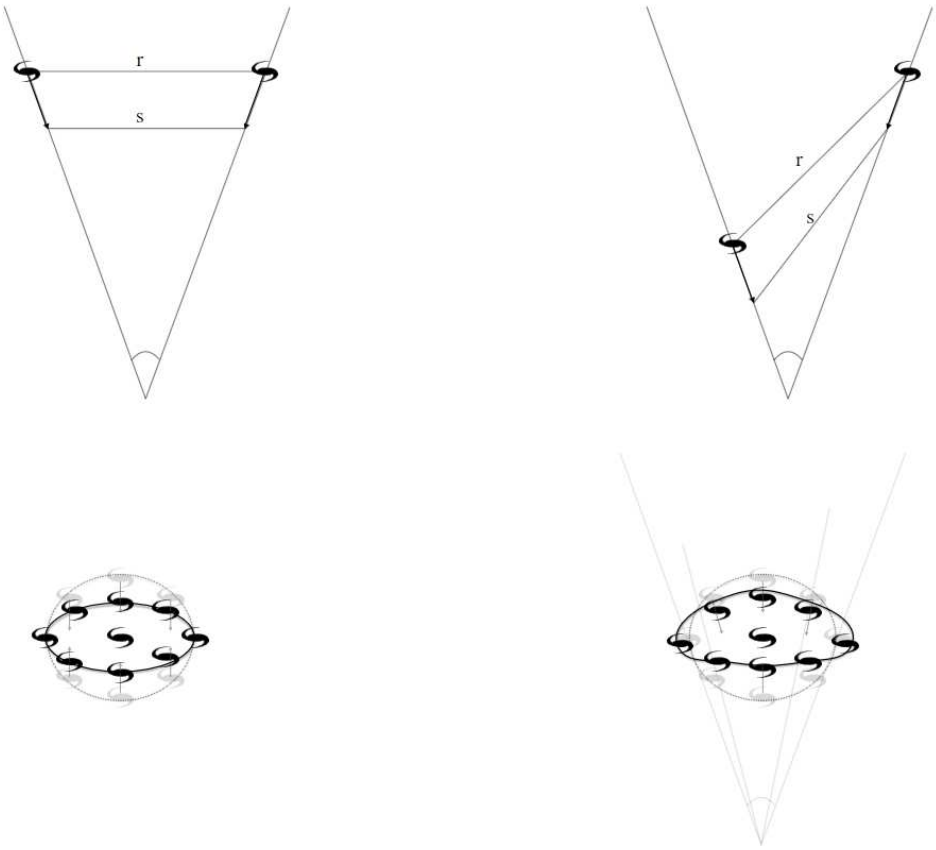


Figure 1.8: Upper panels: Translation from real to redshift-space of a galaxy pair in the transverse and non-transverse case, in the wide-angle case. Note that, even in the case of $|\mathbf{v}_1| = |\mathbf{v}_2|$, the presence of the observer changes the apparent scale s in the transverse case, and both the apparent scale and the apparent angle w.r.t. the line of sight in the non-transverse case. Bottom panels: Large scale apparent modification of a spherical overdensity region. In the plane-parallel approximation (left panel), the Kaiser effect induces the so-called "Pancakes of God", so a spherical distribution of galaxies in real space will appear flattened in the radial direction in redshift space. In the wide-angle case (right panel), the introduction of an observer modifies the shape into a curved croissant-like shape that depends on the angular separation. Picture and description taken from [3].

To do so, we study the null geodesic followed by a photon reaching us from a galaxy. In an unperturbed universe, we refer to this geodesic x_r^μ as in **real space**, and we describe it in conformal units [9, 11, 51] :

$$x_r^\mu = (\eta_r, \mathbf{x}_r) = (\eta_0 - \chi_r, \chi_r \mathbf{n}_r)\,, \tag{1.109}$$

where η_0 is the present conformal time, χ_r is the comoving distance from the observer along the trajectory and $\mathbf{n}_r = \mathbf{x}_r/\chi_r$ is a unit vector pointing towards the emitting galaxy. From now up to the rest of this section we will refer to ∂_μ as the partial derivatives over the conformal units $(\eta, \mathbf{x})$. We investigate how this geodesic changes in a metric perturbed at first order. We refer to these trajectories x_s^μ as in **redshift space**, and we will write them as:

$$x_s^\mu = (\eta_s, \mathbf{x}_s) = (\eta_0 - \chi_s, \chi_s \mathbf{n}_s)\,, \tag{1.110}$$

where $\chi_s = |\mathbf{x}_s|$ is the observed comoving distance and $\mathbf{n}_s = \mathbf{x}_s/\chi_s$. The observed redshift z_s is related to the comoving distance via the relation $\chi_s = \chi(z_s)$. We define a mapping between the two positions as:

$$x_r^\mu(\chi_r) = x_s^\mu(\chi_s) + \Delta x^\mu(\chi_s)\,. \tag{1.111}$$

Defining also $\delta x^\mu(\chi) = x_s^\mu(\chi) - x_r^\mu(\chi)$ and $\delta\chi = \chi_s - \chi_r$ at linear order we get

$$x_r^\mu(\chi_r) = x_s^\mu(\chi_s) + \delta x^\mu(\chi_r) = x_s^\mu(\chi_s) + \frac{dx_s^\mu}{d\chi_s}\,\delta\chi + \delta x^\mu(\chi_s)\,, \tag{1.112}$$

where we use χ as the affine parameter $d/d\chi_s = -\partial/\partial\eta + n_s^i \partial/\partial x_s^i$. We can then write the linear redshift-space distortions as:

$$\Delta x^0(\chi_s) = -\delta\chi + \delta x^0(\chi_s)\,, \tag{1.113}$$

$$\Delta x^i(\chi_s) = n_s^i \delta\chi + \delta x^i(\chi_s)\,. \tag{1.114}$$

The functions δx^0, $\delta\chi$ and δx^i will depend on the characteristics of the perturbed space, i.e. the metric perturbations and the velocities of the source and observer.

1.4.9 Metric perturbations

We describe the perturbations of the FLRW metric in the conformal-Newtonian gauge [30, 68]:

$$ds^2 = a^2(\eta)\left[(1+2\Psi)\,d\eta^2 - (1-2\Phi)\delta_{ij}\,dx^i\,dx^j\right], \tag{1.115}$$

where Ψ and Φ are the Bardeen potentials [68]. This gauge is particularly useful for the purpose of this project because, as we will see later in section (2.1.1), allows to write the Poisson equation for the potential of the simulation in a simple formalism.

By solving the geodesic equations one can work out the formulas for the shifts [8–11]:

$$\begin{aligned}\delta\chi =& -\left(\chi_s + \frac{1}{\mathcal{H}}\right)\left[\Psi_o - (n_s^i v_i)_o\right] + \frac{1}{\mathcal{H}}\left[\Psi_e - (n_s^i v_i)_e\right] \\ &+ \int_0^{\chi_s} [2\Psi + (\chi_s - \chi)\partial_0(\Phi + \Psi)]\, d\chi \\ &+ \frac{1}{\mathcal{H}}\int_0^{\chi_s} \partial_0(\Phi + \Psi)\, d\chi\,,\end{aligned} \tag{1.116}$$

$$\begin{aligned}\delta x^0 =& -\chi_s\left[\Psi_o - (n_s^i v_i)_o\right] + 2\int_0^{\chi_s} \Psi\, d\chi \\ &+ \int_0^{\chi_s} (\chi_s - \chi)\partial_0(\Phi + \Psi)]\, d\chi\,,\end{aligned} \tag{1.117}$$

$$\begin{aligned}\delta x^i =& -\left(v_o^i + \Phi_o n_s^i\right)\chi_s + 2n_s^i\int_0^{\chi_s} \Phi\, d\chi \\ &+ \int_0^{\chi_s} (\chi_s - \chi)\delta^{ij}\partial_j(\Phi + \Psi)]\, d\chi\,,\end{aligned} \tag{1.118}$$

$$\delta \ln a = -\frac{\delta z}{1+z} = -\Phi_o + (n_s^j v_j)_o + \Phi_e - (n_s^j v_j)_e + \int_0^{\chi_s} \partial_0(\Phi + \Psi)\, d\chi\,, \tag{1.119}$$

where $\mathcal{H} = \partial \ln H/\partial \ln a$, and the suffix "$o$" denotes quantities evaluated at the observer space-time position, while "e" denotes the ones evaluated at the emitter space-time position. We can recognize different types of contributions:

- The velocity shifts, which are the terms proportional to the velocities of the observer v_o and of the emitter v_e and are a local contribution.
- The Sachs-Wolfe effects (SW) [69], which are the terms proportional to the potentials outside the integrals, measured at the observer's location (Φ_o, Ψ_o) and at the emitter's location (Φ_e, Ψ_e). They can be explained as the variation in the photon's energy as it starts and finishes at two different potential wells. These effects represent a local correction too.
- The integrated Sachs-Wolfe effects (ISW) [69], which are the integrals along the geodesic of the time derivative of the potentials. They can explained similarly to the non-integrated case: along the geodesic, when a photon enters a potential well, the said potential evolves in time, and by the time the photon exits the well the energy needed to climb it may be different than the one gained while entering.
- The Shapiro time delay, which are the integrals along the geodesic of the Ψ potential and describe the delay caused by the space-time dilation in presence of a gravitational potential [70].

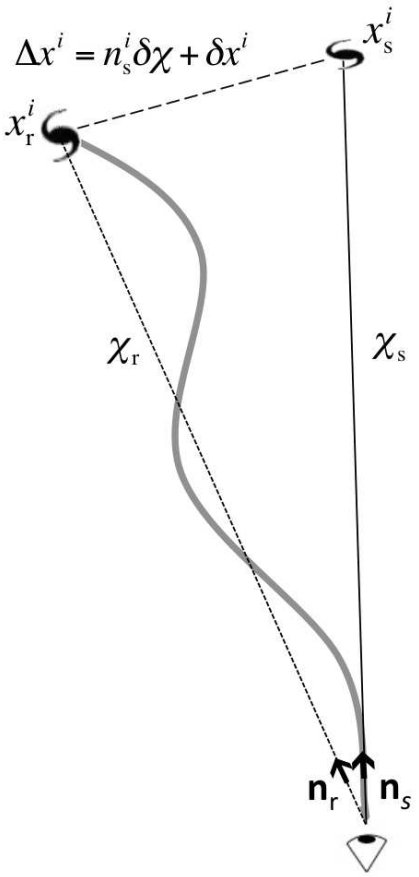


Figure 1.9: Real and redshift-space perspectives. A galaxy with real-space position x_r^i (top left), located at distance χ_r from the observer (bottom), is assigned an apparent position x_s^i in redshift space (top right) at distance χ_s. Since the photon path to the observer in real space is not straight, the observed position of the galaxy on the sky, $\mathbf{n}_s$ does not coincide with its actual one, $\mathbf{n}_r$. Picture and description taken from [17].

- The gravitational lensing effects, which are the integrals along the geodesic of the gradient of the potentials. They describe the deviation of the photon path due to the gravitational potential between the source and the observer [71].

1.4.10 Density perturbation

Now we want to see how the galaxy overdensity changes under the $x_r^\mu \to x_s^\mu$ mapping. The observed number of galaxies contained within a volume V^s defined in terms of the redshift space coordinates is given by a gauge-invariant integral [8–11]

$$N = \int_{V^s} \sqrt{-g}\, n_g(x^\alpha) \epsilon_{\mu\nu\rho\sigma} u^\mu(x^\alpha) \frac{\partial x^\nu}{\partial x^{s,1}} \frac{\partial x^\rho}{\partial x^{s,2}} \frac{\partial x^\sigma}{\partial x^{s,3}}\, d^3x^s\,, \tag{1.120}$$

where ϵ is the Levi-Civita tensor and n_g is the number density of galaxies in comoving coordinates, which depends on both position and conformal time. We can describe the same quantity in function of the redshift-space number density $n_{g,s}$, by writing:

$$N = \int_{V^s} a^3(z^s)\, n_{g,s}(\mathbf{x}^s, z^s)\, d^3x^s\,. \tag{1.121}$$

By comparing these two quantities we can retrieve the relation between the density in real and redshift space [9]. Up to linear order in the perturbations the galaxy density contrast is written as [8–11]

$$\delta_{g,ss} = \delta_{g,r} + \delta_g^{\mathrm{RSD}}\,, \tag{1.122}$$

where

$$\begin{aligned}\delta_g^{\mathrm{RSD}} = &\left(\frac{\partial_0\mathcal{H}}{\mathcal{H}^2 + \frac{2}{\chi_s\mathcal{H}}}\right)\delta\ln a + \Psi_e - 2\Phi_e + \frac{(\partial_0\Phi)_e}{\mathcal{H}} + 3\mathcal{H}\psi_v \\ &- \frac{1}{\mathcal{H}}\left[n_s^i\partial_i\left(n_s^j v_j\right)\right]_e + \frac{2}{\chi_s}\int_0^{\chi_s}(\Phi+\Psi)\,d\chi + 2\left(n_s^i v_i\right)_o - 2\kappa\end{aligned} \tag{1.123}$$

is the perturbation to the overdensity, and κ is the **convergence field**

$$\kappa = \frac{1}{2}\int_0^{\chi_s}(\chi_s - \chi)\frac{\chi}{\chi_s}\nabla_\perp^2(\Phi+\Psi)\,d\chi\,, \tag{1.124}$$

which is related to the weak gravitational lensing effect, as we will briefly show in 1.4.10. The operator $\nabla_\perp^2$ is defined as:

$$\nabla_\perp^2 = \nabla^2 - \left(n_s^i\partial_i\right)^2 - \frac{2}{\chi}n_s^i\partial_i\,. \tag{1.125}$$

The ψ_v term that appears in equation (1.123) is the velocity potential defined in equation (1.48) and appears because the derivation of the overdensity was done in the syncronous-comoving gauge, and then transformed in the conformal-Newtonian gauge.

We can also apply equation (1.122) for the dark matter component, and then, by following a biasing approach, it can be shown that [8–11]:

$$\delta_g^{ss} = (b-1)\,\delta^r + \delta^{ss}\,, \tag{1.126}$$

where b is the linear bias parameter and δ^r, δ^s are the matter overdensity fields in real and redshift-space. Note that there are further effects that need to be accounted for in this equation, related to the fact that in the case of selection effects, the number counts of the tracer are not conserved in the mapping from real to redshift space: this is explored further in section (1.4.12).

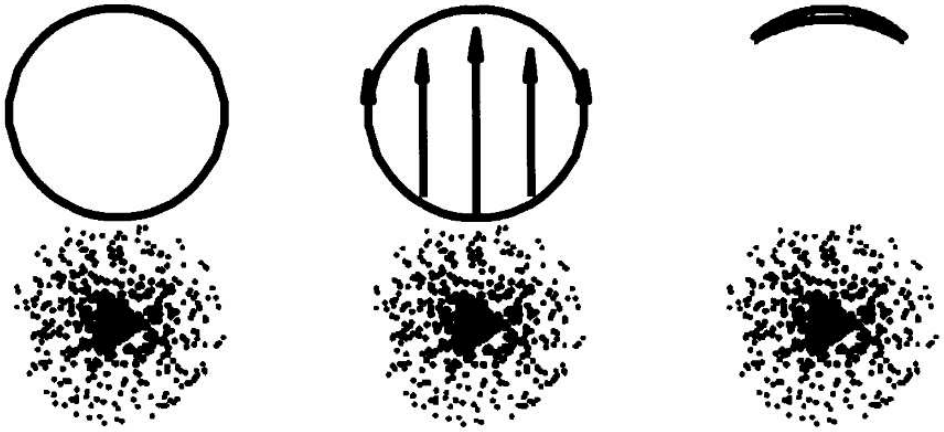


Figure 1.10: (a) Circular galaxy, the source, sits behind a foreground mass distribution represented by points at bottom. The observer is out of the page so that the foreground mass is between the observer and the source, (b) Light rays from source are deflected as they pass by mass distribution. Rays traveling closest to mass get deflected the most, (c) Resulting image is an arc. Picture and description taken from [30].

Weak lensing convergence

We will now investigate on the origin of the convergence term κ that contributes to redshift-space distortions and how it is linked to weak gravitational lensing, following the approach presented in [30]. Gravitational lensing distorts the galaxy shapes due to the foreground mass clustering, which bends the space curve the photons' trajectories. In figure (1.10) we see how a foreground mass distribution bends a circular image: the rays closer to the distribution will be bent more than the ones in the opposite direction causing a distortion in the circular shape, which will now appear as an arc. We use the term weak lensing to describe the situation in which the distortions are small and can be treated perturbatively.

To link these distortions to the underlying gravitational fields we study the photon geodesics at 1st order in perturbations. We describe the position of a photon reaching us from a galaxy with the set of spherical coordinates $\{\chi, \theta\}$, where θ represents the angular position in the sky. We will use the conformal Newtonian gauge as in equation (1.115) and we will neglect anisotropic stresses, so that $\Phi = \Psi$. Furthermore we will assume small angles θ^i, which will limit our study to small angular scales. The geodesic equation then reads [30]:

$$\frac{d^2}{d\chi^2}\left(\chi\theta^i\right) = 2\partial_i\Phi\,. \tag{1.127}$$

We can see that for a constant potential this equation implies that $d(\chi\theta^i)/d\chi =$ const., which means that the angular direction remains constant along the geodesic.

For $\partial_i \Phi \neq 0$ we notice that the equation states that the photon geodesic deviates towards the direction in which Φ grows, which means that light will be curved towards a matter overdensity. By integrating two times, and fixing the constant terms to account for boundary conditions on the geodesic, we get [30]

$$\theta_s^i = \theta_r^i + 2 \int_0^\chi \partial_i \Phi \left(\mathbf{x}(\chi')\right) \left(1 - \frac{\chi'}{\chi}\right) d\chi' \,. \tag{1.128}$$

We recognize this equation in the third term of the right-hand-side of equation (1.118), which is in fact, the lensing term of redshift-space distortions.

To study the magnification effects due to these distortions we define the transformation matrix

$$\begin{aligned} A_{i,j} \equiv & \frac{\partial \theta_s^i}{\partial \theta_r^j} \equiv \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \\ & = \delta_{i,j} + 2 \int_0^\chi \partial_{i,j} \Phi \left(\mathbf{x}(\chi')\right) \chi' \left(1 - \frac{\chi'}{\chi}\right) d\chi' \,, \end{aligned} \tag{1.129}$$

where κ in the already mentioned convergence, which describes how a source is magnified, while γ_1 and γ_2 are the two components of the shear, which describe the distortions in the shape of the galaxies. From equation (1.129) we retrieve

$$\begin{aligned} \kappa = & - \frac{A_{1,1} + A_{2,2}}{2} + 1 \\ & = - \int_0^\chi \left[\partial_1^2 \Phi \left(\mathbf{x}(\chi')\right) + \partial_2^2 \Phi \left(\mathbf{x}(\chi')\right)\right] \frac{\chi'}{\chi} \left(\chi - \chi'\right) d\chi' \,. \end{aligned} \tag{1.130}$$

Comparing this with the definition in equation (1.124) we notice that the differential operator is different: this is due to the fact that we limited ourselves in the study of small angular scales, where $\nabla_\perp^2 \approx \nabla^2$, while the full relativistic treatment of redshift-space distortions does not depend of such assumptions. We will see in the next subsection that the total magnification of a source takes contributions also from other effects beside lensing.

1.4.11 Magnification

Since redshift space distortions alter the distance of a galaxy, we have to treat carefully the flux we measure [72, 73]. Suppose we measure a flux F from a given galaxy at an observed redshift z_s: if the true redshift z_r is bigger than z_s, it means that the galaxy is actually brighter that we inferred, while if $z_r < z_s$ it is fainter.

In general, lensing effects can distort the solid angle of the galaxy we measure, changing the observed flux also if no radial shift were present. One example of this is the convergence term we mentioned in section (1.4.10).

In conclusion, velocity and general relativistic effects alter the apparent flux from a source. In a flux limited survey these effects can bring a galaxy above or below the flux cut, changing the measured clustering signal. We can model this in terms of the luminosity distance (see equation (1.9))

$$d_{\mathrm{L}}(z) = (1 + z)\chi(z) \,. \tag{1.131}$$

Relativistic effects can distort this quantity, with the consequence that within a flux-limited survey some sources may appear/disappear due to redshift-space distortions. We define the magnification as

$$\mathcal{M} \equiv \left(\frac{d_{\rm L}}{\bar{d}_{\rm L}(z_s)} \right)^2 , \tag{1.132}$$

where $\bar{d}_{\rm L}(z_s)$ is the luminosity distance we infer by assuming an unperturbed space and using the observed (distorted) redshift z_s, while $d_{\rm L}$ is the actual luminosity distance of the source.

From linear perturbation theory the resulting magnification is [8–11]

$$\begin{aligned} \mathcal{M} =& 1 - 2\left(1 - \frac{1}{\mathcal{H}\chi_s}\right)\left[\Psi_o - \left(n_o^i v_i\right)_o\right] \\ &+ 2\left(1 - \frac{1}{\mathcal{H}\chi_s}\right)\left[\int_0^{\chi_s} (\chi_s - \chi)\partial_0(\Phi + \Psi)]\, d\chi + \Psi_e - \left(n_s^i v_i\right)_e\right] \\ &+ 2\Psi_e - 2\left(n_s^i v_i\right)_o - \frac{2}{\chi_s}\int_0^{\chi_s} (\Phi + \Psi)\, d\chi + 2\kappa\,. \end{aligned} \tag{1.133}$$

We see that with the full general-relativistic treatment, further corrections beside the convergence term arise to the magnification of an image. We recognize the contribution from the peculiar velocities, the SW and ISW effects and the Shapiro time delay.

1.4.12 Population biases

In equation (1.120) we assumed the conservation of the number of galaxies in the mapping between real and redshift space. In a realistic survey there will be selection effects that will decrease the number of observed galaxies: for example this can be due to a lower flux limit $F_{\rm cut}$ under which the survey's instrument won't detect any galaxy or similarly due to a shape limit. Taking into account only flux selection effects we define $n_g(> L)$ as the number density of galaxies with luminosity greater than L, and we also define

$$L_{\rm min} = 4\pi d_{\rm L}^2(z) F_{\rm cut}\,, \tag{1.134}$$

which is the minimum luminosity that a galaxy at a cosmological redshift z needs to have to appear over the flux cut for an observer at $z = 0$.

Furthermore, the number of tracers above certain luminosity can change in redshift, both due to their formation/merging and of the evolution of their emitters.

These properties, when combined with the perturbations in redshift and magnification, can change the number counts of the tracer under study. Taking that into account we finally update equation (1.126)

$$\delta_{g,ss} = (b-1)\,\delta_r + \delta_s + \mathcal{Q}\,(\mathcal{M} - 1) + \mathcal{E}\,(\delta \ln a - \mathcal{H}\psi_v)\,, \tag{1.135}$$

where

$$\mathcal{Q} = -\frac{\partial \ln \bar{n}_g(> L)}{\partial \ln L}|_{L=L_{\rm lim}} \tag{1.136}$$

is the **magnification bias** and describes the change in tracers' density with the luminosity cut, while

$$\mathcal{E} = -\frac{\partial \ln \bar{n}_g(>L)}{\partial \ln(1+z)} \tag{1.137}$$

is the **evolution bias** and describes the change in number density with redshift, without accounting for the flux cut evolution. It can be shown that the relation between $\mathcal{Q}$ and $\mathcal{E}$ is [16, 65]

$$-\frac{\mathrm{d} \ln \bar{n}_g(z)}{\mathrm{d} \ln(1+z)} = 2\mathcal{Q}(z)\left[1 + \frac{c\,(1+z)}{H(z)\chi(z)}\right] + \mathcal{E}(z)\,, \tag{1.138}$$

where c is the speed of light and $\chi(z)$ is the comoving distance at redshift z. We see that the total derivative of the number counts in redshift depends on both the biases, since they capture different contributions to the number counts.

1.4.13 Observer effects

In this subsection we model the observer's contribution to the power spectrum. We will consider only the velocity term in this contribution, since it dominates over the potential term. The shift caused by the observer's velocity is [1, 63, 64]

$$\Delta\mathbf{x} = -\frac{1}{a(z)H(z)}(\mathbf{v}_{\rm obs}\cdot\hat{\mathbf{x}}_\mathbf{r})\,\hat{\mathbf{x}}_\mathbf{r}\,. \tag{1.139}$$

This term generates a dipole contribution in the redshift-space overdensity [16,63,64]

$$\delta_{\rm obs} = \delta_{\rm cmb} + \frac{\alpha_{\rm obs}[z(\chi)]}{\chi}\frac{\mathbf{v}_{\rm obs}\cdot\hat{\mathbf{x}}_\mathbf{r}}{a[z(\chi)]H[z(\chi)]}\,, \tag{1.140}$$

where $\delta_{\rm cmb}$ is the overdensity in the CMB frame of reference and we refer to the second term as $\delta_{\rm dip}$. The α_s function that appears is a generalization of the one presented in equation (1.92), that accounts for general relativistic effects [74]

$$\alpha_{\rm obs}(z) = 2(1-\mathcal{Q}) - \frac{\chi H}{c(1+z)}\mathcal{E} + \frac{\chi H}{c(1+z)}\left[3 - \frac{1+z}{H}\frac{dH}{dz}\right]\,, \tag{1.141}$$

in which we see a contribution from the evolution and magnification biases. We cross-correlate the dipole term with itself

$$D(\mathbf{k}) = \frac{\int A_1 A_2 (\mathbf{v}_{\rm obs}\cdot\mathbf{u}_{\mathbf{x}_1})(\mathbf{v}_{\rm obs}\cdot\mathbf{u}_{\mathbf{x}_1}) e^{i\mathbf{k}\cdot(\mathbf{x}_2-\mathbf{x}_1)}\, d^3x_1\, d^3x_2}{\int \bar{n}^2(x)\, d^3x}\,, \tag{1.142}$$

with $A_i = \bar{n}_i \alpha_{{\rm obs},i}/(x_i a_i H_i)$. The reason why we weight the overdensity with $\bar{n}$ in the integral is related with the FKP estimator, described on section (2.4). Expanding the angular part of the integral in spherical Bessel functions we obtain [16]

$$D(k, \hat{\mathbf{v}}_{\rm obs}\cdot\hat{\mathbf{k}}) = 16\pi^2 \frac{(\mathbf{v}_{\rm obs}\cdot\hat{\mathbf{k}})^2}{H_0^2}\frac{I^2(k)}{\int \bar{n}^2(x)\, d^3x}\,, \tag{1.143}$$

where

$$I(k) = \int \frac{x\bar{n}\alpha_{\rm obs}H_0}{aH} j_1(kx)\, dx\,, \tag{1.144}$$

and j_1 is the $n = 1$ spherical bessel function. Averaging over all the directions of $\mathbf{k}$ we get

$$D_{\rm iso}(k) = \frac{16\pi^2}{3}\frac{v_{\rm obs}^2}{H_0^2}\frac{I^2(k)}{\int \bar{n}^2(x)\, d^3x}\,, \tag{1.145}$$

which is a decaying oscillating function in k, as it will be shown in section (3.6).

We assume that the auto-correlation of $\delta_{\rm dip}$ dominates over the cross-correlation between $\delta_{\rm cmb}$ and $\delta_{\rm dip}$ (as it was tested in [63]), implying that the contribution of the observer's velocity on the power spectrum monopole is

$$\Delta P_0(k) = P_0^{\rm (obs)}(k) - P_0^{\rm (cmb)}(k) \approx D_{\rm iso}(k)\,. \tag{1.146}$$

In conclusion, the observer's effect on the monopole of the galaxy power spectrum follows a decaying oscillating function that depends on the selection function of the galaxy population, on their magnification and evolution biases, and on the expansion history of the Universe.

Chapter 2

Methods

In this chapter we will start by reviewing the current version of the LIGER code in section (2.1), used to generate mock galaxy catalogues that account for relativistic projection effects on the past lightcone of the observer. In sections (2.2,2.3) we will introduce the main topic of this project: the implementation of a new update on the LIGER code. We will see that, beside other minor improvements, this implementation creates catalogues that allow to probe smaller scales with respect to the current approach, and to avoid assuming a description *at priori* of the tracers under study. In general, the result of this method will be a more realistic catalogue, which allows to emulate better what a survey observes and how it selects the targets. Next, in section (2.4) we will give a brief overview of the power spectrum estimators for a galaxy dataset, which we will use in our analysis to test our implementation.

2.1 LIGER method

A mock galaxy catalogue that accounts for relativistic projection effects can be created in two ways: one way is to simulate structure formation using equations derived from general relativity, solving the geodesics equations and relativistic potentials [75, 76]; the other is to run a Newtonian cosmological N-body simulation and then apply the corrections *a posteriori*. The latter is the method implemented in the LIGER code [16, 17], currently available at `https://astro.uni-bonn.de/~porciani/LIGER/`, which:

- Takes as input a Newtonian N-body simulation.
- Selects an observer, characterized by its comoving position and velocity.
- Shifts all the galaxy positions in the simulation in redshift space, accounting for all the relativistic projection effects at 1st order in the perturbations, following the equations (1.116-1.119).
- Applies the correct magnification to each galaxy following equation (1.133).
- Gives as an output the distribution of the galaxies as they would be observed in the observer's perturbed past light-cone.

The code produces different light-cones that account for different contributions to the distortions described in equations (1.113-1.119, 1.133), labelled as:

- **realspace**: real space output, obtained by setting $\mathcal{M} = 1$ and $\Delta x = 0$.
- **vRSD**: obtained by setting $\mathcal{M} = 1$ and by considering only the $(n_s^i v_i)_e$ terms in the position shifts. This mock emulates the non-relativistic treatment of redshift space distortions presented in section (1.4.1).
- **vRSD_obs**: obtained by setting $\mathcal{M} = 1$ and by considering only the $(n_s^i v_i)_e$ and $(n_s^i v_i)_o$ terms in the position shifts. This mock is the extension of vRSD for the case of a moving observer.
- **GRRSD**: considers all the contributions beside $\Phi_o = \Psi_o = v_o^i = 0$, thus non accounting for the observer velocity and potential.
- **GRRSD_obs**: full redshift space output, all the contributions are considered.

In this section we will discuss the code implementation, i.e. how the equations that describe the shifts are applied on the Newtonian simulation. The product of a cosmological simulation is divided into snapshots of different comoving time instants, for which particle data such as positions and velocities is available. The code will use these snapshots as an input to compute both the underlying potential and the particle distribution.

2.1.1 Simulation Gauge

To evaluate the quantities in equations (1.116-1.119) we need to compute the underlying gravitational potential. Since the equations are applied to a simulation, we have to derive the potential from the particle distribution, which corresponds to using the matter density contrast in the synchronous comoving gauge, i.e. $\delta_{\rm sim} \equiv \delta_{\rm syn}$.

To linear order in the perturbations of a universe with a ΛCDM background, it can be shown that for a pressureless fluid, the source equation in the conformal-Newtonian gauge can be re-written in terms of $\delta_{\rm syn} = \delta_{\rm sim}$ as the standard Poisson equation (1.33) [18]

$$\nabla^2 \phi = 4\pi G a^2 \bar{\rho}_m \delta_{\rm sim}, \tag{2.1}$$

with the mapping

$$\phi = \Phi = \Psi, \qquad r^i = r^i_{\rm sim}, \qquad v^i = v^i_{\rm sim}. \tag{2.2}$$

This means that if we treat our corrections in the Newtonian gauge we can directly use the simulation positions and velocities, without having to recur to any gauge transformation.

2.1.2 Implementation

In this subsection we will discuss how the shifts of equations (1.116-1.119) are implemented and applied to the particles.

The particle velocity contributions can be applied in a straightforward way, since the particle data is already available. We shift each particle's radial position using the velocity evaluated at its snapshot's redshift. In a similar way, having chosen an observer, its velocity contributions are also easy to implement. The terms depending on the potential require more work due to the non-local quantities.

From equations (2.1) we see that to solve the Poisson equation with a numerical approach we need to compute the particle overdensity from the simulation data. We do so by implementing the particle-mesh with the cloud-in-cell method (CIC) [77] for a three-dimensional grid with N_{grid} cells per side. Then, by using a Fast Fourier Transform (FFT) library [78] we obtain the potential evaluated on the said grid, with its spatial derivatives. Currently the code utilizes the FFTW [79] library for all its Fourier transform-related computations. Finally, by using the data for the same cell at different snapshots, we can also compute its time derivative. The local potential terms are then retrieved by assigning to the particle the potential value of the cell in which resides, repeating the same with the observer.

To compute the integrals of the potential for the non-local contributions we need to integrate along the line of sight, which should be performed in redshift space. Since we are working at linear order, we can use the real-space coordinates instead, integrating from the observer's position to the particle's redshift space position. One problem related to this approach is that to get the redshift-space position of the particles we need also the non-local terms, which cannot be computed without knowing it at prior. Practically, the main contributions to the redshift space position are given by the velocity shift, so we limit ourselves to integrate up to the position obtained by applying just velocity contributions. To find the grid cells along crossed by the light ray, LIGER makes use of the fast voxel transversal algorithm by [80].

To avoid shifting all the galaxies at all times, which would only slow down the code, for each particle we select a few snapshots around the time in which it would intersect the lightcone of the observer, without considering any metric perturbation. Then, we compute the redshift displacement at each one of these selected snapshots. Finally, we use these shifted positions to interpolate the position of the particle around the crossing of the lightcone.

The magnification $\mathcal{M}$ and redshift perturbation $\delta \ln a$ are computed for each particle by implementing the equations (1.119,1.133), using the same procedures just presented to compute the integrated quantities.

In figure (2.1) we show a schematic of the shifting procedure.

2.1.3 Field approach

The transverse size of a lightcone rapidly increases with redshift, requiring large simulations to cover the wide opening angle. This creates a problem in building deep lightcones from high-resolution simulations, that would require drastically more computational time. To solve this problem the BUILDCONE code was implemented

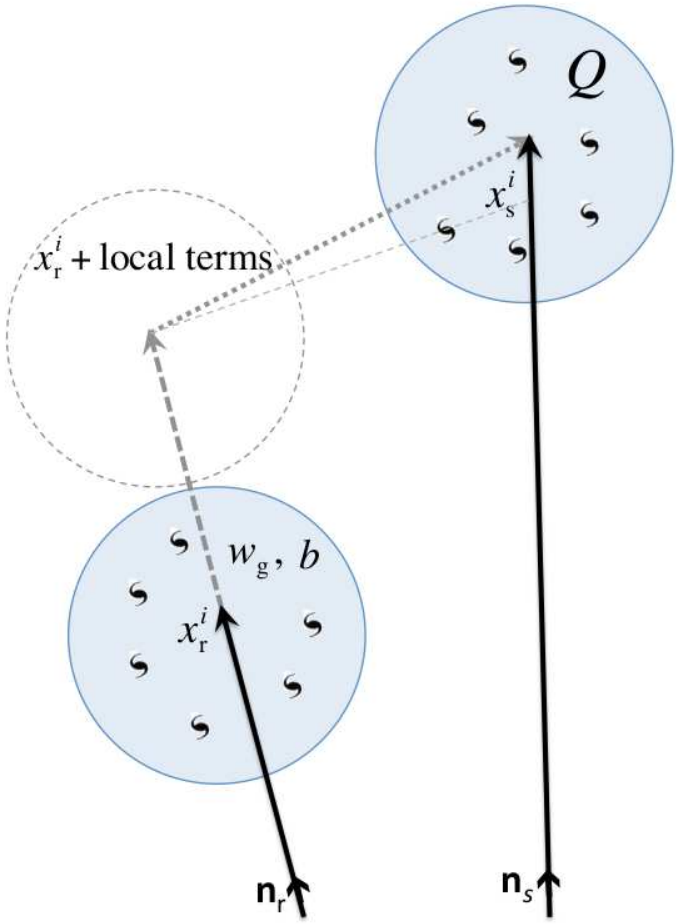


Figure 2.1: Schematic summarising how galaxies or N-body particles are shifted to build the lightcones. We first apply the correction due to local terms (dashed arrow) and then compute the shift produced by the non-local contributions (dotted arrow). In runs with low mass resolution, each particle 'contains' w_g galaxies. The linear bias coefficient b and w_g are calculated at the real-space position x_r^i while the magnification bias $\mathcal{Q}$ is computed at the redshift-space position x_s^i. Picture and description taken from [17].

in [16].

The code takes as an input the output of LIGER applied on a low-resolution dark matter only simulation, to then compute the matter density field in redshift space δ_s on a cubic grid via a CIC scheme. Similarly, it computes the $\mathcal{M}$ and $\delta\ln a$ fields on the same grid, using the same CIC scheme.

Finally, it implements the biasing relation in equation (1.135) to construct the tracer number density in each grid cell $\delta_{g,s}$. This method does not account for the velocity field ψ_v that multiplies the evolution bias, assuming to be working on scales where $|\mathcal{H}\psi_v| \ll |\delta \ln a|$, such that the implemented formula is

$$\delta_{g,ss} = (b-1)\,\delta_r + \delta_{ss} + \mathcal{Q}\,(\mathcal{M}-1) + \mathcal{E}\,\delta \ln a. \tag{2.3}$$

Additionally, one can estimate the effect of Poissonian shot-noise by applying a Poisson sampling on the number counts of tracers in each cell $N_g(\mathbf{x}) = \bar{n}_g(z)[1 + \delta_g(\mathbf{x})]V_{\rm cell}$, where $V_{\rm cell}$ is the cell's volume. This procedure ensures that, according to the number density, in each cell an integer number of tracers is present.

To use this method one needs to know the **survey functions** that describe the evolution in redshift of the tracer under study, i.e. the linear bias $b(z)$, the radial mean density $\bar{n}(z)$, and the magnification and evolution biases $\mathcal{Q}(z)$, $\mathcal{E}(z)$. As we will show later, these functions can be estimated directly from a simulated galaxy catalog, but also estimates from analytical models or observation are possible (see e.g. the work done in [16], starting from [81]).

The product of the code is then a density mesh of the tracer described by the survey functions, that includes all the relativistic projection effects at linear order.

Limits of the field approach

While it allows for a fast production of many galaxy catalogues, there are some limits in this approach.

First of all, we lose any clustering information regarding scales smaller than the cells of the mesh used by BUILDCONE. In many cases this is not a problem, since the code is used to investigate large scales effects, but the size of the cells we can use strongly depends on the characteristics of the survey function chosen. For example, suppose we want to apply the biasing procedure in real-space, so that we account only for the density $\bar{n}_g$ and linear bias b. Then, the painted density in a given cell would be

$$n_g = \bar{n}_g\,(1 + b\,\delta), \tag{2.4}$$

where δ is the underlying matter overdensity field. Suppose now to have a (realistic) linear bias value of $b = 2$: if in an under-dense region we have $\delta < -0.5$ we would get $n_g < 0$, which is not physical. We are thus forced to consider scales for which $b\,\delta \geq -1$, which puts a limit in the maximum number of cells per side $N_{\rm bc}$ that we can use. This can cause problems when computing Fourier-based statistics of highly biased tracers, where spurious effects may arise at larger scales due to the coarse meshing of the density

Second, the way we add shot-noise to the produced sample is not realistic and can generate a spurious signal as we probe smaller scales, as we will see in our analysis.

Third, with this procedure we can not model complex selection effects such as shape selection or more complicated flux cuts such as the one described in [82].

Last, we need to input the survey functions that describe the tracer, which means knowing at prior the tracer linear bias $b(z)$ and luminosity function $\phi(L, z)$.

2.2 Liger direct approach

In this work, I adapt LIGER to work directly on the tracers in a simulation, i.e. it shifts them in redshift space without recurring to the biasing relations as in the field approach does. We can use this version on an Newtonian N-body simulation with a high enough resolution such that it is possible to identify haloes or galaxies and for which the underlying gravitational potential is available.

One of the main differences in the implementation of this approach is to take into account the "broken" trajectories of the tracers and their mergers. We consider as an example the case of a halo: it will be formed at a given snapshot, possibly different than the earliest one given to the code, and will possibly merge (or be disrupted) at a snapshot different than the last one, slightly complicating the trajectory's light-cone interpolation. This was not an issue in the field approach because all dark matter particles exist at any given snapshot of the simulation.

This problem is approached by making use of the merger tree to account correctly for the halo (or galaxy) trajectories. As for now, to allow compatibility with different halo-finders and semi-analytical galaxy codes, we implement a pre-processing to the catalogues, to write them in a form which can be easily treated by the code. Details regarding this procedure will be explained in section (2.2.1).

The code then produces the tracer lightcone directly, by returning also the observed fluxes in case of a galaxy. With this approach, by making use of the magnification of each tracer and its luminosity, the evolution and magnification biases can already accounted for by applying a flux cut in post-processing, following what happens with a real survey. We will show that the code correctly captures these effects in the results.

Benefits of the direct approach

There are several benefits in the implementation of a direct approach. First of all, we are not limited by the meshing procedure of the field approach: we can build a mesh with a number of cells per side that fits our needs to avoid aliasing effects, or to finely sample the boundary of the surveyed region. Second, selection effects can be applied directly on the tracers, without having to recur to biasing models, allowing to consider also more realistic and complicated selection criteria . Third, we are not limited by the size of the bias of the tracer under study, since we do not recur in the "painting" procedure that limits our mesh size. Fourth, the Poisson shot noise is already captured by the catalogue and does not have to be added artificially on the density field. Last, we do not have to model at priori the linear bias and luminosity

function of the tracer, because their effects are already encoded in the simulation's structure formation history and semi-analytical code that produces the catalogues.

2.2.1 Trajectory definition

Different halo finders (e.g. ROCKSTAR [83], AHF [84]) produce different catalogues, with merger trees that use different formalisms. In order to allow compatibility with any halo finder, LIGER takes as input a pre-processed catalogue, in which we define the trajectories in a certain way. To do so we make use of two common halo catalogue entries: "ismmp" and "descID" (or analogue acronyms), which state if the halo is a *most massive progenitor* (MMP) and contain the ID of its descendant (for more details, see e.g. [83]). The procedure we follow for a halo catalogue is the following:

- For each halo we build a main trajectory considering the most massive progenitor (MMP), and we assign a unique ID to the halo for all the snapshots.
- At the end of each MMP "life", we check if the halo is still a MMP: in the negative case, we still check if it merges with another halo in the next snapshot, saving the descendant's ID.
- After the trajectories and mergers are defined, we order the haloes based on their ID. This procedure greatly speeds up the procedure of looking for a halo in each snapshot in the loading phase of LIGER.

This new catalogue is then used by the code to properly define the tracer trajectories and mergers, needed for a correct interpolation of the positions (and other quantities, such as masses and luminosities) on the intersection with the lightcone.

This process we showed was dedicated for the building of trajectories of a halo catalogue, but in a possible future usage on a semi-analytic galaxy catalogue (e.g. the Galacticus catalogue at [85, 86]) a similar procedure could be followed, e.g. by making use of the "hostID" entry that links a galaxy to the host halo ID of the associated halo catalog.

2.2.2 Magnification of sources

Treating directly the tracers allows to apply magnification and selection effects in a more realistic way. While we apply an eventual flux cut in a post-processing phase, we implement magnification directly in the code.

As we mentioned, the computation of the magnification $\mathcal{M}$ and redshift distortion $\mathrm{d}\ln a$ of each particle were already implemented in the code, they just were not used directly on the particles, but instead they were used to build a field. With this approach we use these already available quantities to compute the observed flux of a given input tracer, making use of its position on the lightcone and input luminosity. In order, for a given tracer we:

- Compute the luminosity $L(z)$ interpolated on the lightcone using the unshifted realspace positions to get the crossing redshift z.

- Compute the observed redshift z_s by making use of the realspace redshift z and of the $\delta \ln a$ via $z_s = z - (1+z)\,\delta \ln a$.
- Retrieve the inferred unperturbed luminosity distance $\bar{d}_{\rm L}(z_s)$, using the background relation of equation (1.9) and the observed redshift z_s. This is the luminosity distance infer if we do not account for magnification effects.
- Get the correctly magnified observed flux via $F_s = \mathcal{M}\, L / \left[4\pi \bar{d}_L(z_s)^2\right]$.

There are some details regarding this approach that are worth mentioning. First, we stress that we used the *realspace* positions to interpolate the luminosity, not the redshift space one. This is correct, since redshift space distortions do not interfere with the time t at which a source crosses the lightcone, but only deviate the path that the light emitted at that instant t follows.

Second, we also stress that we need to strictly follow this procedure to correctly account for every contribution: one might think that using directly the luminosity distance evaluated at the realspace redshift z, without then multiplying by $\mathcal{M}$, would be enough to account for magnification, but this is not true. Besides the magnification sourced by the redshift distortion $\delta \ln a$, we also have to account for other terms, such as the convergence magnification, which would act on the observed flux even without redshift space distortions.

2.3 Other updates

In this section we list further minor updates done to the code during this work.

2.3.1 Potential optimization

A further modification to the code was done to optimize the storing of the potential during a run. Before the update, the code used to keep loaded in the run memory the full cubic mesh of the potential for each snapshot, and then to select the cells that the photon path crossed to compute the integrals. Since we consider only particles close to the snapshots around the past lightcone of the observer, we only need to know the potential in the cells near that region.

With this update we discard the potential data from all the cells that are far from the lightcone intersection, reducing noticeably the memory consumption, but without losing any information that is going to get actually used. We show in figure (2.2) the heap memory consumption of a test run with a potential mesh of $N_{\rm grid} = 2^8$ cells per side and 16 snapshots. The increasing oscillating pattern at the first part of the run is caused by the potential computation phase, in which the code loads all the dark matter particles from one snapshot, it computes the potential and it frees the particle data. This is the same procedure for both the runs, and takes up the same memory. In the case in which the potential is already computed and stored somewhere, the code will load that directly, without looking for the dark matter particles, and thus the oscillating feature will not appear.

After the potential loading phase we notice that the heap memory of the optimized run drops: this happens because it frees the unused cells, which does not

happen in the standard procedure. With these settings the optimization saves around 0.8 GiB, over the 2.4 GiB used by the potential storage, keeping only 2/3 of the total memory. This value can change based on the total number of snapshots and on their "coarseness" in the time sampling, but in general allows to save a substantial fraction of memory, making the code less computationally costing.

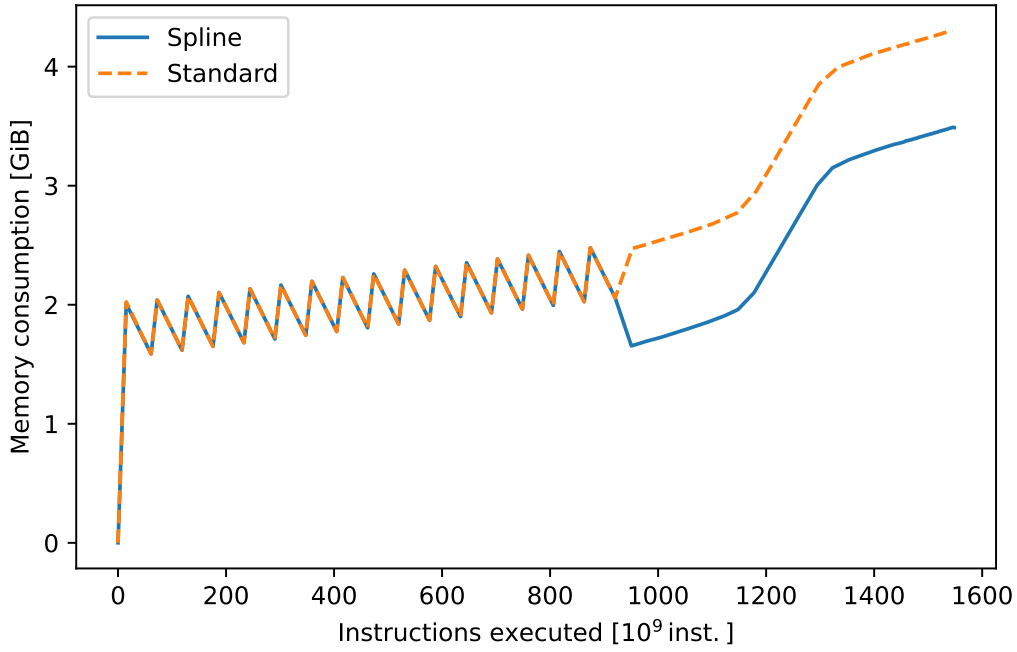


Figure 2.2: LIGER heap memory consumption comparison. The x-axis represents the number of instructions executed by the code. Orange dashed line: memory consumption of a test run without the potential optimization. Blue line: memory consumption of a test run with the potential optimization.

2.3.2 Non-fullsky lightcones

As we already mentioned, one of the main issues regarding the construction of lightcones from a cosmological simulation is the fact that, the deeper the redshift we want to reach, the bigger the simulation box must be. For example, if we want to build a fullsky lightcone up to a redshift z, the box size must be at least $2 \times d_c(z)$.

Up until now LIGER has worked by producing a fullsky lightcone centered on the input position of the observer. It makes use of periodic boundary positions to shift the box center to the lightcone center, so that the cone region would not intersect the box edges. More specifically, for an observer at a x coordinate $x_{\rm obs}$ in the box, and a particle at x, the code would map the particle to

$$x' = \begin{cases} x & |x - x_{\rm obs}| \leq \frac{L}{2} \\ x - L & |x - x_{\rm obs}| > \frac{L}{2} \end{cases} , \tag{2.5}$$

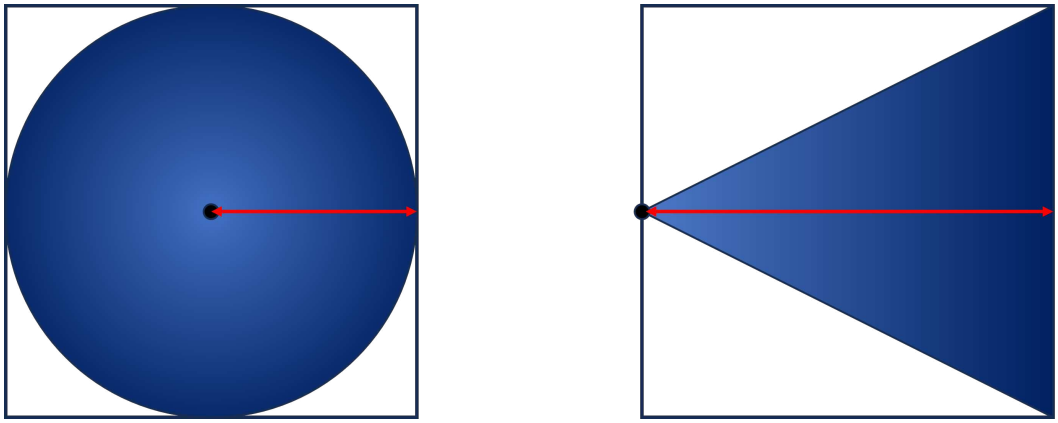


Figure 2.3: Sketch that shows how from the same simulation box we can reach larger distances using a non-fullsky lightcone shape.

where L is the simulation box size. With this procedure the lightcone produced can reach redshifts up to $z_{\rm lim}$, such that $d_c(z_{\rm lim}) < L/2$.

In this work we implemented a new option, which avoids using periodic conditions and simply keeps the coordinate as the same. This means that if we position the observer at the center of one side of the box we would be able to study a conical sky region with an opening angle $\alpha \in [0, \pi/2]$, but with a depth that, depending on α could reach redshifts up to $z_{\rm max}$, such that $d_c(z_{\rm max}) < L$. This set up is sketched in figure (2.3).

2.4 Power spectrum estimation

One way to study the effects of redshift space distortions on the galaxy distribution is to look at summary statistics such as the power spectrum monopole. In this section we will give a brief overwiev on how to estimate it from a simulation snapshot or a galaxy catalogue. As we showed in section (1.3.3), the operative definition of the power spectrum monopole and of the traditional power spectrum coincide, so in this section we will refer both of them as "power spectrum".

To compute the power spectrum from a set of particles we make use of the discrete Fourier transform (DFT): this means that from a particle distribution we need to estimate the density field $n(\mathbf{x})$ on a 3-d mesh, in order to correctly follow the procedure. We thus implement an interpolation scheme to go from the true particle number density in a simulation

$$n_0(\mathbf{x}) = \sum_{i=1}^{N_p} \delta^D(\mathbf{x} - \mathbf{x}_i), \tag{2.6}$$

where $\mathbf{x}_i$ is the i-th particle position, to the one computed on a mesh. The interpolation scheme is determined by the **mass assignment scheme** (MAS), which defines how a particle will contribute to the density of the surrounding grid cells. In this

work we utilize the *Cloud in Cell* (CIC) [78] method, in which the particle's mass is distributed uniformly in a cubic shape of the same size as the grid size, but centered at its position. With this choice mass assignment the interpolated density evaluated on a grid point centered in $\mathbf{x}_p$ is expressed in terms of a convolution

$$n_{\mathrm{CIC}}(\mathbf{x}_p) = \int_V d^3x' n_0(\mathbf{x}') W_{\mathrm{CIC}}(\mathbf{x}_p - \mathbf{x}'), \tag{2.7}$$

where the convolving function $W_{\mathrm{CIC}}(\mathbf{x})$ is called window function, and quantifies how much a particle contributes to the density in a grid point separated by $\mathbf{x}$. The one dimensional CIC window function is given by

$$W_{\mathrm{CIC}}(x) = \begin{cases} 1 - |x|/H & \text{if} |x| \leq H \\ 0 & \text{if} |x| > H \end{cases}, \tag{2.8}$$

where H is the grid size. The three dimensional window function is simply given by the product of three one dimensional window functions

$$W_{\mathrm{CIC}}(\mathbf{x}) = W_{\mathrm{CIC}}(x_1) W_{\mathrm{CIC}}(x_2) W_{\mathrm{CIC}}(x_3). \tag{2.9}$$

From the interpolated density we can then compute the interpolated overdensity as

$$\delta_{\mathrm{CIC}}(\mathbf{x}_p) = \frac{n_{\mathrm{CIC}}(\mathbf{x}_p) - \bar{n}(\mathbf{x}_p)}{\bar{n}(\mathbf{x}_p)}, \tag{2.10}$$

where $\bar{n}$ is the background number density, which for example can be either estimated from the whole grid in case of a simulation snapshot, or from a series of radial bins as it is in the case of data from a lightcone.

Having defined a method to interpolate the particle density on the grid, we can compute the Fourier transform, and using the convolution theorem we see that

$$\tilde{\delta}_{\mathrm{CIC}}(\mathbf{k}) = \tilde{\delta}_0(\mathbf{k}) \tilde{W}_{\mathrm{CIC}}(\mathbf{k}), \tag{2.11}$$

i.e. the actual transform of the particle overdensity is obtained by *deconvolving* the interpolated overdensity field

$$\tilde{\delta}_0(\mathbf{k}) = \frac{\tilde{\delta}_{\mathrm{CIC}}(\mathbf{k})}{\tilde{W}_{\mathrm{CIC}}(\mathbf{k})}. \tag{2.12}$$

The transform of the CIC window function is given by [87]

$$\tilde{W}(\mathbf{k}) = \left[\mathrm{sinc}\left(\frac{\pi k_1}{2k_N}\right) \mathrm{sinc}\left(\frac{\pi k_2}{2k_N}\right) \mathrm{sinc}\left(\frac{\pi k_3}{2k_N}\right) \right]^2, \tag{2.13}$$

where k_i the components of the wave vector $\mathbf{k}$ and $k_N = \pi/H$ is called the **Nyquist frequency**.

Now that we know how to estimate the density field on a cubic grid we can compute the power spectrum, and to do so we use the FFTW library [79] to calculate

the DFT of the overdensity field $\delta^{\mathrm{FFT}}(\mathbf{n}_k)$. Here, the discrete array $\mathbf{n}_k$ describes the cell position on the grid and it is associated to a wave mode with $\mathbf{k} = k_F \mathbf{n}_k$, where $k_F = 2\pi/L$ is the **fundamental frequency**. k_F, given a box size L, defines the smallest frequency interval that can be sampled by the FFT. Taking into account the normalization introduced by the algorithm and the deconvolution of the window function, the resulting estimate of the power spectrum will be evaluated on a series of k-bins multiples of the fundamental frequency [88]:

$$\hat{P}_{\mathrm{n}}(k_F n_1) = \frac{L^3}{N^6} \left\langle \left| \frac{\tilde{\delta}^{\mathrm{FFT}}}{\tilde{W}} \right|^2 \right\rangle = \frac{L^3}{N^6} \left(\frac{1}{N_k} \sum_{|n_k - n_1| < 1/2} \left| \frac{\tilde{\delta}^{\mathrm{FFT}}(\mathbf{n}_k)}{\tilde{W}(k_F \mathbf{n}_k)} \right|^2 \right), \tag{2.14}$$

where n_1 in a discrete number that defines the k-binning, and the average is computed over all the grid cells associated to frequencies belonging to the n_1-th bin. The choice of binning is not limited to the fundamental frequency, and one is able to use a finer or coarser binning, depending on the purpose of use.

As we stated in section (1.3.2), when we estimate the 2-point correlation function from a discrete set of objects, we account also for the self-correlation of the particles, which generates an additive term $+\delta^D(\mathbf{r} - \mathbf{r}')/\bar{n}(\mathbf{r})$, called **Poisson shot-noise**, singular for $\mathbf{r} = \mathbf{r}'$, that needs to be accounted for when moving to Fourier space to get the power spectrum

We consider now the case of a snapshot from an N-body simulation. In this situation $\bar{n}$ is a constant, and the shot noise translates into a contribute to the power spectrum equal to

$$\hat{P}_{\mathrm{SN}} = \frac{1}{\bar{n}}. \tag{2.15}$$

The correct estimation of the power spectrum would then be

$$\hat{P}(k_F n_1) = \hat{P}_{\mathrm{n}}(k_F n_1) - \hat{P}_{\mathrm{SN}}. \tag{2.16}$$

We can also compute the error on the estimator of $\hat{P}$. If we assume that the overdensity field is a Gaussian random variable, which as mentioned in section (1.3.4) is a fair assumption at large scales, it can be shown that the standard deviation of the estimation is given by [88]

$$\sigma_P(k) \equiv \langle [\hat{P}(k) - P(k)]^2 \rangle^{1/2} = \sqrt{\frac{1}{N_k}} P(k), \tag{2.17}$$

where $P(k)$ is the true power spectrum and N_k is the number of modes in the k-bin. We notice how the error decreases with N_k, the number of modes over which we compute the average: as a result, our estimates for the largest scales, the ones comparable to the grid box size L, will be less precise, due to the smaller number of **k**-modes we can average over. This effect is known as **sample variance**, and it limits the information we can extract from a single sample of data.

2.4.1 FKP estimator

As we mentioned, LIGER reproduces the galaxy distribution that a redshift survey would observe, and this allows us to study its output in the same way we would do with a real survey dataset.

When considering a survey's data, differently from a simulation snapshot, the galaxy mean number density $\bar{n}(\mathbf{x})$ will not be constant, but a function of the position. This is due to multiple reasons. First, the mean number density of a given tracer evolves with time, and thus with the distance from the observer, causing a dependence of $\bar{n}$ on the radius $|\mathbf{x}|$. Second, a redshift survey could be considering only a fraction of the sky, for example in a pencil-shaped beam or in a full-sky survey that excludes the galactic plane, and this introduces a dependence of $\bar{n}$ on the position in the light-cone $\mathbf{x}$.

The spatial dependence of the mean density contributes to the variance of the power spectrum estimator [89]. For this reason it is common practice to weight galaxies differently depending of the mean number density at their location, in order to give an *optimal* estimation of the power spectrum $\hat{P}$, i.e. minimize its variance.

In this work, we use the FKP estimator [89, 90], which provides an optimal weighting scheme for estimating the power spectrum $P(k)$ under the assumptions that:

- The wavelenght $2\pi/k$ is small compared to the scale of the survey.
- The density fluctuations are Gaussian.

In this framework, instead of working with the overdensity we will work with the effective field

$$F(\mathbf{x}) = \frac{I_{\mathcal{S}}(\mathbf{x})\, w(\mathbf{x})\, [n_g(\mathbf{x}) - \hat{n}(\mathbf{x})]}{\left[\int_{\mathcal{S}} w^2(\mathbf{x}) \hat{n}^2(\mathbf{x})\, d^3x\right]^{1/2}}, \tag{2.18}$$

where $I_{\mathcal{S}}$ is the indicator function of the region of space $\mathcal{S}$ covered by the survey, $\hat{n}(\mathbf{x})$ is an estimate of the mean galaxy density based on the survey data, and $w(\mathbf{x})$ is a weighting function which we will specify later. We can reshape this in terms of the estimated density contrast $\hat{\delta}(\mathbf{x}) = [n_g(\mathbf{x}) - \hat{n}(\mathbf{x})]/\hat{n}(\mathbf{x})$ and the window function

$$W(\mathbf{x}) = \frac{I_{\mathcal{S}}(\mathbf{x}) w(\mathbf{x}) \hat{n}(\mathbf{x})}{\left[\int_S w^2(\mathbf{x}) \hat{n}^2(\mathbf{x})\, d^3x\right]^{1/2}}, \tag{2.19}$$

by writing

$$F(\mathbf{x}) = W(\mathbf{x}) \hat{\delta}(\mathbf{x}). \tag{2.20}$$

The weighting function introduced in the FKP paper is

$$w(\mathbf{x}) = \frac{1}{1 + \hat{n}(\mathbf{x}) \mathcal{P}}, \tag{2.21}$$

where $\mathcal{P}$ is the typical value of the power spectrum in the scales we are interested in. At a fixed k, in regions where $\hat{n}\mathcal{P} \gg 1$, i.e. where we are not limited by shot-noise, w

provides equal weighting per volume, while in regions in which shot noise dominates ($\hat{n}\mathcal{P} \ll 1$) it provides equal weighting per galaxy.

To evaluate the galaxy power spectrum monopole via FTT, we follow the procedure shown at the beginning of this section in equation (2.14), with a few changes. First, we use the weighted field F instead of the overdensity δ_g, and this, due to the W normalization, introduces a multiplying renormalization factor on the power spectrum monopole of $L^3_{\rm FFT}$, where $L_{\rm FFT}$ is the FFT box size. Second, we estimate the radial density $\hat{n}(\mathbf{x})$ by measuring the mean density of galaxies in radial shells with a fixed width Δx and by interpolating the results with a cubic spline. Third, due to the weighting and to the non-constant radial density, we will use the correct shot noise estimation for this implementation, given by [89]

$$P_{\rm SN} = \frac{\int \hat{n}(\mathbf{x})\, w^2(\mathbf{x})\, d^3x}{\int \hat{n}^2(\mathbf{x})\, w^2(\mathbf{x})\, d^3x}, \tag{2.22}$$

where we estimate the integrals with discrete sums over the grid cells. Furthermore, it can be shown that using a selection function $\hat{n}(\mathbf{x})$ estimated from the data itself introduces another contribution, called **integral constraint** [91], that accounts for the fact that this estimation does not coincide with the actual radial density $\hat{n} \neq \bar{n}$. This contribution affects density fluctuations on scales comparable to and larger than the survey size, leading to a biased estimate of the power spectrum monopole. Since in this work we are interested in a comparison between two catalogs, and not with a theoretical prediction, we avoid to correct for this effect, given that it would raise the same contribution in both cases anyway.

We will now look at the effects of the FKP weighting on the actual power spectrum of the overdensity. We will do that in a simplified case, where the translation symmetry of the density field is preserved, such as in the plane-parallel approximation. We consider the power spectra of the continuous underlying density fields $\delta(\mathbf{x})$ and $F(\mathbf{x})$, which we call respectively $P(\mathbf{k})$ and $\tilde{P}(\mathbf{k})$. Then, it can be shown that they are linked via [16, 88]

$$\tilde{P}(\mathbf{k}) = \int \frac{d^3p}{(2\pi)^3} |\tilde{W}(\mathbf{q}-\mathbf{p})|^2 P(\mathbf{p}). \tag{2.23}$$

From equation (2.23) we can see that the weighting scheme acts on the power spectrum as a convolution with the square of the transform of the window function. This convolution operates as a weighted average that mixes the Fourier modes. For a typical survey $|\tilde{W}(\mathbf{k})|^2$ mixes the large scales modes together, smoothing the signal. This smoothing can be an issue when trying to study large scale effects, which would be smoothed out and more difficult to detect.

Under the assumptions of Gaussian density fluctuations and a wide k-binning Δk (see e.g. [89, 92]), the covariance matrix of the FKP estimator monopole is diagonal with variance

$$\frac{\sigma^2_{\rm G}(k)}{P_0^2(k)} \approx \frac{2(2\pi)^3}{V_{\rm eff}(k) V_k}, \tag{2.24}$$

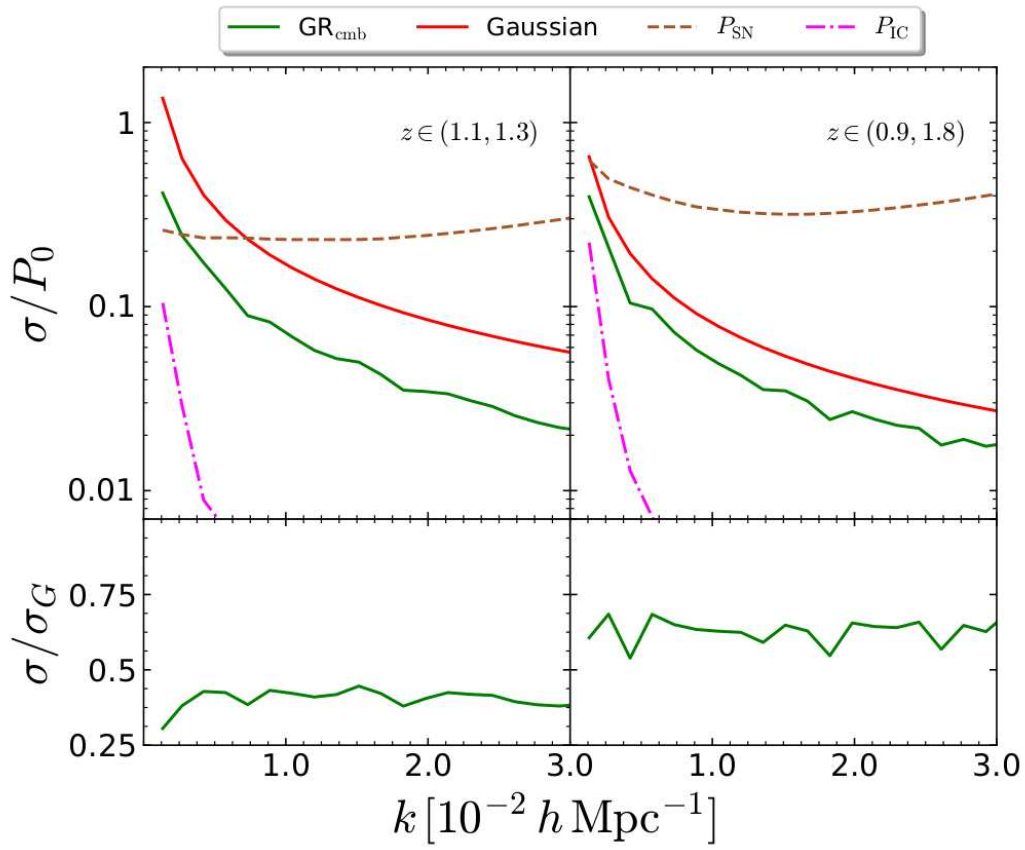


Figure 2.4: Top: True and gaussian errors (in the green and red lines) compared with the shot noise (brown line) and integral constraint (purple line). Bottom: Ratio of the scatter over the Gaussian assumption. Picture taken from [16].

where $P_0(k)$ is the monopole of the true unconvoluted power spectrum $P(\mathbf{k})$ and

$$V_{\rm eff}(k) = \int_{\mathbf{S}} \left[\frac{\bar{n}(\mathbf{x})\, P_0(k)}{1 + \bar{n}(\mathbf{x})\, P_0(k)} \right]^2 d^3x, \tag{2.25}$$

denotes the effective volume probed, while V_k denotes the volume of the k-shell centered around k with witdh Δk.

In [16] it was shown that the approximation for the FKP variance of equation (2.24) is not accurate at large scales when compared to the actual scatter, but it overestimates it instead. This is due to the fact that the equation is obtained by assuming Gaussian density fluctuations, which due to the FKP weighting is not necessarily true for $F(\mathbf{r})$. We show in figure (2.4) the comparison of the estimator with the actual scatter computed in [16]. We can see how the true scatter σ (obtained with Liger) is just a fraction of the Gaussian prediction.

Chapter 3

Results

In this chapter we show the results of our analysis. We start in section (3.1) by describing schematically the code produced and its main differences with the previous implementation. In section (3.2) we show how we use the HugeMDPL simulation [85, 86] to test the new code, by comparing it with the previous version. In section (3.3) we will describe instead the MUSIC [66] runs we use to generate multiple mocks of the previous implementation. Then, in section (3.4) we describe the methods implemented to estimate the tracer's survey functions from the dataset.

To test our implementation we produce tracer catalogues in two different redshift bins, a close one in $z \in [0.2, 0.3]$ and a deeper one in $z \in [0.6, 0.7]$. We compare the power spectrum monopole in both bins, for each one of the LIGER output modes listed in section (2.1). The results are shown in section (3.5).

We do one last test in section (3.6) for the $z \in [0.2, 0.3]$ bin, where we study the observer's velocity contribution to the power spectrum monopole, comparing it with the theoretical prediction, to further validate the new code.

3.1 Code update

Here we present schematically the code produced in this work, for which a detailed explanation is given in sections (2.1 and 2.3). We show in figure (3.1) a schematic of the prescription of the new direct implementation (right) compared to the old field-based one (left). Both the prescriptions use the same coordinate transform described by equations (1.116-1.119), and the main difference in the approaches lies in the input particle to which the transform is applied and in the final product: the field approach starts from a low resolution N-body simulation that does not resolve structure formation, and produces a cubic grid of the tracer density field. The direct approach instead starts from an N-body simulation that resolves structure formation and an associated galaxy/halo catalogue, to then produce a catalogue of the same tracers as they would be observed by a redshift survey. We list the main benefits of the implemented direct approach:

- We are not limited by the meshing procedure of the field approach: we can build a mesh with a number of cells per side that fits our needs to avoid

aliasing effects, or to finely sample the boundaries of the surveyed region.

- Selection effects can be applied directly on the tracers, without having to recur to biasing models, allowing to consider also more realistic and complicated selection criteria.
- We are not limited by the size of the bias of the tracer under study, since we don not recur in the "painting" procedure that limits out mesh size.
- The Poisson shot noise is already captured by the catalogue and doesn't have to be added artificially.
- We don not have to model at priori the linear bias and luminosity function of the tracer, because their effects are already encoded in the simulation's structure formation history and semi-analytical code that produces the catalogues.
- Allows the possibility of building cone-shaped lightcones, which allow to reach up to double the distance of a full sky survey for a given simulation box size, probing higher redshifts.

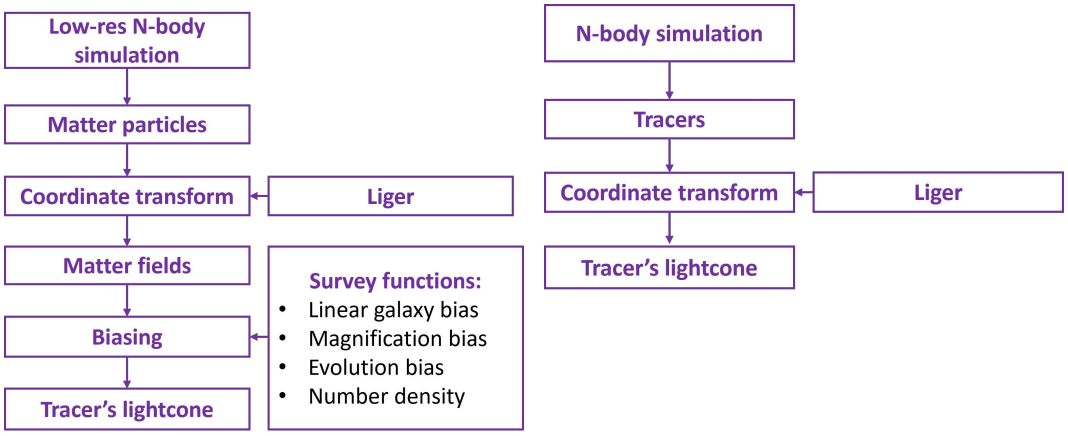


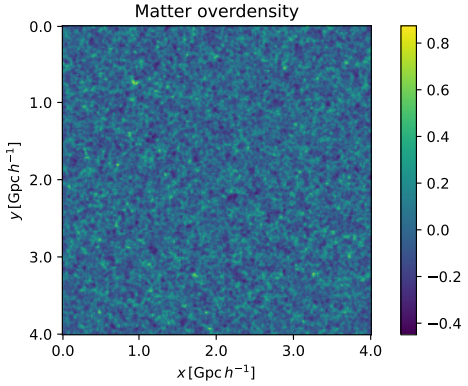
Figure 3.1: Left: Field approach schematic of LIGER. Right: Direct approach implemented in this work.

3.2 HugeMDPL simulation

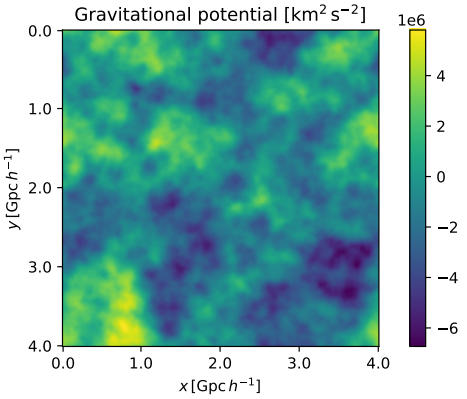
In order to compare the two LIGER implementations, we need to use a cosmological N-body simulation with a high enough resolution to follow galaxy formation (needed for the direct approach), and with a large enough volume to probe the largest scales (needed for the field approach, and in general to observe large-scale effects). There are not many available runs with these characteristics, since the computational cost of high-resolution simulations increases greatly with the simulated volume.

We opt for the HugeMDPL simulation [85, 86], a dark matter-only simulation of size $4\,\mathrm{Gpc}/h$ with 4096^3 particles of mass $m = 7.9 \times 10^{10} M_\odot/h$ and a Planck 2014

cosmology [93] with $(\Omega_0, \Omega_b, \Omega_\Lambda, \sigma_8, n_s, h) = (0.307, 0.048, 0.693, 0.829, 0.96, 0.678)$. This simulation data is hosted on the CosmoSim [85] database, and it has two available "raw" particle snapshots at the redshifts $z = 0.49$ and $z = 0.00$, and ROCKSTAR [83] halo catalogues for a series of snapshots, roughly uniformly distributed in values of the scale factor a from $a_i = 0.10$ to $a_f = 1.00$, meaning from $z_i = 8.58$ to $z_f = 0.00$. In figure (3.2a) we show a slice of the matter overdensity field, obtained from the particle data implementing the CIC scheme in a cubic grid with $N_{\text{grid}} = 2^8$ cells per side, with size $L_{\text{grid}} = 15.625\,\text{Mpc}/h$. In figure (3.2b) we show a slice of the large scale potential field generated from the particle distribution, obtained by solving the Poisson equation in Fourier space. Finally, in figure (3.3) we show the matter power spectrum compared with the linear theory prediction obtained with the CAMB code [94]. The error bars are estimated by equation (2.17) and we can see that the data correctly follows the shape of the linear prediction up to the smallest scales shown.



(a) Slice of the matter overdensity field computed from the $z = 0$ snapshot of the HugeMDPL simulation. The field was interpolated on the grid via the cloud-in-cell mass assignment scheme.



(b) Slice of the gravitational potential field computed from the $z = 0$ snapshot of the HugeMDPL simulation. The field was obtained by numerically solving in Fourier space the Poisson equation, starting from the overdensity field.

Figure 3.2: Slice of the matter overdensity field and the of gravitational potential field it sources.

3.2.1 Potential interpolation

Regarding the gravitational potential, LIGER takes as an input the simulation's large scale potential field evaluated at each galaxy snapshot, interpolated on a cubic grid with an N_{grid} specified as a parameter. Since there are only two available particle snapshots to compute the potential from, we assume a linear evolution in the quantity $D^{(+)}(t)/a(t)$, hinted by the fact that at first order in density perturbations, the large

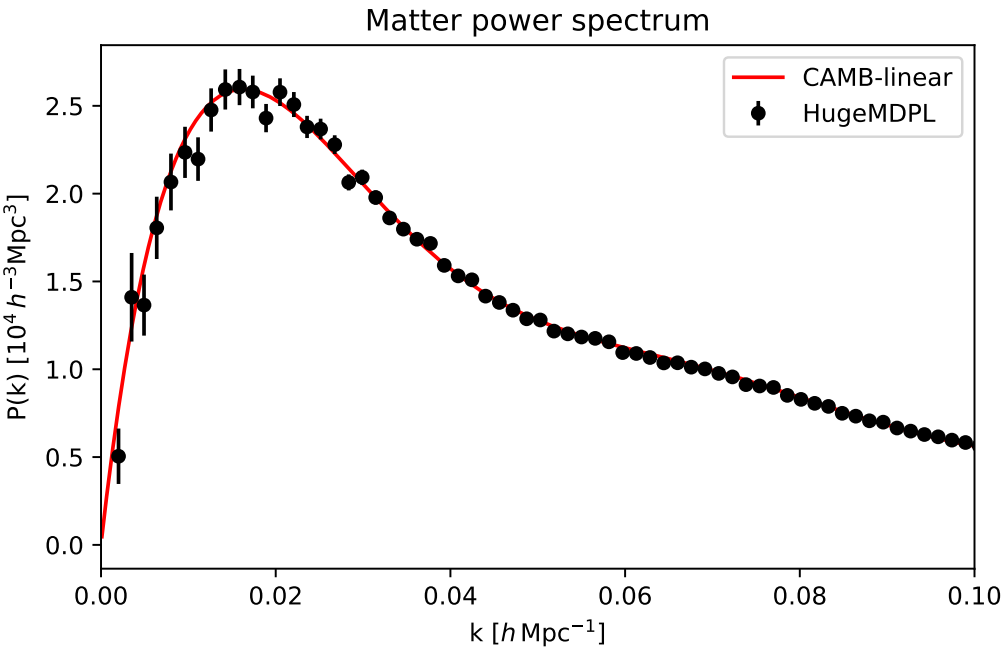


Figure 3.3: Matter power spectrum evaluated from the $z = 0$ snapshot of the HugeMDPL simulation (black dots), with the linear prediction from the CAMB code (red line). The error bars on the estimator are obtained by assuming Gaussian density fluctuations.

scale potential follows a law of the type

$$\phi(t) \propto \frac{D^{(+)}(t)}{a(t)}, \tag{3.1}$$

as derived in equation (1.38).

We test this assumption with the simulation data by computing the potential for the two available particle snapshots on a grid with $N_{\rm grid} = 2^8$, and compare the average of their ratio with the prediction obtained from equtation (3.1). The mean ratio, estimated from all the grid cells in which neither of the two potentials is equal to 0 is

$$\hat{r} \equiv \left\langle \frac{\phi(z_1)}{\phi(z_2)} \right\rangle \Bigg|_{\phi(z_1),\phi(z_2)\neq 0} = 1.15496, \tag{3.2}$$

while the theoretical prediction reads

$$r =\equiv \frac{D^{(+)}[t(z_1)]a[t(z_2)]}{D^{(+)}[t(z_2)]a[t(z_1)]} = 1.15477, \tag{3.3}$$

which means that the relative difference of the ratios is

$$\frac{\hat{r} - r}{r} = 0.016\%. \tag{3.4}$$

We conclude that, at least regarding its large-scale behaviour, the gravitational potential closely follows the linear evolution: we then extrapolate it for each redshift used in the code by utilizing the $z = 0$ value and equation (3.1).

3.2.2 Mock catalogue

Since in this project we are mostly interested in testing the new direct LIGER implementation, and this simulation does not have any semi-analytic galaxy catalogues attached to it, we decide to build a mock set of tracers from the halo data available in the ROCKSTAR catalogue. Each tracer is constructed by taking the halo's comoving position and peculiar velocity, its mass, and its unique and descendant IDs, necessary for the construction of trajectories across different snapshots. We also assign a luminosity to each halo, making use of a mock mass-luminosity relation derived from the low-mass tail of the halo mass-galaxy luminosity relation presented in [95]:

$$L(M) = A\frac{(M/m')^b}{[c+1]^{1/k}}, \tag{3.5}$$

with $A = 5.7 \times 10^9\, L_\odot/h^2$, $m' = 0.77 \times 10^{11}\, M_\odot/h$, $b = 4$, $c = 0.57$, $k = 0.230$. We show the plot of this relation in figure (3.4). Note that this is not a realistic model, since for the large halo masses considered, we would expect them to host multiple sub-haloes with multiple galaxies [95].

The trajectories of the mock tracers are constructed from the merger tree available in the catalogue, following the pre-processing shown in section (2.2.1).

In figure (3.5a) we show the differential halo abundance in function of the mass compared with the Shirasaki mass distribution fit [96]. We observe a discrepancy at large masses $M \gtrsim 10^{14}\,\mathrm{M}_\odot$. However, the calibration of the high-mass range is sensible to the size of the cosmological simulation used (see e.g. section 5.2 of [96]), and in fact the HugeMDPL run has a box size almost four times the size of the biggest simulation used in that work.

In figure (3.5b) we instead plot the luminosity function of the mock tracers, built by applying the mass-luminosity relation in equation (3.5) to the ROCKSTAR catalogue. We also plot the luminosity function derived by combining the Shirasaki mass distribution with the mock mass-luminosity relation of equation (3.5). The discrepancy at a large luminosity is expected, since it is encoded in the mass distribution we used to derive it. Nonetheless, this shows consistency in this luminosity assignment.

We will only treat the haloes with masses larger than $M_{\mathrm{cut}} = 10^{13}\,\mathrm{M}_\odot/h$, in order to consider only the ones made up by at least ≈ 120 particles. This correspond to luminosities of at least $L_{\mathrm{cut}} \approx 2.3 \times 10^{17}\,\mathrm{L}_\odot/h$. However we will set a lower mass cut $M_{\mathrm{select}} = 10^{12}\,\mathrm{M}_\odot$ for the input catalogue. This is done for multiple reasons. First of all, we can only treat haloes that appear at least in two snapshots, due to the interpolating process: applying a mass cut too near to the actual cut we consider, would inevitably lead to the exclusion of some haloes that might rise above the cut at the lightcone intersection. Second, the luminosity we use is directly linked to the mass, and we need to make sure we correctly capture the magnification bias. If we exclude all the galaxies just below the flux cut, we would risk to not account for the ones that would rise above it when magnified, which contribute to $\mathcal{Q}$.

In the bottom plot of figure (3.6) we show in grey the power spectrum of the tracers overdensity field, obtained using a grid with $N\mathrm{grid} = 2^8$ cells per side and the CIC scheme.

3.2.3 Bias relation test

In order to implement the field approach, we have to ensure that at large enough scales, the tracer we want to reproduce obeys a linear bias relation with a scale-independent linear bias b:

$$\delta_g = b \times \delta, \tag{3.6}$$

where δ_g is the tracer overdensity and δ is the matter one. We check this for the $z = 0$ snapshot, for which we have both particle and tracer data, and following equation (1.72), we do so by estimating $b(k)$ from the ratio of the power spectra

$$\hat{b}(k) = \sqrt{\frac{\hat{P}_g(k)}{\hat{P}(k)}} \quad \text{where } k \leq k_0, \tag{3.7}$$

where $\hat{P}_g(k)$ and $\hat{P}(k)$ are the tracer and matter power spectra estimates, while k_0 is an upper limit for k that we choose so that the bias relation holds. We then

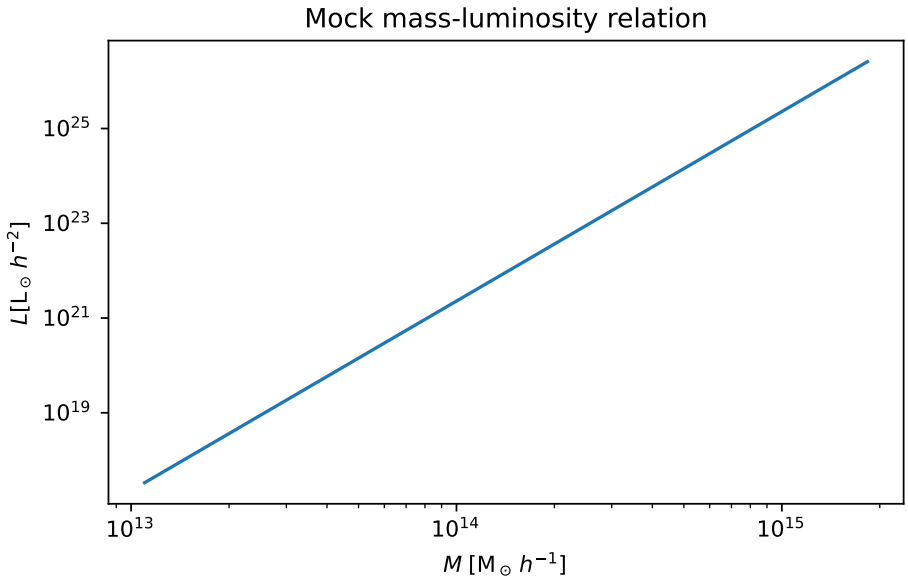


Figure 3.4: Mass-luminosity relation used for the construction of mock tracers. Note that this is not a realistic relation, since at the masses considered the haloes generally host multiple galaxies in its sub-haloes.

compute the weighted average of $\hat{b}$ over k with a weight given by $1/\sigma_b^2$, which using error propagation on equation (3.7) reads

$$\sigma_b(k) = \frac{1}{2\hat{b}(k)} \sqrt{\left[\left(\frac{1}{\hat{P}(k)}\right)^2 \sigma_{\hat{P}_g}(k)^2 + \left(-\frac{\hat{P}_g(k)}{\hat{P}(k)^2}\right)^2 \sigma_P(k)^2\right]}, \tag{3.8}$$

where the error on the power spectrum estimator is given in equation (2.17), assuming Gaussian density fluctuations. In conclusion, the estimate of the linear bias is

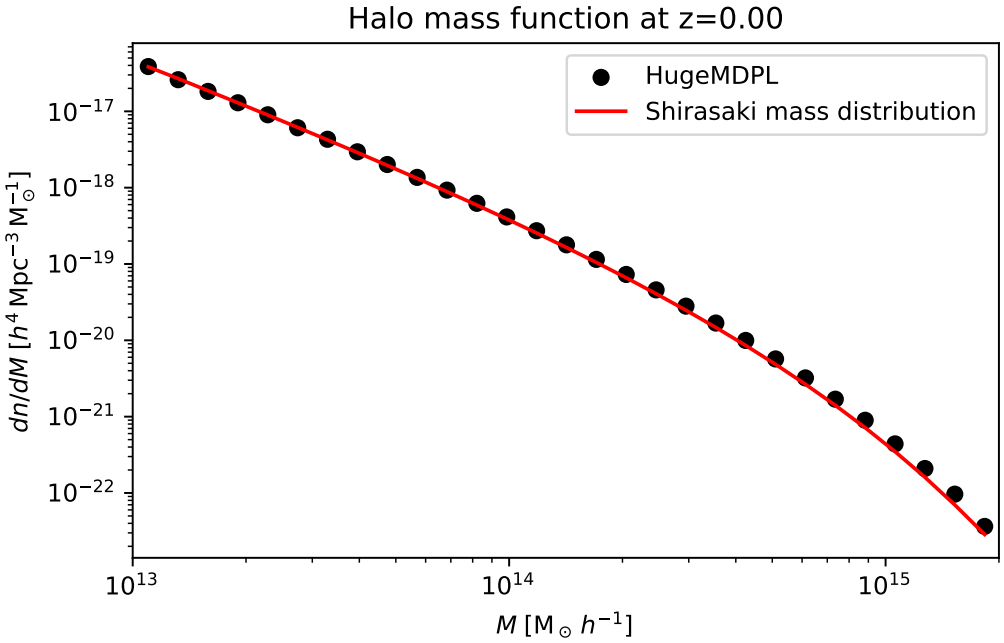
$$\hat{b} = \sum_{k_i \leq k_0} \hat{b}(k_i) \frac{1}{\sigma_b(k_i)^2}, \tag{3.9}$$

where the sum is intended over the scales $k \leq k_0$.

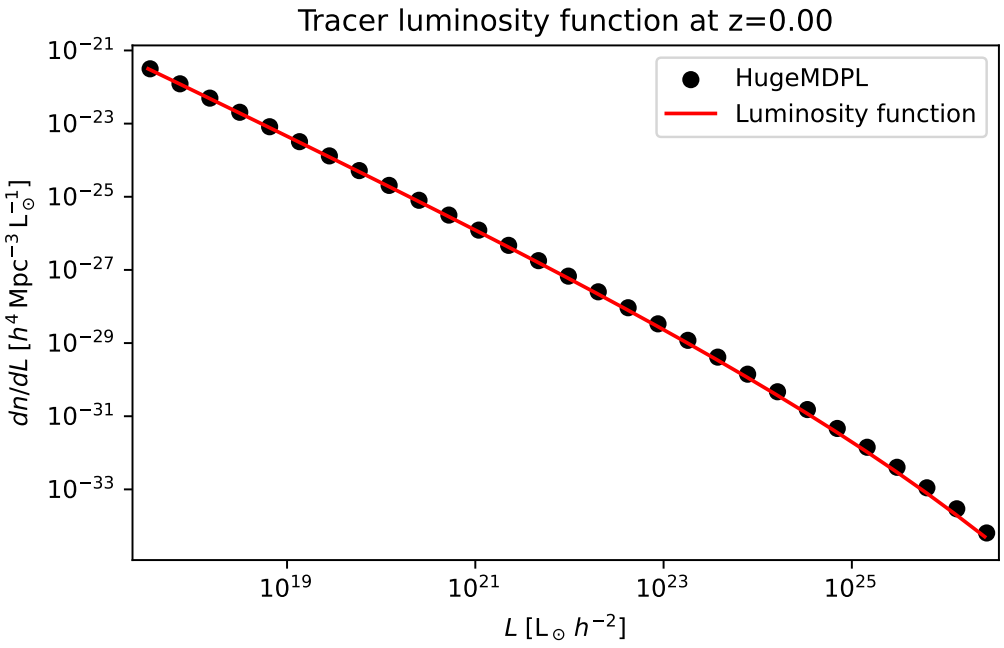
In the top plot of figure (3.6) we show the values of $\hat{b}(k)$ estimated with $k_0 = 0.1\, h/\mathrm{Mpc}$: we can see that taking into account the error bars, the curve is compatible with a constant value of $\hat{b} = 1.42$, therefore we will consider the biasing relation valid for values of k smaller than k_0. We conclude by comparing in the bottom part of figure (3.6) the tracer power spectrum with the biased linear prediction

$$\hat{P}_b(k) = \hat{b}^2 \times P(k), \tag{3.10}$$

where $P(k)$ is the linear prediction of the power spectrum. We can see once again that at the scales considered the data is compatible with the assumption of equation (3.6).



(a) Differential halo abundance obtained from the HugeMDPL halo catalogue (black dots), compared with the Shirasaki fit (red line).



(b) Tracer luminosity function obtained by assigning the mock luminosity of equation (3.4) to the halo catalogue (black dots), compared with the fit prediction (red line).

Figure 3.5: Mass and luminosity functions of the set of mock tracers built for this work.

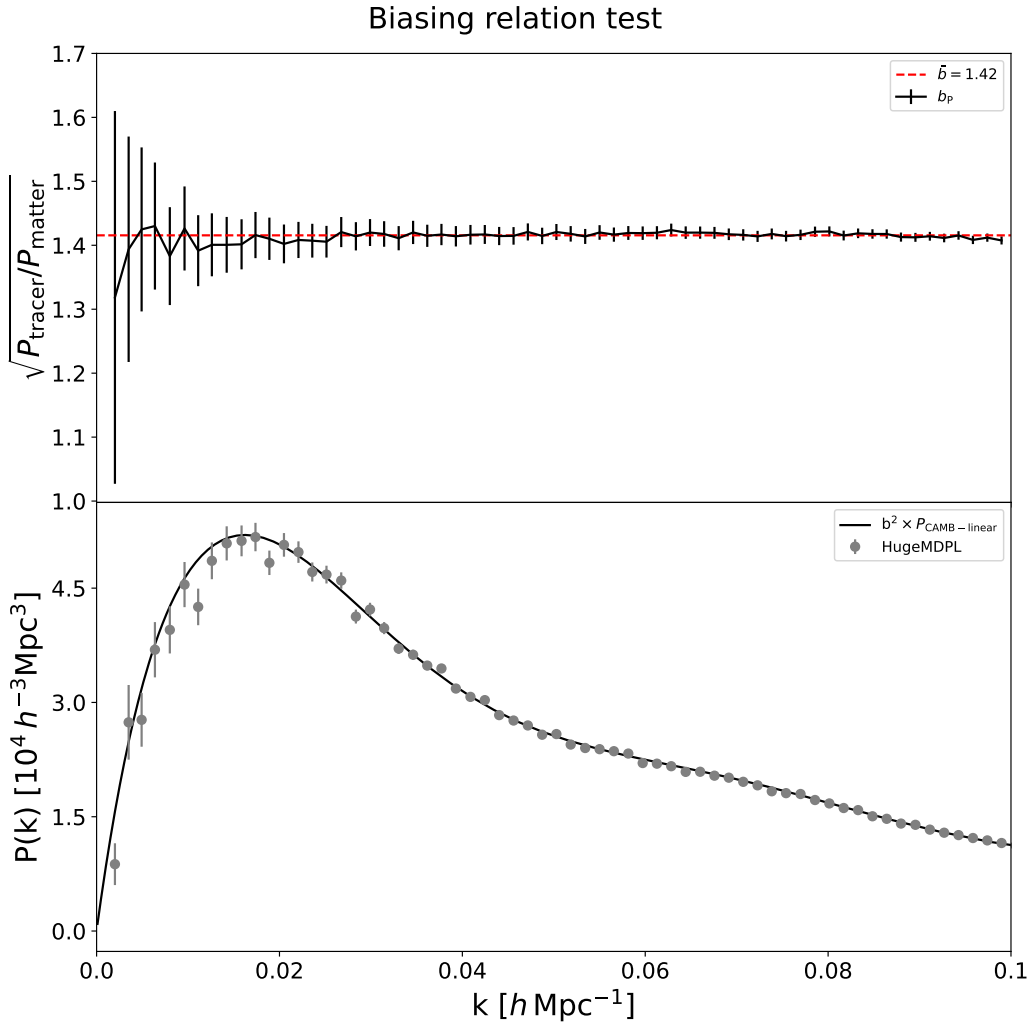


Figure 3.6: Top: Tracer linear bias estimated from $z = 0$ snapshot (red dashed line). The black points show the estimated value for each wavemode k considered, and the error bars are given by equation (3.8). Bottom: Tracer power spectrum evaluated from the $z = 0$ snapshot of the HugeMDPL simulation (grey dots), with the linear prediction from the CAMB code amplified by the linear bias b (black line). The error bars on the estimator are obtained by assuming Gaussian density fluctuations.

3.3 MUSIC runs

In this project the low-resolution dark matter-only simulations, necessary for the implementation of the field approach, were computed via the MUSIC code [66], which implements Lagrangian 2-nd order perturbation theory (2LPT) to simulate the large scale structure of the Universe. This approximation scheme allows to produce many different realizations for a lower computational cost, compared to an N-body simulation.

We run 16 simulations, all with the same cosmological parameters as the HugeMDPL run, using 2^8 particles per side of mass $m_{\rm dm} = 32\ 512 \times 10^{10}\,{\rm M}_\odot/h$, within a box size of $4000\,{\rm Mpc}/h$. From each run, we save the snapshots at the same redshifts as the ones of the HugeMDPL simulation.

In figure (3.9) we compare the power spectrum estimated from one of the MUSIC simulations with the linear theory prediction obtained with the CAMB code. We can see how on wave-modes larger than $k \approx 0.02\,h/{\rm Mpc}$ the MUSIC power spectrum becomes noticeably lower than the linear prediction. This is an expected issue of the 2LPT approximation, which is known to start underestimating the true power spectrum after a certain frequency k, which depends weakly the target resolution (i.e. the number of particles used per simulation box size) and becomes larger with redshift (see e.g. section 3 of [97]). This means that the approximation scheme will become worse as $z \to 0$, since it will start underestimating the power at increasingly larger scales. Regarding other possible causes to the discrepancy as $k \to 0.1\,h/{\rm Mpc}$, we do not expect further deviations from the CAMB spectrum caused by the fact that it is a linear prediction, because as we saw in figure (3.3), it fits well with the HugeMDPL N-body spectrum even up to $k = 0.1\,h{\rm Mpc}$. Finally, we notice that the MUSIC signal grows again close to the CAMB spectrum near $k \approx 0.1\,h/{\rm Mpc}$. We can attribute this to a spurious gain of power due to aliasing effects near the Nyquist frequency, which due to our large box and coarse grid is quite small: $k_N \approx 0.20\,h/{\rm Mpc}$.

We will take all these effects into account in doing our comparisons, as they will only act on the field approach mocks.

3.4 Survey functions estimation

In order to implement the field approach, one needs a set of survey functions that describe the tracer to "paint" the correct density field over the matter one. For this work, since we use the field approach to test the direct approach, we estimate the survey functions from the simulation itself.

In the mock galaxy catalogue we created starting from the ROCKSTAR haloes, we have information about positions, velocities and luminosities. A set of survey functions changes with the chosen flux cut $f_{\rm cut}$, which selects only the galaxies that have an apparent flux $f \geq f_{\rm cut}$. For a snapshot at a redshift z the flux cut will select all the galaxies with a luminosity $L \geq L_{\rm cut}(z) \equiv 4\pi f_{\rm cut} d_L(z)^2$, following equation (1.134). Given a luminosity function $\phi(L, z) = dN(L' > L)/dLdV$, defined such that

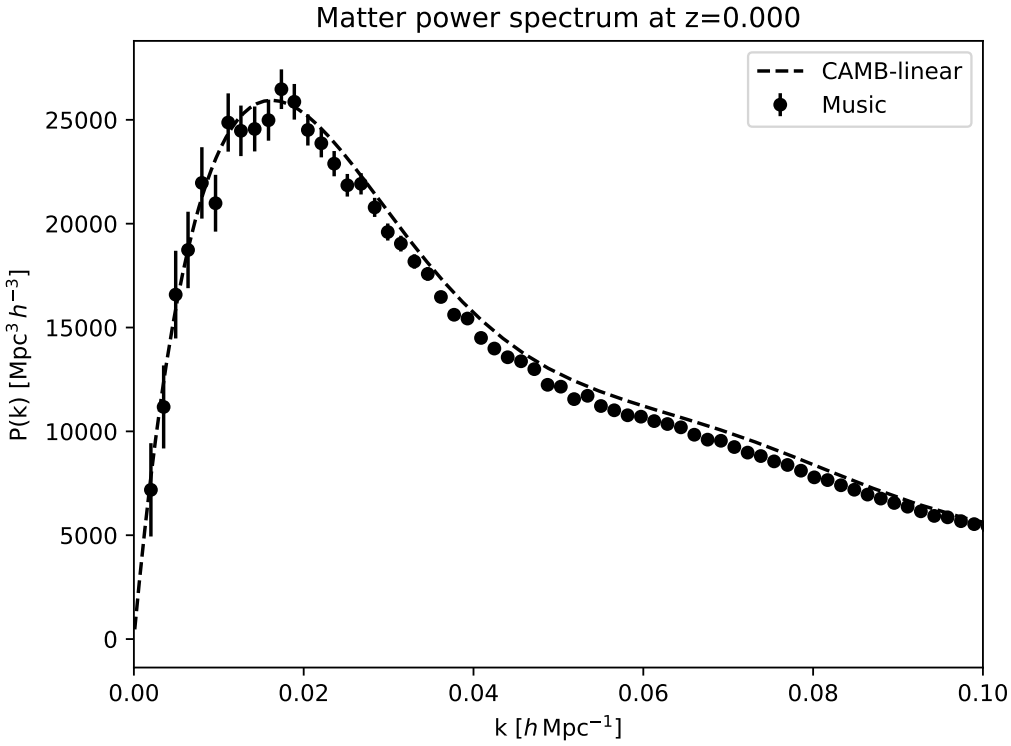


Figure 3.7: Comparison of the matter power spectrum computed from one of the MUSIC runs (black dots) with the prediction from the CAMB code at $z = 0$. The error bars on the estimator are computed assuming Gaussian density fluctuations.

the comoving number density of galaxies brighter than L is

$$n_g(L,z) = \int_L^\infty \phi(L',z)dL, \tag{3.11}$$

the number density of galaxies visible above the flux cut $f_{\rm cut}$ is

$$\bar{n}_g(z) = n(L_{\rm cut}(z),z) = \int_{L_{\rm cut}(z)}^\infty \phi(L',z)dL. \tag{3.12}$$

From the definitions in equations (1.136, 1.137) it follows that [16]

$$\mathcal{Q} = -\frac{L}{n(L,z)}\frac{\partial n(L,z)}{\partial L}\bigg|_{L=L_{\rm cut}(z)} = \frac{L_{\rm cut}(z)\phi(L_{\rm cut}(z),z)}{\bar{n}_g(z)}, \tag{3.13}$$

and

$$\mathcal{E} = -\frac{\partial \ln \int_{L_{\rm cut}(z)}^\infty \phi(L',z)dL}{\partial \ln(1+z)} = -\frac{1}{\bar{n}_g(z)}\int_{L_{\rm cut}(z)}^\infty \frac{\partial\phi(L',z)}{\partial\ln(1+z)}dL. \tag{3.14}$$

These equations link the luminosity function with the two biases. We will now present the method of estimation used for each one of the survey functions.

For a given snapshot at a redshift $\tilde{z}$, we compute on a grid the matter overdensity and the overdensity of the tracers with $L \geq L_{\rm cut}(\tilde{z})$. Then, by implementing the approach described in section (3.2.3), we estimate the linear bias using the power spectra ratio. Since we have only two particle snapshots available, we interpolate the matter overdensity field δ assuming a linear evolution, analogously to what we did for the potential, following the growing mode in equation (1.35)

$$\delta(t,\mathbf{x}) = D^{(+)}(t)\,\delta(0,\mathbf{x}). \tag{3.15}$$

After evaluating the bias at each snapshot available, we estimate $b(z)$ by interpolating the values with a cubic spline.

For the galaxy number density $\bar{n}_g(z)$ the process is straightforward: for each snapshot with redshift $\tilde{z}$ we compute $n(L_{\rm cut}(\tilde{z}),\tilde{z})$ counting the galaxies with $L \geq L_{\rm cut}(\tilde{z})$, and then we interpolate with a cubic spline over the z values.

For the magnification bias $\mathcal{Q}$, we first calculate the luminosity function $\phi(L,z)$. We then evaluate $\mathcal{Q}$ it via equation (3.13). We proceed in this way:

- At a given redshift $\tilde{z}$ we construct several luminosity bins L_i around the value $L_{\rm lim}(\tilde{z})$. We use 50 luminosity bins, evenly spaced on a logarithmic scale between the largest and smallest tracer luminosities, with a logarithmic separation of ≈ 0.20.

- We compute the number density n_i of galaxies with $L \geq L_i$ for each bin.

- We fit the values of n_i with a 3-rd order polynomial in log-log space (this is done to avoid including stochastic fluctuations of the sample, which could generate relevant noise in the computation of a numerical derivative).

- We compute $\phi(L, \tilde{z})$ by taking the derivative of (3.12) with respect of L, which results in computing the analytical derivative of the polynomial fit obtained in the previous step.
- We retrieve $\mathcal{Q}(\tilde{z})$ using equation (3.13) evaluated at $L_{\rm lim}(\tilde{z})$.
- We repeat the process for each value of z we are considering.

In figure (3.8a) we show the tracer number density in function of the luminosity cut at $z = 0.3$. The error bars due to counting statistics are smaller than the points, nevertheless we see that the fitting function chosen closely follows the density evolution without any outlier data point.

Finally, the evolution bias $\mathcal{E}$ is estimated similarly to the magnification bias:

- At a given redshift $\tilde{z}$ we consider the redshift bins around it z_i.
- We compute the number density n_i of galaxies with $L > L_{\rm min}(\tilde{z})$ for each bin.
- We fit the values of n_i with a 2-nd order polynomial.
- We retrieve $\mathcal{E}(\tilde{z})$ using equation (1.137) evaluated at $\tilde{z}$. Even in this case we compute an analytical derivative of the fit estimated in the previous step.
- We repeat the process for each value of z we are considering.

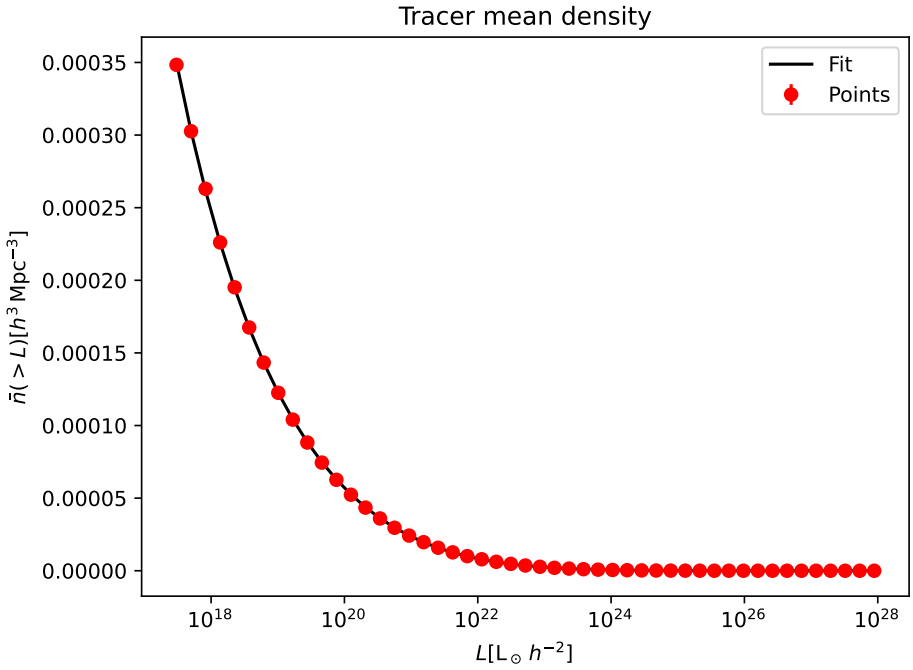
In figure (3.8b) we show the tracer number density in function of the redshift for a luminosity cut of $L = L_{\rm cut}(z = 0.3)$. The error bars due to counting statistics are smaller than the points, nevertheless we see that the fitting function chosen closely follows the density evolution without any outlier data point.

As we mentioned, we will test our code in two redshift bins, a closer one of $z \in [0.2, 0.3]$ and a deeper one of $z \in [0.6, 0.7]$. We will choose two different flux cuts in the two bins, selecting two different tracer populations. The reason we do this lies in the fact that in order to be able to make our tests, we are limited in the choice of flux cut by two conditions. First, since the tracer luminosity is directly linked to the mass via the relation $L(M)$ of equation (3.4), and since we performed a mass cut $M_{\rm cut}$ on the input catalogue, depending of the typical luminosity distance of the bin $d_{\rm L}$ we must not choose a flux cut $f_{\rm cut}$ too low, such that

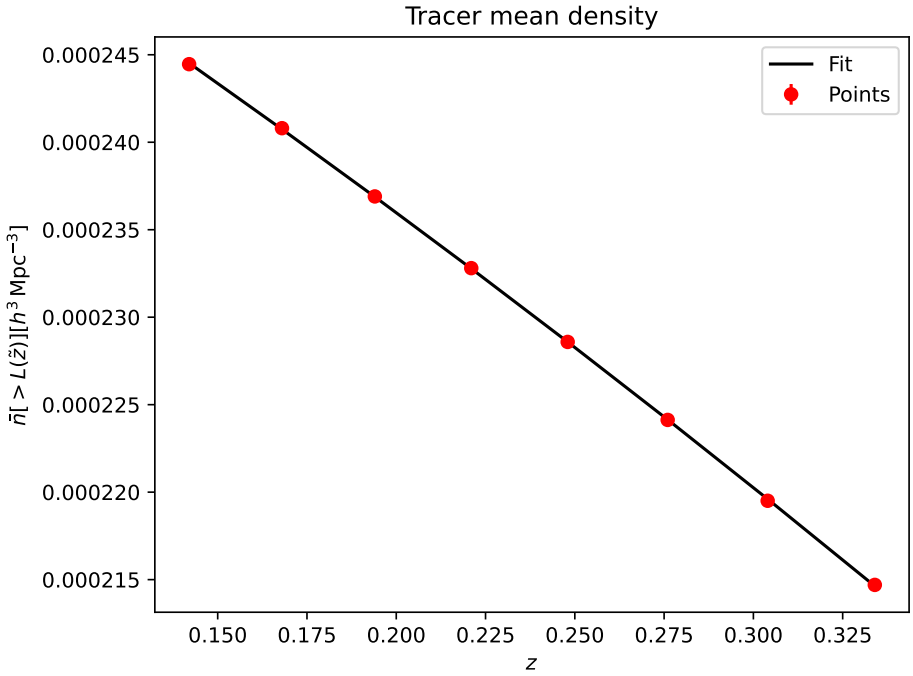
$$L(M_{\rm cut}) \gtrsim 4\pi\, f_{\rm cut}\, d_{\rm L}^2. \tag{3.16}$$

This is because we need to avoid being too close to the flux cut imposed by $M_{\rm cut}$ to correctly estimate the magnification bias, since it involves the computation of $n(> L)$ on a number of bins around $L_{\rm cut}(z)$. Second, we have an upper limit on $f_{\rm cut}$ set by the linear bias of the tracers, which grows with the flux cut. As we already mentioned, the grid size of the field approach mesh has to increase with the bias in order to correctly describe the field, but a too coarse mesh could cause aliasing effects at large scales in the monopole estimator.

Keeping in mind these two limits, we choose a value of $f_{\rm cut} = 4.0{\times}10^{-5}\,{\rm erg/cm^2/s^2}$ for the closer bin, and $f_{\rm cut} = 1.4 \times 10^{-6}\,{\rm erg/cm^2/s^2}$ for the deeper bin. We apply



(a) Tracer number density at $z = 0.3$ in function of the luminosity cut L_{cut}. The red points are estimated from the data following the procedure listed in section (3.4), where the error bars from counting statistics are smaller than the markers. The black line is the best fit with a 3-rd order polynomial in log-log space.



(b) Tracer number density for a luminosity cut of $L = L_{\mathrm{cut}}(z = 0.3)$ in function of redshift. The red points are estimated from the data following the procedure listed in section (3.4), where the error bars from counting statistics are smaller than the markers. The black line is the best fit with a 2-nd order polynomial.

Figure 3.8

these flux cuts in the direct approach catalogues and we use them in the survey function estimation, which we show in figure (3.9).

Looking at the radial density $\bar{n}(z)$ we see that the number of tracers decreases with redshift. There are multiple effects contributing to this quantity. First of all, the flux cut boosts the number of objects detected at a lower redshift, since for a given luminosity they will present a higher flux than the ones farther away. Second, the fact that the haloes grow in mass with time, and thus their luminosity increases as the redshift decreases, further boosts the effects of the flux cut. Last, there is also a balancing between the halo formation and mergers, which would in turn increase and decrease the total number counts as the redshift decreases. For similar reasons we also expect the linear bias to grow with redshift: looking farther away we only see the most massive haloes, which we expect to occupy the highest density peaks regions of the matter field (see e.g. [45]) and thus appear more clustered. If instead we look at the evolution bias, we can get more information regarding the tracer density. Since it is positive we can say that, for a fixed *luminosity* cut L_{cut}, the number of haloes that rise above it increases over time, at least for the mass range considered. Finally, we notice that the magnification bias is approximately constant. This result strictly depends on the halo population and the mock mass-luminosity relation we chose. Even if this is not a realistic galaxy catalog, we mention that a constant magnification bias is consistent with the case of a flux cut applied near the faint-end region of the Schecter luminosity function [98]. In fact, in that region $\phi \propto L^{-\gamma}$ with $\gamma > 1$, and thus $\mathcal{Q} \approx \gamma - 1$.

All these effects listed above are implicitly captured by the direct approach, without requiring any modelling, since they are already accounted for in the input N-body simulation.

To test the self-consistency of the survey functions estimated from the simulation, we make use of the relation of equation (1.138): we compute the left side by taking the total logarithmic derivative of the estimated number density, and the right side using the estimated evolution and magnification bias. We plot the comparison in figure (3.10). The comparison is made at the z values of the HugeMDPL snapshots, where the survey functions are actually estimated from the catalogue and not interpolated. In the upper part we plot the theoretical prediction via $\mathcal{Q}$ and $\mathcal{E}$, following equation (1.138) (blue continuous line), and the direct estimation via the numerical derivative of the radial density $\bar{n}(z)$ (orange dashed line). In the lower part we plot the relative difference between the two. The redshift bin that we use for the analysis is delimited by the red vertical lines.

We notice that in the closer bin the comparison holds with a 1% error, while the deeper bin has a 3% error. We do not expect to get particularly small errors for the following numerical reasons: the direct estimate consists in the numerical derivative of a cubic spline obtained without any fitting procedure. Then, we can expect $\bar{n}(z)$ to be subject to stochastic fluctuations due to its estimation. Moreover, we argue that the discrepancy at the redshift limits is sourced by the cubic spline interpolation scheme, which assigns artificial boundary conditions to the derivatives of the function at its domain limits. In any case we took this into account by estimating the survey functions on a bin larger than the one we will actually be using.

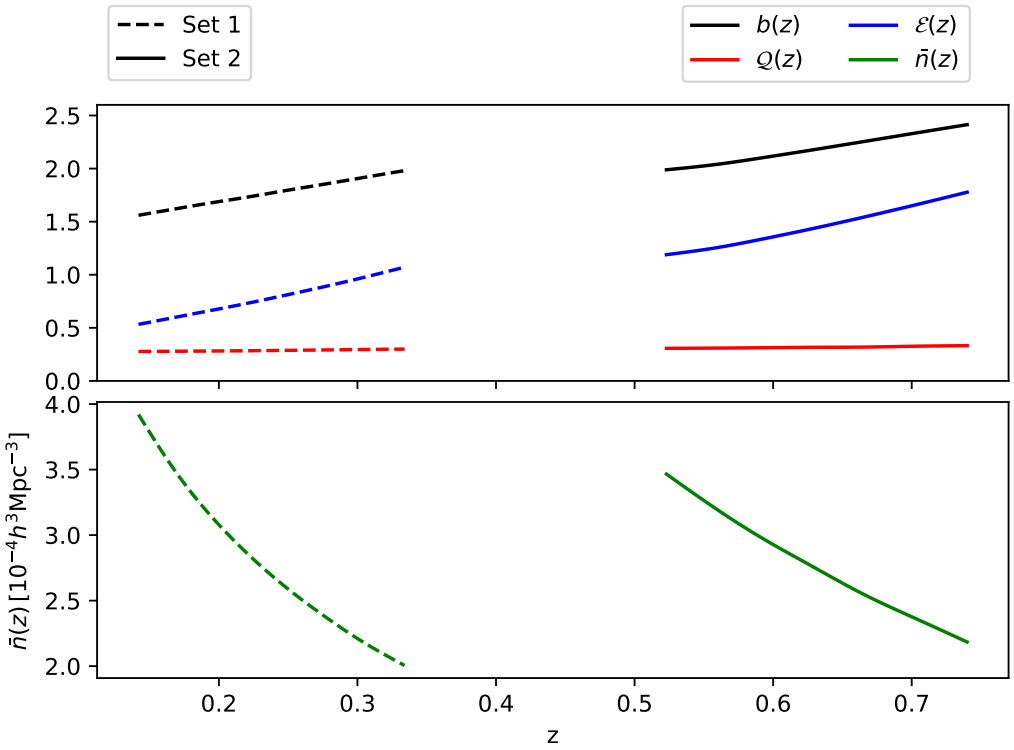
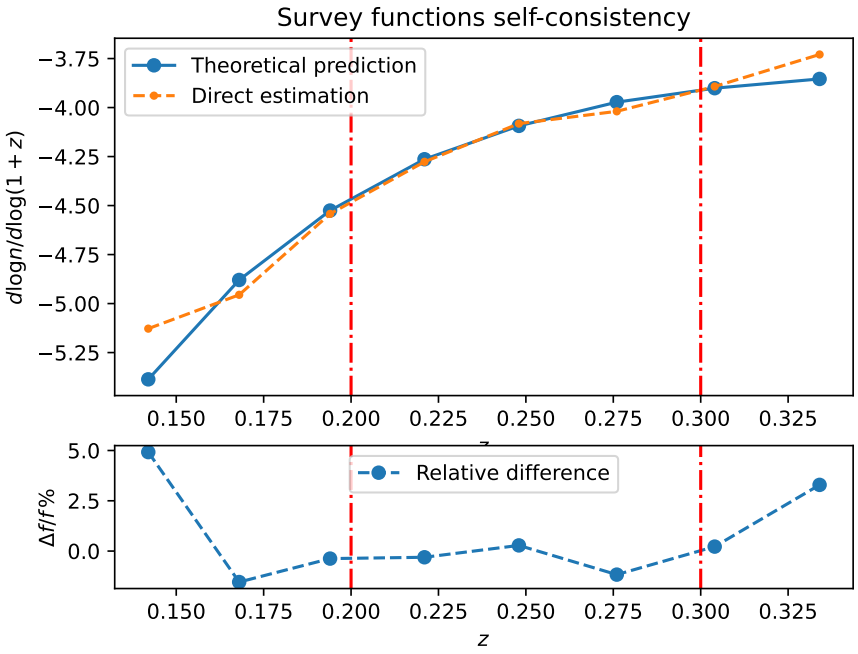
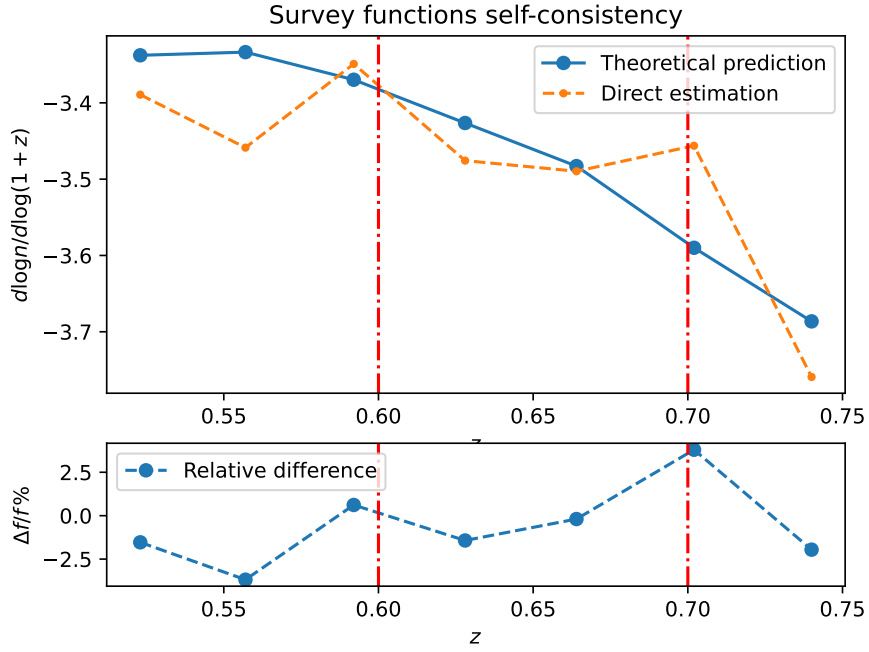


Figure 3.9: Survey functions estimated for the $z \in [0.2, 0.3]$ bin with $f_{\rm cut} = 4.0 \times 10^{-5}\,{\rm erg/cm^2/s^2}$ (set 1), and for the $z \in [0.6, 0.7]$ bin $f_{\rm cut} = 1.4 \times 10^{-6}\,{\rm erg/cm^2/s^2}$ (set 2). Top: Linear bias (black line), the magnification bias (red line) and evolution bias (blue line). Bottom: Tracer radial mean density.



(a) Survey function self-consistency test for the $[0.2, 0.3]$ redshift bin, delimited by the red vertical lines. Top: Total logarithmic derivative comparison, obtained directly from the radial mean density (orange dashed line) and from the magnification and evolution biases (blue line). Bottom: Relative discrepancy of the two estimates. The large discrepancy at the functions' definition boundaries are caused by the artificial boundary conditions set on the derivatives by the cubic spline interpolation scheme.



(b) Survey function self-consistency test for the $[0.6, 0.7]$ redshift bin, delimited by the red vertical lines. Top: Total logarithmic derivative comparison, obtained directly from the radial mean density (orange dashed line) and from the magnification and evolution biases (blue line). Bottom: Relative discrepancy of the two estimates. The large discrepancy at the functions' definition boundaries are caused by the artificial boundary conditions set on the derivatives by the cubic spline interpolation scheme.

Figure 3.10

3.5 Monopole comparison

After having estimated the survey functions we show the power spectrum monopole of our realizations, for both the redshift bins considered and for all the different mocks produced by LIGER.

3.5.1 Deep bin

We start by analyzing the deep $z \in [0.6, 0.7]$ bin. Since we have one realization of the HugeMDPL simulation available, we can use it to build only one mock catalog of the direct approach. Instead, for the field approach we use the 16 MUSIC outputs of the runs we made for this work, producing 16 independent realization of the tracer catalog. As we already mentioned in section (2.4.1), the "Gaussian" covariance estimator for the FKP monopole is not accurate at large scales, so we will not use it for the deep HugeMDPL realization. Luckily, regarding the field approach, we can use the scatter of the 16 realizations, so that a comparison with error bars can still be done.

In figure (3.11) we compare the redshift space power spectra (in red) and the realspace ones (in black) for each one of the mocks produced by LIGER listed in section (2.1). The points are computed from the direct catalogue (HugeMDPL), while the shaded areas are the $1-\sigma$ and $2-\sigma$ regions of the 16 field realizations scatter from the mean value, denoted with a full line. We estimate the radial mean density $\hat{n}(\mathbf{x})$ from each catalogue via concentric shells of $\delta r = 15\,\mathrm{Mpc}/h$ around the observer, to then use it in the FKP computation. The FFT for the power spectrum monopole estimator is computed on a box of size $L_{\mathrm{FFT}} = 4\,\mathrm{Gpc}/h$, which implies a fundamental frequency of $k_{\mathrm{fund}} \approx 1.6 \times 10^{-3}\,h/\mathrm{Mpc}$. We use two different N_{FFT} cells per side values for the two methods. For the field approach we use $N_{\mathrm{FFT}} = 2^7$, which equals to a grid size of $31.25\,\mathrm{Mpc}/h$. We can't use a finer mesh because the field approach realizations cannot produce a finer grid, due to the biasing issues listed in section (2.1.3). This is not a limit of the direct approach, and for this reason we decide to use $N_{\mathrm{FFT}} = 2^9$. In the field approach we add shot noise "artificially" following the cell sampling procedure mentioned in section (2.1.3) (FieldSN). We then correct for the shot noise in the power spectrum estimate for both the samples. We also show the same field approach density realization without the shot noise introduction (Field), and we plot it with a dashed line. For the catalogues with an observer contribution we set $v_{\mathrm{obs}} = 1000\,\mathrm{km/s}$. As we will see later in section (3.6), we choose this value because it generates a signature on the power spectrum monopole comparable with the scatter of the realizations, which will be useful for the rest of the analysis.

Looking at the comparison in figure (3.17), we see that the configurations are consistent for $k < 0.03\,h/\mathrm{Mpc}$, in all the mocks, since most of the HugeMDPL points lie in the $1-\sigma$ region, and a few of them in the $2-\sigma$ region. We also notice that for all the mocks, the relative position of the HugeMDPL points with the respect of the field approach average is consistent going from real to redshift space. There is a large scatter at large scales, comparable with the magnitude of the signal. This is to be expected due to the sample variance of the box, which limits our knowledge at the

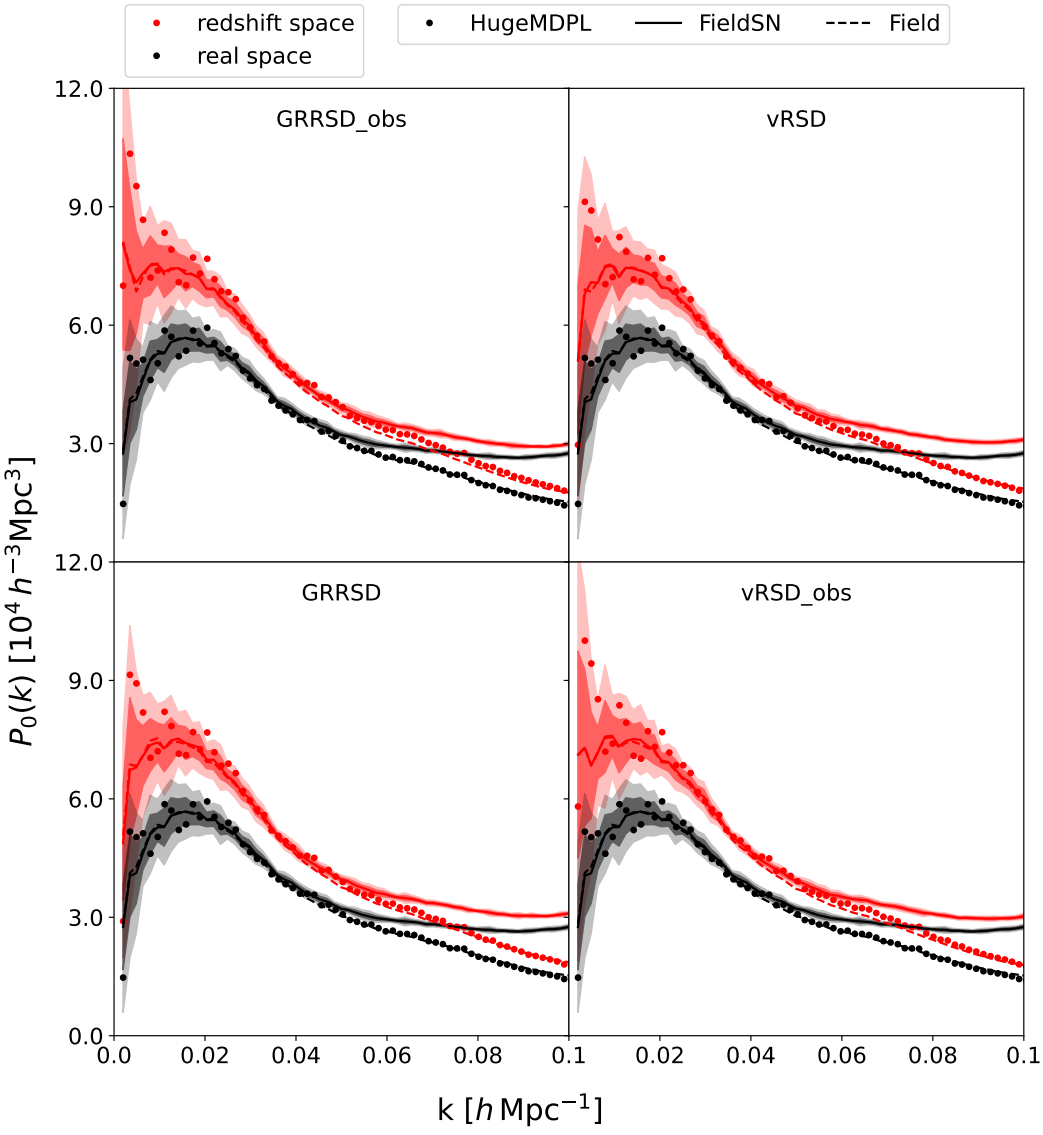


Figure 3.11: Power spectrum monopole test in the $z \in [0.6, 0.7]$ bin, for the real space (black) and redshift space mocks (red), accounting for different contributions. We compare the direct approach result for one realization (points) with the field approach 1 and $2-\sigma$ scatter of 16 realizations (shaded areas) around the mean value (full line). We also plot with a dashed line the average value of the field mocks without the inclusion of shot-noise.

smallest k values, and due to the window function convolution, which flattens the signal. This is unfortunate since the effects we want to test (e.g. the GR contribution or magnification effects) contribute at those scales. We will present an alternative approach to overcome this problem in the next section.

As we probe larger k values, we notice that the approaches start deviating, becoming incompatible. First of all, looking at the FieldSN sample, the comparison stops holding around $k \approx 0.04\,h/\text{Mpc}$, where it starts deviating from both the HugeMDPL and Field mocks. Due to this lack of self consistency we conclude that the shot-noise is not correctly captured by the field approach as we get to smaller scales. We can expect this discrepancy to appear at different scales based on the refinement of the BUILDCONE mesh, possibly producing a better estimate as the grid cell shrinks.

If we compare the HugeMDPL mocks with the Field mocks we notice that after $k \approx 0.04\,h/\text{Mpc}$ the signal of the former becomes higher than the latter, to then converge again near $k = 0.1\,h/\text{Mpc}$. This signature is expected due to the nature of the MUSIC mocks used: if we look at the power spectrum of one MUSIC snapshot at $z = 0$ in figure (3.9), we notice that the signal showed there follows a similar trend compared to the linear prediction, so we conclude that the discrepancy are caused by the fact that MUSIC fails to capture the correct power spectrum at smaller scales.

To show that the direct approach gives instead the correct (or better) results at larger k-s we test the accuracy of the Kaiser boost on the power spectrum monopole, as described in section (1.4.5). We know that this effect is valid only in the plane-parallel limit, where we can consider small separation angles between galaxies. Differently from the derivation made in the $z \to 0$ limit, which also implied $\Delta z < 1$, in our case we have to take into account the Kaiser factor $\mathcal{F}$ evolution on the redshift bin considered. This was implemented in [16] by taking a weighted average of $\mathcal{F}$ over the redshift bin $[z_i, z_f]$, resulting in a factor

$$\bar{\mathcal{F}} = \frac{\int_{z_i}^{z_f} \mathcal{F}\bar{n}_g^2 D_{(+)}^2 \frac{dV_\mathcal{S}}{dz} dz}{\int_{z_i}^{z_f} \bar{n}_g^2 D_{(+)}^2 \frac{dV_\mathcal{S}}{dz} dz}, \tag{3.17}$$

where $D^{(+)}$ is the linear growing mode of density perturbations, $\bar{n}_g$ is the tracer mean radial density and $V_\mathcal{S}$ is the comoving volume withing the past lightcone of the observer.

We compare this value with the ratio between the redshift and real space power spectrum monopoles $\hat{P}_0^s(k)/\hat{P}_0(k)$, for all the mocks produced by LIGER. We do not expect this to hold at very large scales, where wide angle effects take place and further boost the redshift space signal, but instead at intermediate scales, at which we can still work in linear approximation and consider small angular separations.

In the top plot of figure (3.12) we show the result of the comparison: the red line is the theoretical prediction $\bar{\mathcal{F}}$, the black dots are the estimation from the HugeMDPL mock, and the shaded areas indicate the $1-\sigma$ and $2-\sigma$ scatter of the FieldSN estimates around the mean value (full line). We also show the not-noised Field mocks as a dashed line. At large scales $k < 0.02\,h/\text{Mpc}$ we see that both the field and direct estimations do not retrieve the correct Kaiser factor, but this is expected due to the wide-angle effects that act at large scales. Regarding the GRRSD_obs

and vRSD_obs mocks, at the largest scale probed, we recognize a further boost due to the observer's velocity, as it was described in section (1.145).

On smaller, but still large, scales we notice two things: the direct approach seems to be capturing the Kaiser factor better than the FieldSN approach, and this becomes more noticeable starting from $k \approx 0.04\,h/\mathrm{Mpc}$. We partially attribute this large discrepancy to the artificial shot noise issue mentioned above: as we can see looking back to figure (3.11), the Poisson sampled mock starts deviating from the direct approach around $k \approx 0.04\,h/\mathrm{Mpc}$, which is the same scale in which we see it deviate from the plane parallel approximation. However, this is not the only cause: the Field mocks deviate too from the Kaiser prediction, even if less than the "shot-noised" counterpart.

In the bottom plots of figure (3.12) we plot the discrepancy between the estimates and the theoretical value in units of the $1-\sigma$ scatter estimated through the FieldSN mocks. Besides the expected incompatibility with $\bar{\mathcal{F}}$ at small k values, we note that also here the FieldSN mocks quickly deviate from the plane parallel prediction, while the HugeMDPL mocks keep capturing the effect better, up to smaller scales. In this plot we can identify better the scale in which the "not-noised" Field mocks start to deviate, which is around $k \approx 0.04\,h/\mathrm{Mpc}$, the same scale in which the discrepancies with the HugeMDPL mocks started.

We still notice that the HugeMDPL signal underestimates the effect as k grows, but there are a few factors to take into account when considering this. First, $\mathcal{F}$ depends on the linear bias b, which we estimated from a power spectra ratio. Spurious effects like the shot noise removal in the power spectrum can affect this estimate, and consequently the $\mathcal{F}$ factor. Second, $\bar{\mathcal{F}}$ is a weighted average over the redshift bin considered, so we can expect this to only approximate the effect of redshift space distortions. Third, as we reach non-linear scales, the $\bar{\mathcal{F}}$ factor should be computed differently (see e.g. [99]).

3.5.2 Closer bin

As we mentioned, we only have a single realization for the HugeMDPL simulation, and this limits the possibility of estimating the scatter of the power spectrum. One solution to this issue is to subdivide the original $4\,\mathrm{Gpc}/h$ box in a set of contiguous sub-boxes, and treat those as separate realizations. There is one caveat to this approach: by subdividing the box we must consider a closer redshift bin, depending on the number of subdivision. Taking this into account we decide to divide the simulation in 8 equal parts with sides of $2\,\mathrm{Gpc}/h$, which allow us to reach $z = 0.3$ and still have a rough scatter estimate.

We thus repeat the monopole comparison for a $z \in [0.2 - 0.3]$ redshift bin, where this time we also show the $1-\sigma$ scatter of the direct approach over the 8 sub-mocks and do not show the $2-\sigma$ scatters to avoid a confusing plot. The comparison is plotted in figure (3.13), with the same color-labeling as the previous case, but with the addition of an estimate of the $1-\sigma$ scatter for the direct approach.

From the comparison we can make the same observations as the previous bin, but we notice that the HugeMDPL scatter has smaller error bars with respect to

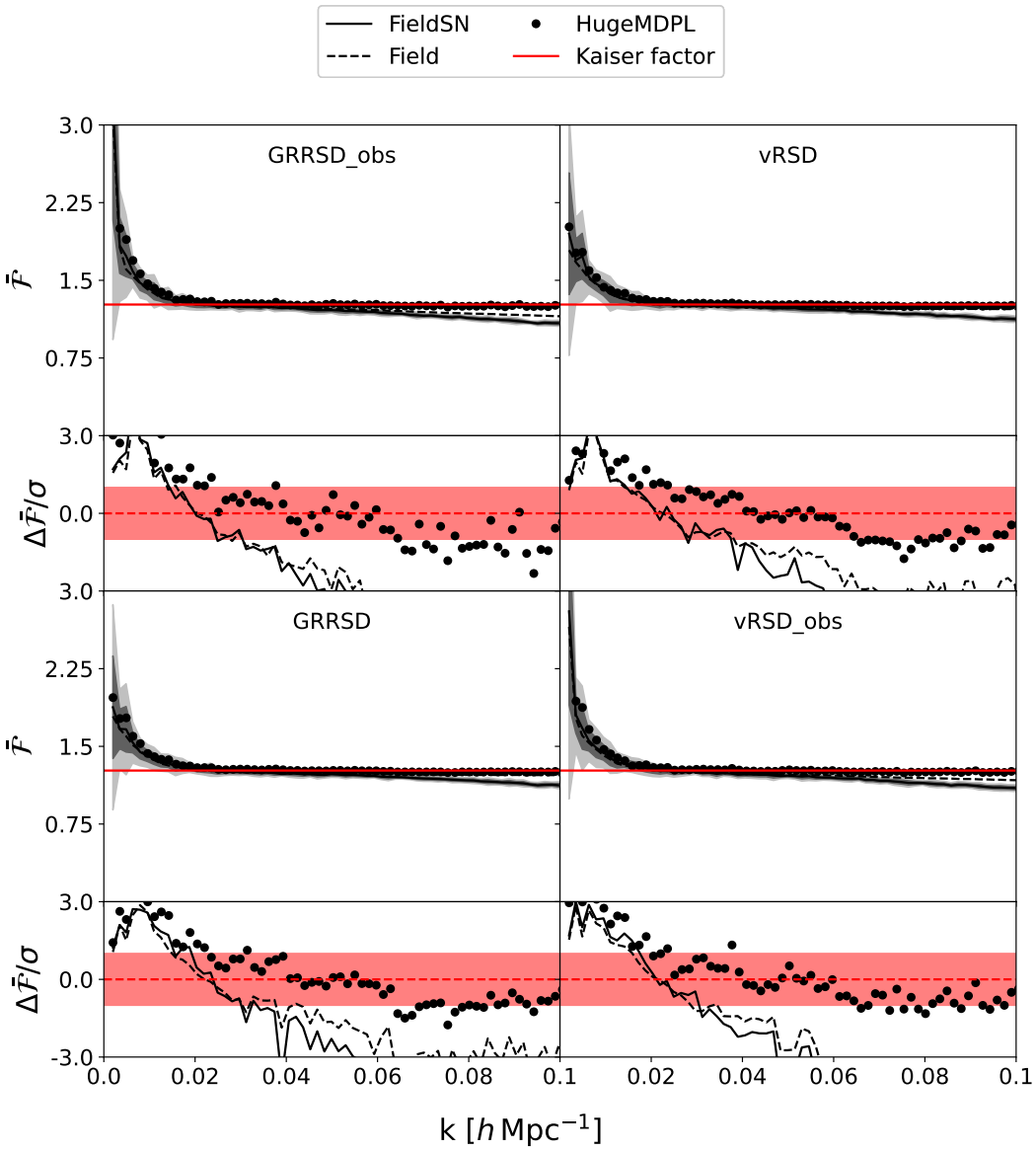


Figure 3.12: Kaiser factor estimation in the $z \in [0.6, 0.7]$ bin (in black), accounting for different contributions, and its theoretical prediction from equation (3.17) (red dashed line). We compare the direct approach result for one realization (points) and the field approach 1 and $2-\sigma$ scatter of 16 realizations (shaded areas and full line). We also plot the average value of the field mocks without accounting for shot-noise (dashed black line).

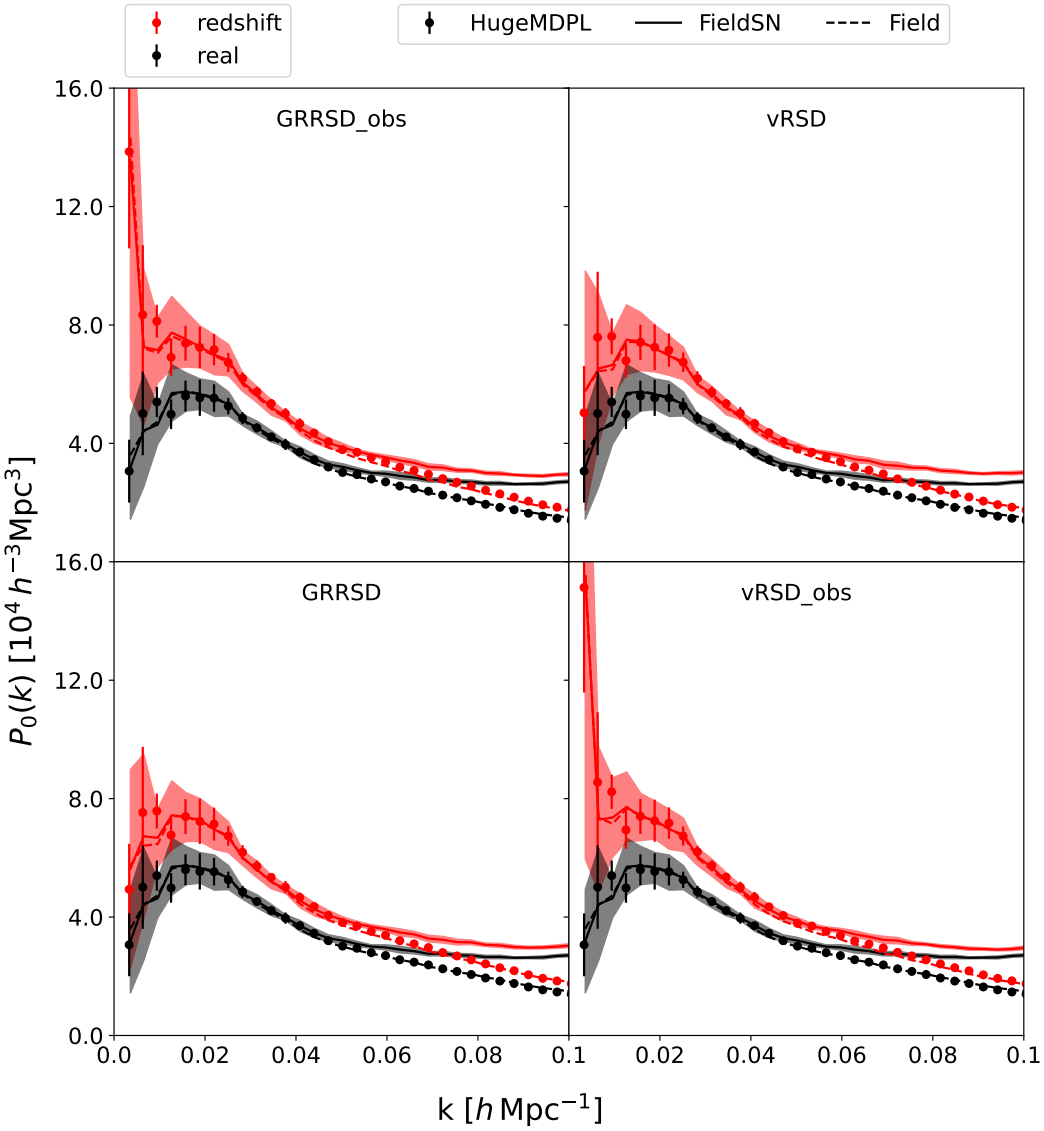


Figure 3.13: Power spectrum monopole test in the $z \in [0.2, 0.3]$ bin, for the real space (black) and redshift space mocks (red), accounting for different contributions. We compare the direct approach result from the 8 sub boxes with their $1-\sigma$ scatter (barred points) and the field approach $1-\sigma$ scatter of 16 realizations (shaded areas). We We also show with a dashed line the average value of the field mocks without the inclusion of shot-noise. We argue that the $1-\sigma$ bars of the direct approach show a reduced variance with respect to the field approach because the sub-boxes from which we extracted the mocks started from the same seeds for the initial conditions, and are thus correlated.

the FieldSN mocks. However, since this happens both in real and redshift space, we expect it to be unrelated to the effects implementation. One of the possible reasons for the smaller scatter of these 8 mocks could be related to the way we built them: since they were all obtained by subdividing the same simulation box, they share the same seeds for the initial conditions (IC) generation. Depending on the algorithm for IC generation we used, we could expect a correlation between the wave modes sampled from different regions of the box, that could reduce the variance of the produced mocks. Regarding the observer's velocity signature in the vRSD_obs and GRRSD_obs mocks, we notice that, on average, the signal in this bin is higher than in the deeper bin. Looking back at equation (1.144) we can attribute this to the behaviour of the Bessel function $j_0(kr)$, which peaks when $k \sim 1/r$: for a closer radial bin then the value of k at the peak is larger, and we can probe it more closely.

Now we repeat the Kaiser effect test for this bin and we show the result in figure (3.14). Due to the correlation of the sub mocks in the first example we still use the field approach error bars to compute relative discrepancy. We notice overall the same behaviours as the previous bin, with less noticeable wide-angle effects at large scales, consistent with the fact that we are probing a smaller comoving volume. Even in this plot we can notice how the error bars of the HugeMDPL mocks are smaller than the FieldSN ones.

3.6 Observer effect

To conclude our tests we estimate the observer's velocity contribution on the power spectrum monopole $\Delta P_{\rm obs}$, and we compare it with the theoretical prediction of equation (1.145). We know that $D_{\rm iso}$ depends on both the magnification and evolution biases, so with this comparison we can test if their effects are correctly captured by the direct approach. Moreover, since $D_{\rm iso} \propto v_{\rm obs}^2$, we can control the signal-to-noise ratio (SNR) by setting an observer velocity sufficiently large. We find that a speed of $v_{\rm obs} = 1000\,{\rm km/s}$ grants us a good SNR, so we assign that value for the observer's velocity in every mock.

We follow the procedure described in [16]: we estimate the monopole $\hat{P}_{0,\rm obs}$ from the **GRRSD_obs** mocks and $\hat{P}_{0,\rm cmb}$ from the **GRRSD** mocks, were the "cmb" label denotes that the observer is steady with respect to the CMB. Then we estimate $\hat{D}_{\rm iso} = \hat{P}_{0,\rm obs} - \hat{P}_{0,\rm cmb}$. To better characterize the oscillatory features of $D_{\rm iso}$ we decide to use a finer k binning with respect of the previous case. We do so by using a very large FFT box of $L_{\rm FFT} = 16\,{\rm Gpc}/h$, which leads to a k-binning of $\Delta k = 3.9 \times 10^{-3}\,h/{\rm Mpc}$. We will do this analysis in the $z \in [0.2, 0.3]$ redshift bin. We estimate $D_{\rm iso}$ for both the field and direct approach, for all the realizations available. Since equation (1.145) does not contain any stochastic term we expect that averaging over many realization the $D_{\rm iso}$ estimator would converge to the theoretical prediction. For this reason we plot the standard error of the mean, obtained by diving the $1-\sigma$ scatter by the square root of the number of mocks. The comparison is shown in figure (3.15), where we plot in black the theoretical prediction, in red the estimation from the field approach mocks, and in blue the direct approach estimate from the 8 HugeMDPL sub-boxes. We see that the two estimates

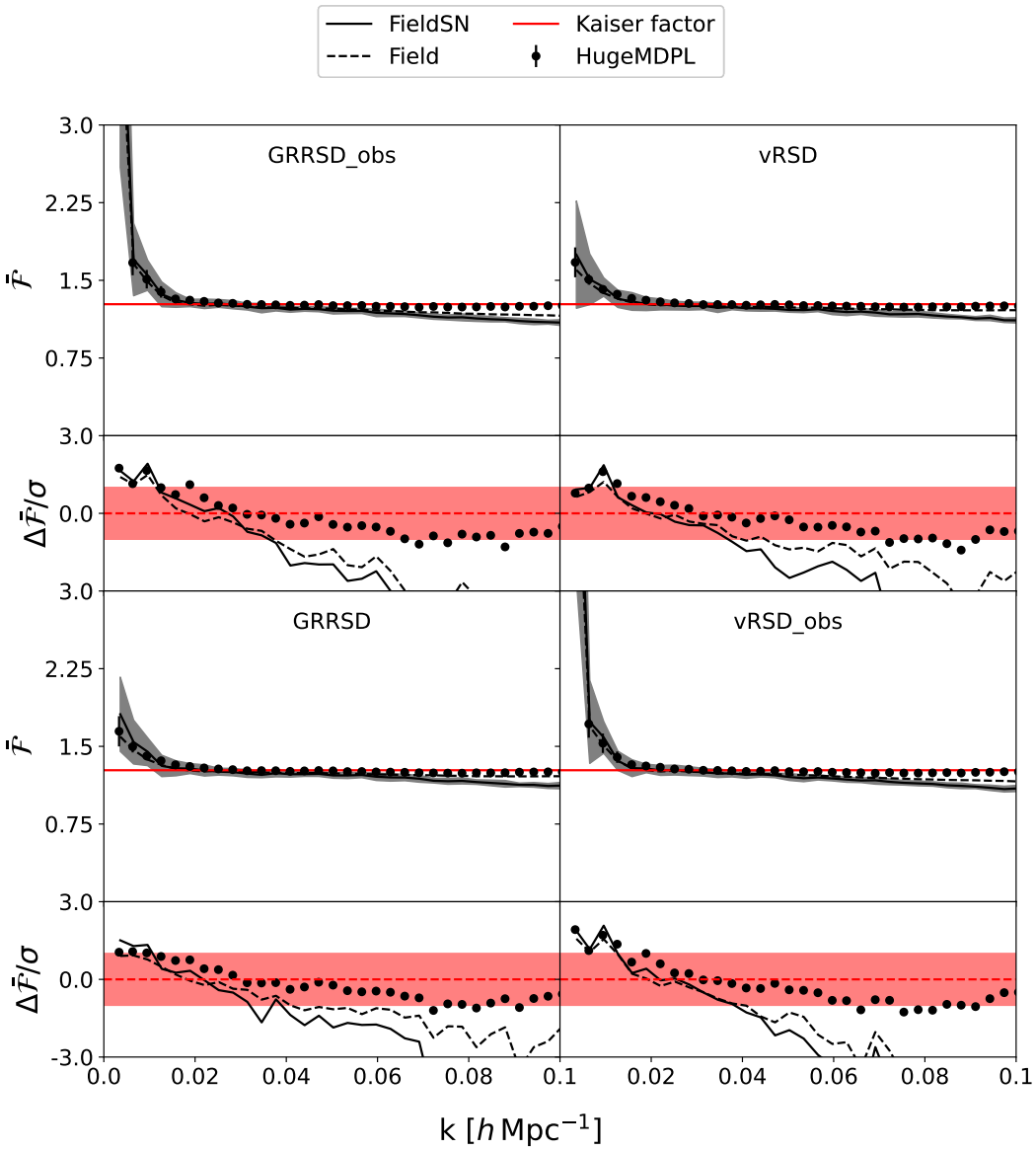


Figure 3.14: Kaiser factor estimation in the $z \in [0.2, 0.3]$ bin (in black), accounting for different contributions, and its theoretical prediction from equation (3.17) (red dashed line). We compare the direct approach result from the 8 sub boxes with their $1-\sigma$ scatter (barred points) and the field approach $1-\sigma$ scatter of 16 realizations (shaded areas and full line). We also plot the average value of the field mocks without accounting for shot-noise (dashed black line).

match the theoretical prediction at large scales, up to $k = 5 \times 10^{-3}\, h/\mathrm{Mpc}$, starting to deviate from it for larger k values. This is expected: the estimated oscillating pattern strongly depends on the binning of the radial mean density $\hat{n}(\mathbf{x})$, which can either affect the largest scales signal for a binning too coarse, or the smaller scales for a binning too fine. We also notice that as k grows the two approaches' estimates start to deviate from each other: we can attribute this to the fact that the $15\,\mathrm{Mpc}/h$ binning used to compute $\hat{n}(\mathbf{x})$ is smaller than the field approach mesh size of $31.25\,\mathrm{Mpc}/h$, and this affects the number counts in each shell, acting on the small scales behaviour of $\hat{n}$. Moreover, the survey functions used are estimated from a data set, and are subjected to noise due to the number counts and numerical derivatives implemented, so we can expect small deviations between the modelled v_{obs} effect D_{iso} and the actual one. Furthermore, the shot noise of this tracer is approximately $3000\,\mathrm{Mpc}^3/h^3$, which is comparable with the signal at the minima of D_{iso}. Lastly, as we said, the HugeMDPL mocks are generated by the same seeds for initial conditions, so we can expect the signal to not converge at the same points as the field approach, since it is not a "true" average over many realizations. Related to this issue, we also notice that the HugeMDPL error bars are smaller that the field approach one, which is still consistent with the correlated mocks explanation.

In conclusion, the D_{iso} estimators are consistent with each other at large scales, and with the theoretical prediction, which means that the direct approach correctly captures the survey functions' effects. The observer's effect could be used in the future to study the expansion history of the Universe, or to measure the observer's peculiar velocity [16]: with the approach implemented in this work we are able to account for more complex selection effects and systematics, which would allow to consider more realistic surveys.

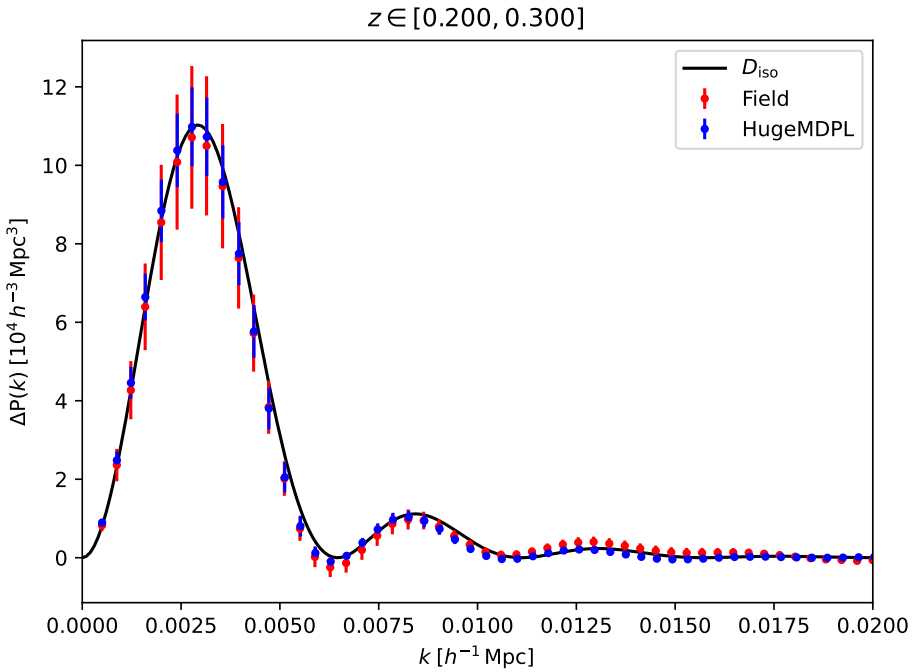


Figure 3.15: Observer's velocity contribution to the power spectrum monopole. In black we plot the theoretical prediction $D_{\rm iso}$ of equation (1.145), while in red and blue we respectively plot the estimations from the field and direct approaches, with error bars given by the error on the mean.

Chapter 4

Conclusions and prospectives

The next generation of galaxy redshift surveys will probe scales comparable to the Hubble radius. At these scales we expect to observe distortions in the measured galaxy distribution sourced by relativistic effects, which add up to the ones sourced by peculiar velocities.

The LIGER code implements a method to produce galaxy lightcone catalogues that account for these relativistic effects at first order. It operates by taking as input an N-body simulation and building upon it the catalogue, by shifting the particles in redshift space, accounting for both velocity, relativistic and magnification effects at linear order in the perturbations.

Up until this work, LIGER was used in a "field" configuration, that consists in shifting the matter particles on a low-resolution dark matter simulation, to then follow a biasing prescription to "paint" the galaxy distribution . This requires previous knowledge of the luminosity function and linear bias of the tracer we want to study, but allows to produce very large lightcones for a low computational cost, which in turn allows to produce many mock catalogues useful for covariance estimation.

However this approach presents a series of caveats that limit its range of applicability. First, as we said, it requires the previous knowledge regarding the tracer's evolution. Second, it implements basic selection effects based on a constant flux cut, without accounting for size selection and other more complicated selection criteria. Third, the painting procedure requires to build the galaxy distribution on a cubic mesh, losing all the clustering information for scales smaller than the mesh cell size. Fourth, the mesh cell size grows with the input linear bias, making Fourier-based analysis difficult for strongly biased tracers. Fifth, it implements an artificial shot-noise effect that fails as we probe smaller scales. Last, it can resolve clustering only on scales where the linear biasing relation holds.

In this project we implement and test a direct configuration for the code, in we start from a high-resolution N-body simulation and shift directly the tracers in redshift space, applying also the correct magnification. This approach solves all the issues mentioned above for the previous version, allowing to produce a realistic galaxy catalogue, to which we can apply realistic selection effects just as a real survey would do. Our work can be summarized as follows:

- We update the LIGER code by adding the option to shift directly the tracers

from a simulated catalogue. In particular this requires to implement a lookup plugin that allows to trace the tracer trajectories and their mergers, in order to correctly account for the Observer's past lightcone intersection. Moreover, this calls for a generalization of the internal routines of the code to account for the fact that the tracer trajectories can exist only for a fraction of the simulation, as opposed to dark matter particles.

- We implement an optimization in the gravitational potential storage that allows to reduce the memory occupation.
- We implement the possibility to produce cone-shaped lightcones, which for a given simulation box, allows to probe higher redshifts than in the full-sky case.
- We use the publicly available HugeMDPL simulation to build a set of mock tracers, starting from the ROCKSTAR halo catalogue attached to it and assigning a $L \propto M^4$ luminosity-mass relation.
- Since the simulation contains raw particle data only for two snapshots, we compute the large scale gravitational potential for $z = 0$ and then extrapolate it for higher redshifts by applying the evolution scheme given by the linear treatment of density perturbations.
- We use the newly implemented code to produce one deep tracer catalogue up to redshift $z = 0.7$ and 8 smaller catalogues up to $z = 0.3$, obtained by dividing the original HugeMDPL box in 8 equal parts.
- We run 16 dark matter-only low-resolution simulations using the MUSIC code.
- We estimate the survey functions that describe the mock catalogue from the snapshot data. In particular, we choose two different flux cuts for the deep and small catalogues, producing two sets of functions.
- Using the low-resolution runs and the survey functions we apply the field approach to produce 16 mock galaxy catalogues.
- We estimate the monopole of the power spectrum from the direct and field catalogues, comparing all the mocks with different contributions as listed in section (2.1). We compute this in a $z \in [0.2, 0.3]$ bin for the smaller catalogues and in a $z \in [0.6 - 0.7]$ bin for the larger one.
- We use the observer's velocity $v_{\rm obs}$ contribution on the power spectrum monopole to further test if the survey functions are correctly captured. Due to the $\propto v_{\rm obs}^2$ dependence of the generate signal, by setting $v_{\rm obs} = 1000\,{\rm km/s}$ we obtain a good signal-to-noise ratio. We compare the estimated quantities with the theoretical prediction, which depends on the magnification and evolution biases.

We find agreement in the large scale power spectrum monopole from the two different implementations for all the different contributions considered. In particular, the comparison holds up to $k \approx 0.5\,h/{\rm Mpc}$ for the close bin and $k \approx 0.4\,h/{\rm Mpc}$ for the

deep bin, For smaller scales, we notice that the artificial shot-noise implementation of the field approach generates a spurious boost in power that grows with k, while the direct implementation does not present this effect. In any case, using a field approach mock that does not account for shot-noise we get a better comparison with the direct implementation, but still with a suppressed power for large k-s, which is characteristic of the low-resolution simulations used, due to the 2LPT approximation scheme implemented by MUSIC.

We test the two implementations' accuracy in recovering the Kaiser boost factor $\bar{\mathcal{F}}$ to the power spectrum monopole in the plane-parallel approximation. We do so by computing the power spectra ratio between the redshift space and realspace mocks. First of all, we observe in both redshift bins the expected signature of wide-angle effects at very large scales, which deviates the ratio from the Kaiser boost factor. Then, we see that as k grows to $0.1\,h/\mathrm{Mpc}$, the field approach quickly drops from the theoretical value $\bar{\mathcal{F}}$. We attribute this discrepancy to the artificial shot-noise effect mentioned before, as it arises at the same scales. If we look at the direct approach instead, we see a constant trend in the powers spectra ratio, consistent at one σ with $\bar{\mathcal{F}}$. From this comparison we also assert further the consistency of the distortions implementation at very large scales, as the power spectra ratios of the direct approach mocks lie roughly within the $1-\sigma$ scatter area of the field realizations for $k < 0.02\,h/\mathrm{Mpc}$.

Looking at the observer's velocity signature on the power spectrum, we find a good agreement between the two approaches and the theoretical prediction up to $k \approx 0.004\,h/\mathrm{Mpc}$. Then the two estimations deviate from it but remain in agreement up to $k \approx 0.006\,h/\mathrm{Mpc}$.

With these test we conclude that the newly implemented approach produces consistent results with the previous implementations of LIGER, but does not suffer from the same limitations. We plan to include this code in a new version of LIGER, integrating it with the previous approach [1]. Further tests regarding the covariance of the scatter consistency could be performed by making use of a large set of high-resolution simulations, such as the Quijote suite [100].

Regarding future implementations of this now tested method, we mention a possible combination with a Halo Occupation Distribution model (HOD), or a more complex model. An HOD characterizes how galaxies populate dark matter haloes, including their spatial distribution and velocity distribution within these haloes. For a given halo one can then use an HOD to populate it with a central galaxy and satellites. This approach could be combined with the code in the following way. Starting from a simulation that resolves dark matter haloes, we can use LIGER to produce the halo catalogue in redshift space. Next, via an HOD we can populate the haloes with galaxies. Then we can correct the galaxies' redshift space distribution by adding the contributions of their peculiar velocities to their position and magnification. In conclusion, this process would generate a catalogue that:

- Accounts for relativistic effects at scales larger than the largest haloes size, i.e. in a given halo all the galaxies get the same relativistic contributions to the

[1] The code is currently available at `https://astro.uni-bonn.de/~porciani/LIGER/`

shift and magnification;

- Accounts for the peculiar velocities of the single galaxies, that contribute to both the position and magnification.

The advantage of this approach lies in the fact that using an HOD (or a more complex model) to generate the galaxy catalogue allows us to start from lower resolution N-body simulations, which are faster to run, and to shift only the haloes, which is faster to do in LIGER with respect of shifting every single galaxy.

Acknowledgements

I would like to thank Dr. Daniele Bertacca for the opportunity of working on this project and for supervising it. I extend the thanks to Prof. Cristiano Porciani for the really important insights given during this work. Finally, I thank Mohamed Yousry Elkhashab for the great help given to me throughout the whole project.

Bibliography

[1] A. J. S. Hamilton. Linear redshift distortions: A review. In *Astrophysics and Space Science Library*, pp. 185–275. Springer Netherlands, 1998.

[2] N. Kaiser. Clustering in real space and in redshift space. , 227:1–21, July 1987.

[3] A. Raccanelli et al. Doppler term in the galaxy two-point correlation function: wide-angle, velocity, doppler lensing and cosmic acceleration effects, 2016.

[4] L. Samushia et al. Interpreting large-scale redshift-space distortion measurements. *Monthly Notices of the Royal Astronomical Society*, 420(3):2102–2119, jan 2012.

[5] S. Andrianomena et al. Testing general relativity with the doppler magnification effect. *Monthly Notices of the Royal Astronomical Society*, 488(3):3759–3771, jul 2019.

[6] D. Bertacca et al. Beyond the plane-parallel and newtonian approach: wide-angle redshift distortions and convergence in general relativity. *Journal of Cosmology and Astroparticle Physics*, 2012(10):025–025, oct 2012.

[7] P. McDonald. Gravitational redshift and other redshift-space distortions of the imaginary part of the power spectrum. *Journal of Cosmology and Astroparticle Physics*, 2009(11):026–026, nov 2009.

[8] A. Challinor and A. Lewis. Linear power spectrum of observed source number counts. *Physical Review D*, 84(4), aug 2011.

[9] D. Jeong et al. Large-scale clustering of galaxies in general relativity. *Physical Review D*, 85(2), jan 2012.

[10] J. Yoo. General relativistic description of the observed galaxy power spectrum: Do we understand what we measure? *Physical Review D*, 82(8), oct 2010.

[11] J. Yoo et al. New perspective on galaxy clustering as a cosmological probe: General relativistic effects. *Physical Review D*, 80(8), oct 2009.

[12] Y. Zhang et al. Testing general relativity on cosmological scales at redshift z 1.5 with quasar and CMB lensing. *Monthly Notices of the Royal Astronomical Society*, 501(1):1013–1027, 11 2020.

[13] J. Amiaux et al. Euclidmission: building of a reference survey. *SPIE Proceedings*, Aug 2012.

[14] M. E. Levi et al. The dark energy spectroscopic instrument (desi), 2019.

[15] R. Maartens et al. Cosmology with the ska – overview, 2015.

[16] M. Y. Elkhashab et al. The large-scale monopole of the power spectrum in a euclid-like survey: wide-angle effects, lensing, and the 'finger of the observer'. *Monthly Notices of the Royal Astronomical Society*, 509(2):1626–1645, oct 2021.

[17] M. Borzyszkowski et al. liger: mock relativistic light cones from newtonian simulations. *Monthly Notices of the Royal Astronomical Society*, 471(4):3899–3914, jun 2017.

[18] N. E. Chisari and M. Zaldarriaga. Connection between newtonian simulations and general relativity. *Phys. Rev. D*, 83:123505, Jun 2011.

[19] C. Bonvin and R. Durrer. What galaxy surveys really measure. *Phys. Rev. D*, 84:063505, Sep 2011.

[20] J. Yoo. Complete treatment of galaxy two-point statistics: Gravitational lensing effects and redshift-space distortions. *Phys. Rev. D*, 79:023517, Jan 2009.

[21] R. Maartens et al. Magnification and evolution biases in large-scale structure surveys. *Journal of Cosmology and Astroparticle Physics*, 2021(12):009, dec 2021.

[22] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3):168–173, March 1929.

[23] J. R. G. III et al. A map of the universe. *The Astrophysical Journal*, 624(2):463–484, may 2005.

[24] M. Colless. First results from the 2dF Galaxy Redshift Survey. In G. Efstathiou and et al., editors, *Large-Scale Structure in the Universe*, volume 357, pp. 105, January 1999.

[25] E. Hawkins et al. The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the Universe. , 346(1):78–96, November 2003.

[26] F. Beutler et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space. , 464(3):3409–3430, January 2017.

[27] A. Lawrence et al. The QDOT all-sky IRAS galaxy redshift survey. , 308(4):897–927, October 1999.

[28] M. A. Strauss et al. A Redshift Survey of IRAS Galaxies. VII. The Infrared and Redshift Data for the 1.936 Jansky Sample. , 83:29, November 1992.

[29] J. Loveday et al. The stromlo-APM redshift survey. III. redshift space distortions, omega, and bias. *The Astrophysical Journal*, 468:1, sep 1996.

[30] S. Dodelson. *Modern Cosmology*. Academic Press, Amsterdam, 2003.

[31] D. Green et al. Messengers from the early universe: Cosmic neutrinos and other light relics, 2019.

[32] A. H. G. Peter. Dark matter: A brief review, 2012.

[33] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, January 1933.

[34] S. Perlmutter et al. Measurements of and from 42 high-redshift supernovae. *The Astrophysical Journal*, 517(2):565, jun 1999.

[35] B. P. Schmidt et al. The high-z supernova search: Measuring cosmic deceleration and global curvature of the universe using type ia supernovae. *The Astrophysical Journal*, 507(1):46–63, nov 1998.

[36] A. Achúcarro et al. Inflation: Theory and observations, 2022.

[37] L. Kofman et al. Towards the theory of reheating after inflation. *Physical Review D*, 56(6):3258–3295, sep 1997.

[38] J. A. Vazquez Gonzalez et al. Inflationary cosmology: from theory to observations. *Revista Mexicana de Física E*, 17(1 Jan-Jun):73–91, Jan. 2020.

[39] S. Dodelson. *Modern Cosmology*. 2003.

[40] R. H. Wechsler and J. L. Tinker. The Connection Between Galaxies and Their Dark Matter Halos. , 56:435–487, September 2018.

[41] A. V. Kravtsov and S. Borgani. Formation of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 50(1):353–409, sep 2012.

[42] S. Walker et al. The physics of galaxy cluster outskirts. *Space Science Reviews*, 215(1), jan 2019.

[43] B. Ryden. *Introduction to Cosmology*. Cambridge University Press, 2 edition, 2016.

[44] G. Goldhaber and D. B. Cline. The acceleration of the expansion of the universe: A brief early history of the supernova cosmology project (SCP). In *AIP Conference Proceedings*. AIP, 2009.

[45] J. A. Peacock. *Cosmological Physics*. Cambridge University Press, 1998.

[46] D. W. Hogg. Distance measures in cosmology, 2000.

[47] M. Dainotti et al. The hubble constant tension: current status and future perspectives through new cosmological probes, 2023.

[48] P. Schneider. *Extragalactic Astronomy and Cosmology.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[49] H. Mo et al. *Galaxy Formation and Evolution.* Cambridge University Press, 2010.

[50] V. Sahni. Approximation methods for non-linear gravitational clustering. *Physics Reports*, 262(1-2):1–135, nov 1995.

[51] J. Yoo et al. Going beyond the kaiser redshift-space distortion formula: A full general relativistic account of the effects and their detectability in galaxy clustering. *Phys. Rev. D*, 86:063514, Sep 2012.

[52] H. E. S. Velten et al. Aspects of the cosmological "coincidence problem". *The European Physical Journal C*, 74(11), nov 2014.

[53] K. Yamamoto et al. The cosmological light-cone effect on the power spectrum of galaxies and quasars in wide-field redshift surveys. *The Astrophysical Journal*, 527(2):488, dec 1999.

[54] Abdurro'uf et al. The seventeenth data release of the sloan digital sky surveys: Complete release of manga, mastar, and apogee-2 data. *The Astrophysical Journal Supplement Series*, 259(2):35, mar 2022.

[55] A. S. Szalay et al. Redshift-space distortions of the correlation function in wide-angle galaxy surveys. *The Astrophysical Journal*, 498(1):L1, apr 1998.

[56] S. Zaroubi and Y. Hoffman. Clustering in Redshift Space: Linear Theory. , 462:25, May 1996.

[57] R. Scoccimarro. Fast estimators for redshift-space clustering. *Physical Review D*, 92(8), oct 2015.

[58] N. Bartolo et al. Non-gaussianity from inflation: theory and observations. *Physics Reports*, 402(3-4):103–266, nov 2004.

[59] and Y. Akrami et al. iplanck/i2018 results. *Astronomy &amp Astrophysics*, 641:A7, sep 2020.

[60] J. A. Peacock and A. F. Heavens. The statistics of maxima in primordial density perturbations. , 217:805–820, December 1985.

[61] N. Kaiser. On the spatial correlations of Abell clusters. , 284:L9–L12, September 1984.

[62] H. D. Politzer and M. B. Wise. Relations between spatial correlations of rich clusters of galaxies. , 285:L1–L3, October 1984.

[63] B. Bahr-Kalus et al. The kaiser-rocket effect: three decades and counting. *Journal of Cosmology and Astroparticle Physics*, 2021(11):027, nov 2021.

[64] D. Bertacca. Generalization of the kaiser rocket effect in general relativity in the wide-angle galaxy 2-point correlation function. *International Journal of Modern Physics D*, 29(12):2050085, sep 2020.

[65] D. Bertacca. Observed galaxy number counts on the light cone up to second order: III. magnification bias. *Classical and Quantum Gravity*, 32(19):195011, sep 2015.

[66] O. Hahn and T. Abel. Multi-scale initial conditions for cosmological simulations. *Monthly Notices of the Royal Astronomical Society*, 415(3):2101–2121, jul 2011.

[67] J. Huchra et al. A survey of galaxy redshifts. IV - The data. , 52:89–119, June 1983.

[68] J. M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev. D*, 22:1882–1905, Oct 1980.

[69] R. K. Sachs and A. M. Wolfe. Perturbations of a Cosmological Model and Angular Variations of the Microwave Background. , 147:73, January 1967.

[70] I. I. Shapiro. Fourth test of general relativity. *Phys. Rev. Lett.*, 13:789–791, Dec 1964.

[71] M. Bartelmann. Gravitational lensing. *Classical and Quantum Gravity*, 27(23):233001, nov 2010.

[72] L. Hui et al. Anisotropic magnification distortion of the 3d galaxy correlation. II. fourier and redshift space. *Physical Review D*, 77(6), mar 2008.

[73] M.-A. Breton et al. Impact of lensing magnification on the analysis of galaxy clustering in redshift space. , 661:A154, May 2022.

[74] R. Maartens et al. The kinematic dipole in galaxy redshift surveys. *Journal of Cosmology and Astroparticle Physics*, 2018(01):013, jan 2018.

[75] J. Adamek et al. General relativistic<mml:math xmlns:mml="http://www.w3.org/1998/math/mathml" display="inline"><mml:mi>n</mml:mi></mml:math>-body simulations in the weak field limit. *Physical Review D*, 88(10), November 2013.

[76] E. Bentivegna and M. Bruni. Effects of nonlinear inhomogeneity on the cosmic expansion with numerical relativity. *Physical Review Letters*, 116(25), June 2016.

[77] R. W. Hockney and J. W. Eastwood. *Computer simulation using particles*. 1988.

[78] M. Heideman et al. Gauss and the history of the fast fourier transform. *IEEE ASSP Magazine*, 1(4):14–21, 1984.

[79] M. Frigo and S. G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[80] J. Amanatides and A. Woo. A Fast Voxel Traversal Algorithm for Ray Tracing. In *EG 1987-Technical Papers*. Eurographics Association, 1987.

[81] L. Pozzetti et al. Modelling the number density of hi/iemitters for future spectroscopic near-IR space missions. *Astronomy &amp Astrophysics*, 590:A3, apr 2016.

[82] L. Wenzl et al. Magnification bias estimators for realistic surveys: an application to the boss survey, 2023.

[83] P. S. Behroozi et al. THE ROCKSTAR PHASE-SPACE TEMPORAL HALO FINDER AND THE VELOCITY OFFSETS OF CLUSTER CORES. *The Astrophysical Journal*, 762(2):109, dec 2012.

[84] S. R. Knollmann and A. Knebe. AHF: AMIGA's HALO FINDER. *The Astrophysical Journal Supplement Series*, 182(2):608–624, may 2009.

[85] K. Riebe et al. The MultiDark Database: Release of the Bolshoi and MultiDark cosmological simulations. *Astronomische Nachrichten*, 334(7):691–708, August 2013.

[86] F. Prada et al. Halo concentrations in the standard Λ cold dark matter cosmology. , 423(4):3018–3030, July 2012.

[87] Y. P. Jing. Correcting for the alias effect when measuring the power spectrum using a fast fourier transform. *The Astrophysical Journal*, 620(2):559–563, feb 2005.

[88] D. Jeong. Cosmology with high (z1) redshift galaxy surveys. phd, University of Texas at Austin, djeong@astro.as.utexas.edu, 2010.

[89] H. A. Feldman et al. Power-Spectrum Analysis of Three-dimensional Redshift Surveys. , 426:23, May 1994.

[90] A. J. S. Hamilton. Power spectrum estimation i. basics, 2005.

[91] J. A. Peacock and D. Nicholson. The large-scale clustering of radio galaxies. *Monthly Notices of the Royal Astronomical Society*, 253(2):307–319, 11 1991.

[92] M. Tegmark et al. Measuring the galaxy power spectrum with future redshift surveys. *The Astrophysical Journal*, 499(2):555, jun 1998.

[93] and P. A. R. Ade et al. iplanck/i2013 results. XVI. cosmological parameters. *Astronomy &amp Astrophysics*, 571:A16, oct 2014.

[94] A. Lewis. Efficient sampling of fast and slow cosmological parameters. , 87:103529, 2013.

[95] A. Vale and J. P. Ostriker. Linking halo mass to galaxy luminosity. *Monthly Notices of the Royal Astronomical Society*, 353(1):189–200, 09 2004.

[96] M. Shirasaki et al. Virial halo mass function in the planck cosmology. *The Astrophysical Journal*, 922(1):89, nov 2021.

[97] F. Leclercq et al. One-point remapping of lagrangian perturbation theory in the mildly non-linear regime of cosmic structure formation. *Journal of Cosmology and Astroparticle Physics*, 2013(11):048–048, nov 2013.

[98] P. Schechter. An analytic expression for the luminosity function for galaxies. , 203:297–306, January 1976.

[99] D. Pryer et al. The galaxy power spectrum on the lightcone: deep, wide-angle redshift surveys and the turnover scale. *Journal of Cosmology and Astroparticle Physics*, 2022(08):019, August 2022.

[100] F. Villaescusa-Navarro et al. The quijote simulations. *The Astrophysical Journal Supplement Series*, 250(1):2, aug 2020.