Alma Mater Studiorum · Università di Bologna

School of Science
Department of Physics and Astronomy
Master Degree Programme in Astrophysics and Cosmology

Towards a joint modelling of second- and third-order clustering statistics of the Large Scale Structure

Graduation Thesis

Presented by:
Alessandro Benati

Supervisor:
Dott. Michele Moresco
Co-supervisor:
Prof. Federico Marulli
Dott. Alfonso Veropalumbo

Academic Year 2022/2023
Graduation Date III

Abstract

The next generation of galaxy redshift surveys will provide us data with unprecedented levels of precision, allowing us to extract new cosmological information from both lower and higher-order clustering statistics. For this reason, new theoretical models capable of effectively describing these data sets, as well as new statistical strategies able to reach an advanced accuracy, are required.

This Thesis project lies within this context, intending to develop the necessary framework and tools to investigate the combination of two-point correlation function (2PCF) and three-point correlation function (3PCF) of astrophysical objects. Specifically, we proposed an extension of the redshift-space bispectrum model of Scoccimarro et al., 1999 (hereafter SCF99), including the tidal bias b_t as an additional parameter, and a revision of the redshift-space 3PCF model of Slepian and Eisenstein, 2017 (hereafter SE17). After having theoretically derived the new 3PCF model and having implemented it in the *free software* C++/Python libraries for cosmological calculations CosmoBolognaLib (CBL; Marulli et al., 2016), we performed a thorough validation through a comparison with the 3PCF model used by Veropalumbo et al., 2022. After that, we explored the necessary components to develop a statistical and numerical framework for the probe combination of the 2PCF and 3PCF, a technique which allows us to reduce the statistical errors in the estimates of the cosmological parameters, while breaking some parameter degeneracies. In particular, we applied our pipeline to the 298 MINERVA mock catalogues (Grieb et al., 2016), using the extended Taruya-Nishimichi-Saito (eTNS) model for the 2PCF and our new model for the 3PCF. We carried out a Bayesian analysis separately on the 2PCF and on the 3PCF. The analysis in real space demonstates the goodness of our developed models, and shows the potential of the combination of 2PCF and 3PCF in lifting degeneracies in cosmological parameters determination and in increasing the accuracy of the constraints. The analysis in redshift space paves the way to an actual application of this kind of analysis to real data, with several interesting expansion and possible extension in view of the next large spectroscopic surveys.

Contents

Introduction

Over the last few decades, an exponential growth in astrophysical data has been facilitated by remarkable advancements in computational technology, enabling the processing of substantial amounts of information, commonly referred to as Big Data. Within this context, cosmology has entered a new precision era, benefiting from the availability of extensive astrophysical data sets (see e.g. Huterer and Shafer, 2018).

Large-scale clustering analyses of astrophysical probes, such as galaxies and galaxy clusters, has emerged as a crucial observational tool for constraining cosmological parameters (see e.g. Alam et al., 2021). Measurements of two-point statistics of the spatial distribution of these objects, either the power spectrum $P(k)$ in Fourier space or the two-point correlation function (2PCF) in configuration space, have been decisive in defining the Λ-cold dark matter (ΛCDM) model, i.e. the concordance cosmological model (see e.g. Mohammad et al., 2018; Icaza-Lizaola et al., 2019; Wang et al., 2020). This model is based on Einstein's theory of General Relativity (Einstein, 1915). It claims the existence of cold dark matter, which is non-relativistic matter that does not interact with the electromagnetic force and makes up for approximately 26% of the energy content of the Universe, and dark energy, defined as the component of the Universe responsible for its accelerated expansion, which makes up for roughly the 69% of the energy content of the Universe (Planck Collaboration et al., 2020). These two dark components completely dominate our cosmos, yet we are far from a satisfactory understanding of their nature. For this reason, they represent some of the most intriguing open questions of modern cosmology.

The large-scale structure (LSS) of the Universe contains information about the matter peculiar velocity field, which induces the so-called redshift-space distortions (RSD) on the clustering properties of matter. Furthermore, the LSS encodes information about the cosmological parameters e.g. in the form of over-densities of objects at a well-defined distance (around $100\,h^{-1}$Mpc), inherited from the initial distribution of matter in the primordial Universe. These over-densities appear in the $P(k)$ in the form of wiggles, called baryonic acoustic oscillations (BAO), and in the 2PCF in the form of a single peak. Both RSD and BAO have been demonstrated to be fundamental features to constrain the cosmological model (see e.g. Bautista et al., 2020; Gil-Marín et al., 2020; Marulli et al., 2021).

While the 2PCF and $P(k)$ can fully describe a perfectly Gaussian field, going

beyond these two-point statistics is fundamental to capture the non-Gaussian behavior of structures, which is caused by multiple factors, such as the nonlinear evolution of matter due to gravity and measurement limitations. For this reason, the moments of the density field beyond the second one are crucial, containing valuable cosmological information. As a result, n-point statistics have acquired great significance in the last decades. The simplest one is represented by three-point statistics, either in the form of the bispectrum $B(k)$ in Fourier space, or the three-point correlation function (3PCF) in configuration space, which are sensitive to deviations from Gaussianity in the primordial density perturbations. In recent years, three-point statistics arose as an essential cosmological probe (see e.g. Slepian and Eisenstein, 2017; Moresco et al., 2021). For this reason, numerous works have tackled the issue of developing theoretical models for both the 3PCF and $B(k)$. Many models for the redshift-space bispectrum have been proposed, either derived analytically (see e.g. Heavens et al., 1998; Smith et al., 2008), or through numerical simulations (see e.g. Gil-Marín et al., 2015). SCF99 derived a model for the redshift-space bispectrum using Perturbation Theory (PT) going up to second-order in the linear density field. SE17 proposed a model for the redshift-space 3PCF derived from PT, basing their analysis on the Fourier space model of SCF99. However, we found some calculation errors in the 3PCF model of SE17. For this reason, one of the major goals of this Thesis work was to develop a new 3PCF model, based on the bispectrum model of SCF99. This model has been implemented within the *CosmoBolognaLib* (CBL), a set of C++ and python libraries for cosmological calculations.

So far, very few works exploited the full potential of the combination of second- and third-order statistics, and mostly in Fourier space (see e.g. Sefusatti et al., 2006; Yankelevich and Porciani, 2019; Veropalumbo et al., 2021). These papers showed the advantages of combining both statistics, which include reducing the statistical errors when estimating cosmological parameters, and breaking some parameter degeneracies. An important goal of this project was to develop the necessary framework to exploit the joint likelihood analysis of the 2PCF and 3PCF. We used the CBL to build a computational infrastructure with the purpose of modelling the 2PCF and 3PCF and, by sampling the posterior distribution using a Monte Carlo Markov Chain (MCMC) method, to estimate the cosmological model parameters. We then proceeded by preparing the necessary environment for the joint analysis of 2PCF and 3PCF. We applied our models and pipelines on 298 mock catalogues of dark matter haloes obtained from the MINERVA N-body simulations, analyzing both real and redshift space. The real space analysis demonstrate the strength of this approach, showing how the combination of 2PCF and 3PCF can be crucial to precisely determine the bias parameters of the analyzed samples and significantly reduce the error bars on the derived parameters. The natural continuation of this work would be the analysis of the joint constraints through a MCMC analysis, but it was not possible to include the results in this work due to CPU constraints; however a full analysis is currently ongoing.

We organize our work as follows:

- in Chapter 1 we present an introduction to the cosmological framework upon which the concordance cosmological model is built;
- in Chapter 2 we introduce the concepts of 2PCF and 3PCF, escribing the cosmological information they encode;
- in Chapter 3 we discuss the model used in our analysis for the modelling of the 2PCF;
- in Chapter 4 we present the 3PCF model we developed, based on an extension of the SCF99 model for the bispectrum and a revision of the SE17 model for the 3PCF, showing our mathematical derivation;
- in Chapter 5 we describe the implementation of the codes developed in this work, and in particular the derivation of the covariance matrices for the joint 2PCF and 3PCF analysis, and the modelling of the 2PCF and 3PCF, including the addition of the new 3PCF model;
- in Chapter 6 we present and discuss the results of our cosmological analysis exploring separately the analysis of the 2PCF and of the 3PCF, and comparing their results. We conclude discussing the developed framework for the joint analysis of 2PCF and 3PCF, and presenting the obtained results.
- in Chapter 7 we summarize our results and discuss future developments of this work;
- finally, in Appendices A, B and C we discuss in more detail the implementation of new CBL functions regarding covariance matrices, 2PCF modelling, and 3PCF modelling, respectively.

Chapter 1

Cosmological background

In this chapter, we present the cosmological background, which serves as the fundamental cornerstone for the foundation of this Thesis work. First, we provide insight into the mathematical framework that underlies modern cosmology by introducing the core principles of Einstein's theory of General Relativity. Subsequently, we focus on both the geometry of the Universe, illustrating the Friedmann-Lemaître-Robertson-Walker (FLRW) metric and the Hubble-Lemaître Law, and on its content. Then, we proceed to describe the Friedmann equations, which are found by applying Einstein's field equation to a homogeneous Universe. Afterwards, the concordance ΛCDM model is introduced. Finally, we give an insight on the Jeans theory of cosmic structure formation starting from density perturbations.

1.1 Rudiments of General Relativity

At cosmological scales, the dominating interaction is gravity, which is remarkably described by General Relativity (GR), that connects the geometrical properties of spacetime to the energy content of the Universe. GR constitutes a classic field theory, in which interactions are transported at a finite velocity, the speed of light, by a field, which is spacetime itself.

GR is based on the so-called equivalence principle, which states that in a small region of the Universe the effects of gravity are indistinguishable from the effects of an acceleration. As a consequence, an experiment carried out inside an accelerated laboratory should yield the same results as if it was performed inside a laboratory subject to a gravitational field of the same measure.

A notable peculiarity of GR is that it is background independent: this represents a significant departure from Newtonian and special relativity. For this reason, GR is formulated using tensorial equations, which allow for a manifestly covariant formalism. Therefore, some first considerations concerning coordinates should be made.

Let us consider a 2-dimensional surface Σ with a point P on it and the plane

tangent to Σ in P. We can project the points of the plane onto the surface Σ, so that we can write the Cartesian local coordinates of the point P:

$$X_P^i = (X_P, Y_P). \tag{1.1}$$

If x_P^μ are arbitrary coordinates of the point P, the Jacobian of the map $X_P^i(x^\mu)$ between the two coordinate systems is:

$$e^i{}_\mu = \left.\frac{\partial X_P^i(x^\mu)}{\partial x^\mu}\right|_{x^\mu = x_P^\mu}. \tag{1.2}$$

This quantity is defined as the frame field and it captures the distance between points.

If we consider now the two points x^μ and $x^\mu + \mathrm{d}x^\mu$ on Σ, we can write:

$$\mathrm{d}X^i = \frac{\partial X^i(x^\mu)}{\partial x^\mu}\mathrm{d}x^\mu = e^i{}_\mu(x)\mathrm{d}x^\mu. \tag{1.3}$$

Then the distance between these two points is:

$$\mathrm{d}s^2 = \delta_{ij}\mathrm{d}X^i\mathrm{d}X^j = \delta_{ij}e^i{}_\mu(x)e^j{}_\nu(x)\mathrm{d}x^\mu\mathrm{d}x^\nu, \tag{1.4}$$

where we use the Einstein summation convention, that implies the summation over the repeated indices in a formula, for brevity. If we define the metric tensor as:

$$g_{\mu\nu}(x) \equiv \delta_{ij}e^i{}_\mu(x)e^j{}_\nu(x), \tag{1.5}$$

then we can write:

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\mathrm{d}x^\mu\mathrm{d}x^\nu. \tag{1.6}$$

The metric tensor $g_{\mu\nu}$ is a symmetric, hence $g_{\mu\nu} = g_{\nu\mu}$, rank $(0,2)$ tensor field and it specifies the local geometric structure of spacetime at each point.

With the metric tensor we are able to derive the length of a curve γ on the spacetime manifold:

$$L[\gamma] = \int \mathrm{d}s = \int \sqrt{g_{\mu\nu}\Big(x(\tau)\Big)\mathrm{d}x^\mu\mathrm{d}x^\nu} = \int \sqrt{g_{\mu\nu}\Big(x(\tau)\Big)\dot{x}^\mu\dot{x}^\nu}\mathrm{d}\tau, \tag{1.7}$$

where τ is the variable through which the curve is parametrized.

By means of the metric tensor we can also define angles. Indeed, we have that:

$$v^\mu \equiv e^\mu{}_i(x)v^i, \tag{1.8}$$

from which we can define the angle θ:

$$\cos\theta = \mathbf{v}\cdot\mathbf{w} = \delta_{ij}v^iw^j = g_{\mu\nu}(x)v^\mu w^\nu, \tag{1.9}$$

where equation (1.5) was used.

For simplicity we have considered the surface Σ, which is a 2-dimensional manifold, but all these considerations can be generalized to the case of an n-dimensional manifold.

The shortest path between two points in a curved space is called a geodesic. In the context of GR, it represents the path of an object subject to gravity only. We can obtain the equation of a geodesic by differenciating the length of a curve in equation (1.7):

$$\delta L[\gamma] = \int \delta \sqrt{g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} \mathrm{d}\tau. \tag{1.10}$$

After some calculations, we can derive the geodesic equation:

$$\ddot{x}^\mu + \Gamma^\mu_{\alpha\beta}\dot{x}^\alpha \dot{x}^\beta = 0, \tag{1.11}$$

where $\Gamma^\mu_{\alpha\beta}$ are the Christoffel symbols, defined as:

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\nu}\big(\partial_\alpha g_{\nu\beta} + \partial_\beta g_{\alpha\nu} - \partial_\nu g_{\beta\alpha}\big). \tag{1.12}$$

Let us consider a vector v^ν. We define its covariant derivative as:

$$\nabla_\mu v^\nu = \partial_\mu v^\nu + \Gamma^\nu_{\mu\alpha} v^\alpha. \tag{1.13}$$

On the other hand, the covariant derivative of the covariant vector w_ν is defined as:

$$\nabla_\mu w_\nu = \partial_\mu w_\nu - \Gamma^\alpha_{\mu\nu} w_\alpha. \tag{1.14}$$

In a flat, Euclidean space, we have $\Gamma^\mu_{\alpha\beta} = 0$, meaning that the covariant derivative coincides with the partial derivative. Conversely, in a curved space the Christoffel symbols are non-zero and they encode information about how tangent spaces are connected as one moves along a curve on the manifold. Indeed, if the covariant derivative of a tensor is equal to zero, it means that the vector is parallel transported along a curve on the manifold.

It can be proven that:

$$\nabla_\mu \nabla_\nu v^\alpha - \nabla_\nu \nabla_\mu v^\alpha = R^\alpha{}_{\beta\mu\nu}, \tag{1.15}$$

where $R^\alpha{}_{\beta\mu\nu}$ is the Riemann tensor, defined as:

$$R^\alpha{}_{\beta\mu\nu} = \partial_\mu \Gamma^\alpha_{\beta\nu} - \partial_\nu \Gamma^\alpha_{\beta\mu} + \Gamma^\alpha_{\mu\rho}\Gamma^\rho_{\beta\nu} - \Gamma^\alpha_{\nu\rho}\Gamma^\rho_{\beta\mu}. \tag{1.16}$$

If $R^\alpha{}_{\beta\mu\nu} = 0$ in a region of spacetime, it implies that the curvature of spacetime in that region is flat.

The main result of GR is the Einstein's field equation:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1.17}$$

where G is the gravitational constant, c is the speed of light, and $R_{\mu\nu}$ and R are the Ricci tensor and Ricci scalar, respectively, defined as:

$$R_{\mu\nu} \equiv R^\rho{}_{\mu\nu\rho}, \tag{1.18}$$

$$R = g^{\mu\nu} R_{\mu\nu}. \tag{1.19}$$

$T_{\mu\nu}$ is the energy-momentum tensor and it quantifies the matter and energy content of the Universe. Equation (1.17) implies that the curvature of the Universe, described on the left-hand side, is determined by its matter and energy content, specified on the right-hand side.

1.2 The Friedmann-Lemaître-Robertson-Walker metric

Cosmology is founded on the so-called Cosmological Principle, which states that the Universe is homogeneous and isotropic at large scales (tipically hundreds of Mpc). In other words, there is no preferred position or direction in the Universe.

If we consider two events in spacetime (t_1, x_1, y_1, z_1) and (t_2, x_2, y_2, z_2), their distance is:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = g_{00}\mathrm{d}t^2 + 2g_{0i}\mathrm{d}t\mathrm{d}x^i + g_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{1.20}$$

where $g_{00}\mathrm{d}t^2$ is the time component, $g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ is the spatial component (i and j are the spatial indices and they can assume the values of $\{1, 2, 3\}$) and $2g_{0i}\mathrm{d}t\mathrm{d}x^i$ is the mixed component. As a consequence of the Cosmological Principle, the mixed component is equal to zero. For light, we have $\mathrm{d}s^2 = 0$, which implies that $g_{00} = c^2$, where c is the speed of light. Therefore, we can write:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 + g_{ij}\mathrm{d}x^i\mathrm{d}x^j = c^2\mathrm{d}t^2 - \mathrm{d}l^2. \tag{1.21}$$

The line element $\mathrm{d}l^2$ depends on the geometry of the Universe. If we consider the spacetime as a plane the line element is:

$$\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2, \tag{1.22}$$

which in polar coordinates, becomes:

$$\mathrm{d}l^2 = \mathrm{d}\rho^2 + \rho^2\mathrm{d}\phi^2, \tag{1.23}$$

where $0 \leq \rho < \infty$ and $0 \leq \phi \leq 2\pi$. If we define $\rho \equiv ar$, where a has the dimensions of a length and r is dimensionless:

$$\mathrm{d}l^2 = a^2\Big[\mathrm{d}r^2 + r^2\mathrm{d}\phi^2\Big]. \tag{1.24}$$

If we consider the spacetime as a spherical surface of radius R, then we have:

$$\mathrm{d}l^2 = R^2\Big[\sin^2\theta\,\mathrm{d}\phi^2 + \mathrm{d}\theta^2\Big], \tag{1.25}$$

where $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$. If we define $R \equiv a$ and $\sin\theta \equiv r$, then:

$$\mathrm{d}l^2 = a^2\Big[r^2\mathrm{d}\phi^2 + \frac{1}{1-r^2}\mathrm{d}r^2\Big]. \tag{1.26}$$

If we consider the spacetime as a hyperbolic surface of radius R, then we have:

$$\mathrm{d}l^2 = R^2\Big[\sinh^2\theta\,\mathrm{d}\phi^2 + \mathrm{d}\theta^2\Big]. \tag{1.27}$$

If we define $R \equiv a$ and $\sinh\theta \equiv r$, then:

$$\mathrm{d}l^2 = a^2\Big[r^2\mathrm{d}\phi^2 + \frac{1}{1+r^2}\mathrm{d}r^2\Big]. \tag{1.28}$$

Therefore, we can write in general:

$$\mathrm{d}l^2 = a^2\Big[r^2\mathrm{d}\phi^2 + \frac{1}{1-kr^2}\mathrm{d}r^2\Big], \tag{1.29}$$

where k can only assume the values of 0, +1 or −1.

We can move to 3-dimensions by replacing the angle ϕ in equation (1.29) with the solid angle Ω, knowing that $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$:

$$\mathrm{d}l^2 = a^2\Big[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\big(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\big)\Big]. \tag{1.30}$$

Thus, we can write the distance in the 4-dimensional spacetime as:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a^2(t)\Big[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\big(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\big)\Big], \tag{1.31}$$

also known as the FLRW metric. The parameter a is called scale factor and it encodes information about the expansion of volumes in the Universe, while the parameter k is called curvature parameter and, depending on the geometry of spacetime, it can assume three values:

- $k = 0$: flat geometry (flat Universe);
- $k = +1$: spherical geometry (closed Universe);
- $k = -1$: hyperbolic geometry (open Universe).

We define the proper distance as the distance between two points when $\mathrm{d}t = 0$ and by rotating the reference frame so that $\mathrm{d}\theta = \mathrm{d}\phi = 0$, we can write:

$$d_{pr}(t) = a(t)\int_0^r \frac{dr'}{\sqrt{1-kr'^2}} = a(t)\chi(r), \tag{1.32}$$

where the first term contains the temporal part and the second term contains the spatial one. Depending on the geometry of the Universe the spatial term $\chi(r)$ can assume different values:

$$\chi(r) = \begin{cases} r, & k = 0 \\ \arcsin r, & k = +1 \\ \operatorname{arcsinh} r, & k = -1 \end{cases}. \tag{1.33}$$

The comoving distance is the proper distance computed at the present time t_0:

$$d_c = \frac{a_0}{a(t)}\, d_{pr}(t), \tag{1.34}$$

where $a_0 \equiv a(t_0)$. A comoving reference frame is a reference frame which is comoving with the variation of $a(t)$, so with the expansion of the Universe.

Two other quantities can be introduced: the particle horizon and the event horizon. The particle horizon is defined as the largest comoving distance from

which a photon emitted by a particle can have travelled to the observer in the age of the Universe, so it sets the size of the observable Universe:

$$R_{ph}(t) = a(t) \int_0^t \frac{c\,\mathrm{d}t'}{a(t')}. \tag{1.35}$$

On the other hand, the event horizon identifies the largest comoving distance that a photon emitted now can ever reach an observer in the future, then it sets the maximum size reached by the Universe in the future (potentially it could be infinite):

$$R_{eh}(t) = a(t) \int_t^{t_{end}} \frac{c\,\mathrm{d}t'}{a(t')}. \tag{1.36}$$

1.3 Hubble-Lemaître Law and redshift

The expansion of the Universe causes the scale factor a to change as a function of cosmic time. As a consequence, also the proper distance separating two points grows over time, and by calculating its time derivative, we can deduce the radial velocity between these points:

$$v_r(t) = \frac{\mathrm{d}}{\mathrm{d}t} d_{pr}(t) = \dot{a}(t) F(r) = \frac{\dot{a}}{a} d_{pr}(t). \tag{1.37}$$

Therefore, the radial velocity is proportional to the distance, meaning that points that are more distant from one another recede from each other with greater velocity. By defining the Hubble parameter as $H \equiv \dot{a}/a$, we can write the well-known Hubble-Lemaître Law:

$$v_r(t) = H(t)\, d_{pr}(t). \tag{1.38}$$

If we compute the Hubble parameter at the present time we obtain the Hubble function $H_0 \equiv H(t_0)$. Its value, as measured e.g. by Wang et al., 2017 using the latest baryonic acoustic oscillation measurements from the eBOSS survey, is:

$$H_0 = (67.27 \pm 1.55)\,\mathrm{km\,s^{-1}Mpc^{-1}}. \tag{1.39}$$

Conventionally, we introduce the dimensionless Hubble parameter h as follows:

$$H_0 \equiv 100\,h\,\mathrm{km\,s^{-1}Mpc^{-1}}. \tag{1.40}$$

The inverse of the Hubble constant represents a first-order approximation for the age of the Universe. It is important to underline that one of the open questions of modern cosmology is the so-called Hubble tension, which refers to a discrepancy in the value of H_0 measured in the late and early Universe (for a detailed review see Verde et al., 2019; Abdalla et al., 2022). This inconsistency might potentially be an indication of the existence of new physics.

The expansion of the Universe causes objects to recede from us, but this does not mean that we are the center of the expansion. Indeed, the cosmological principle states that there is not a preferred position in the Universe.

Consequently, from any point of spacetime we would see objects receding from us. This implies that the observed spectrum of such receding objects undergoes a reddening, caused by the shift towards longer wavelengths. The so-called redshift is defined as:

$$z = \frac{\lambda_{\rm obs} - \lambda_{\rm em}}{\lambda_{\rm em}} = \frac{\Delta\lambda}{\lambda}, \tag{1.41}$$

where $\lambda_{\rm obs}$ is the observed wavelength and $\lambda_{\rm em}$ is the emitted wavelength. Since the objects are moving away from us, $\lambda_{\rm obs}$ is always greater than $\lambda_{\rm em}$, meaning that the redshift z is always positive.

Let us consider a photon which is emitted at time $t_{\rm em}$ with wavelength $\lambda_{\rm em}$ and is observed at time $t_{\rm obs}$ with wavelength $\lambda_{\rm obs}$. Since photons move along null geodesics by definition, we have $\mathrm{d}s^2 = 0$ and by choosing a suitable reference frame we also have $\mathrm{d}\theta = \mathrm{d}\phi = 0$, then:

$$\int_{t_{\rm em}}^{t_{\rm obs}} \frac{c\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}} = F_k(r). \tag{1.42}$$

We now introduce a second photon, emitted at time $t_{\rm em} + \delta t_{\rm em}$ with wavelength $\lambda_{\rm em}$ and observed at time $t_{\rm obs} + \delta t_{\rm obs}$ with wavelength $\lambda_{\rm obs}$. Analogously, we can write:

$$\int_{t_{\rm em}+\delta t_{\rm em}}^{t_{\rm obs}+\delta t_{\rm obs}} \frac{c\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}} = F_k(r), \tag{1.43}$$

since the spatial part remains unchanged. Therefore, if $\delta t_{\rm em} \ll 1$ and $\delta t_{\rm obs} \ll 1$, so that $a(t)$ can be considered roughly constant:

$$\frac{\delta t_{\rm obs}}{a_{\rm obs}} = \frac{\delta t_{\rm em}}{a_{\rm em}}, \tag{1.44}$$

from which, knowing that $\delta t \propto 1/\nu$ and $\lambda\nu = c$, we derive:

$$\frac{a_{\rm em}}{\lambda_{\rm em}} = \frac{a_{\rm obs}}{\lambda_{\rm obs}}. \tag{1.45}$$

Remembering the definition of redshift in equation (1.41), we obtain:

$$\frac{a_{\rm obs}}{a_{\rm em}} = 1 + z. \tag{1.46}$$

If we consider an observer at present time and a photon emitted at time t we find the more general relation:

$$\frac{a_0}{a(t)} = 1 + z. \tag{1.47}$$

These considerations regarding distances and redshift will be extremely useful to understand the distortions in the clustering properties of cosmic tracers (see section 2.3).

1.4 Content of the Universe

From now on we simplify the notation by using units in which:

$$c = \hbar = k_B = 1, \tag{1.48}$$

where $\hbar$ is the reduced Planck constant and k_B is the Boltzmann constant.

After having familiarized with the geometry of spacetime, contained in the left-hand side of equation (1.17), we now turn to the content of the Universe, described by the right-hand side of the same equation. Let us consider the energy-momentum tensor $T^{\mu\nu}$. We consider the content of the Universe to be a perfect fluid, in which the particles have a mean free path much smaller than their physical scales of interaction. As a consequence, we can write the energy-momentum tensor as:

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}, \tag{1.49}$$

where ρ is the energy density, p is the pressure and u^i is the i-th component of the 4-velocity of the fluid element. Knowing that in Cartesian comoving coordinates:

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix} \tag{1.50}$$

and:

$$T^\mu{}_\nu = g_{\alpha\nu}T^{\mu\alpha}, \tag{1.51}$$

we derive that:

$$T^\mu{}_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{1.52}$$

If we take the covariant derivative of the energy-momentum tensor we obtain the following condition:

$$\nabla_\mu T^\mu{}_\nu = 0, \tag{1.53}$$

which is the GR version of the continuity and Euler equations. Equation (1.53) actually consists of 4 equations since $\nu = 0, 1, 2, 3$. In particular, if we take the $\nu = 0$ component we obtain:

$$\frac{\partial T^\mu{}_0}{\partial x^\mu} + \Gamma^\mu_{\alpha\mu}T^\alpha{}_0 - \Gamma^\alpha_{0\mu}T^\mu{}_\alpha = 0. \tag{1.54}$$

We know that $T^i{}_0 = 0$ since we are assuming isotropy, consequently the μ index in the first term and the α index in the second term must be equal to zero:

$$-\frac{\partial \rho}{\partial t} - \Gamma^\mu_{0\mu}\rho - \Gamma^\alpha_{0\mu}T^\mu{}_\alpha = 0. \tag{1.55}$$

Knowing that the only non-zero components of the Christoffel symbols are:

$$\Gamma^0_{ij} = \delta_{ij}\dot{a}a, \tag{1.56}$$

$$\Gamma^i_{0j} = \Gamma^i_{j0} = \delta_{ij}\frac{\dot{a}}{a}, \tag{1.57}$$

we can conclude that $\Gamma^\mu_{0\mu} = 0$ and $\Gamma^\alpha_{0\mu}$ does not vanish only in the case in which α and μ are spatial indices. Then we obtain:

$$\frac{\partial\rho}{\partial t} + \frac{\dot{a}}{a}\big[3\rho + 3p\big] = 0, \tag{1.58}$$

which can be rewritten as:

$$a^{-3}\frac{\partial[\rho a^3]}{\partial t} = -3\frac{\dot{a}}{a}p. \tag{1.59}$$

We can now define the equation of state parameter w_s:

$$w_s \equiv \frac{p_s}{\rho_s}, \tag{1.60}$$

where the subscript s refers to a generic component of the Universe. Recalling that the sound velocity is defined as:

$$v_s^2 = \left(\frac{\partial p}{\partial \rho}\right)\Bigg|_{S=const}, \tag{1.61}$$

where S is the entropy, we deduce that, since v_s must be smaller than c:

$$0 \leq w_s < 1. \tag{1.62}$$

We can integrate equation (1.58) and, by assuming that w_s does not change with time, we derive:

$$\rho_s(a) \propto a^{-3(1+w_s)}. \tag{1.63}$$

Therefore, depending on the value of w_s, we have different evolutions of the different components of the Universe:

- matter ($w_m = 0$): $\rho_m \propto a^{-3} \propto (1+z)^3$, which is not surprising as the matter energy density scales like a cubic length, due to the expansion of volumes in the Universe;
- radiation ($w_r = 1/3$): $\rho_r \propto a^{-4} \propto (1+z)^4$, in which an extra $(1+z)$ term with respect to the matter case is present, which can be understood as due to the shift of the photon's energy towards greater wavelengths.

Understanding the evolution of matter with cosmic time is essential to study its clustering properties, described in chapter 2.

1.5 Friedmann equations

We have already discussed about Einstein's field equation in section (1.1), deriving equation (1.17). If we apply it to the special case of a homogeneous Universe, we can derive the so-called Friedmann equations, which are fundamental in cosmology. Let us begin with the calculation by assuming the FLRW metric and $k = 0$. We can conveniently write the Ricci tensor, defined in equation (1.18), as:

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} - \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{\mu\nu} - \Gamma^{\alpha}_{\beta\nu}\Gamma^{\beta}_{\mu\alpha}, \tag{1.64}$$

where the comma indicates a derivative with respect to x, so that $\Gamma^{\alpha}_{\mu\nu,\alpha} \equiv \partial\Gamma^{\alpha}_{\mu\nu}/\partial x^{\alpha}$.

We can derive the time-time component of $R_{\mu\nu}$:

$$R_{00} = \Gamma^{\alpha}_{00,\alpha} - \Gamma^{\alpha}_{0\alpha,0} - \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{00} - \Gamma^{\alpha}_{\beta 0}\Gamma^{\beta}_{0\alpha}. \tag{1.65}$$

Recalling that the Christoffel symbol vanishes when the two lower indices are equal to zero, the first and third term of the right-hand side of the previous equation are cancelled; for the same reason, the indices α and β in the second and fourth terms must be spatial. Then:

$$R_{00} = -\Gamma^{i}_{0i,0} - \Gamma^{i}_{j0}\Gamma^{j}_{0i}, \tag{1.66}$$

which, if we use equation (1.57), becomes:

$$\begin{aligned} R_{00} = &- \delta_{ij}\frac{\partial}{\partial t}\left(\frac{\dot{a}}{a}\right) - \left(\frac{\dot{a}}{a}\right)^2 \delta_{ij}\delta_{ij} \\ &- 3\left(\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}\right) - 3\left(\frac{\dot{a}}{a}\right)^2 \\ &- 3\frac{\ddot{a}}{a}, \end{aligned} \tag{1.67}$$

where in the second line we have used the fact that $\delta_{ij}\delta_{ij} = \delta_{ii}$ and after summing we get a 3 factor. We can now derive the spatial part of $R_{\mu\nu}$:

$$R_{ij} = \delta_{ij}\Big[2\dot{a}^2 + a\ddot{a}\Big]. \tag{1.68}$$

The Ricci scalar, defined in equation (1.19), becomes:

$$R = g^{\mu\nu}R_{\mu\nu} = g^{00}R_{00} + g^{ij}R_{ij} = -R_{00} + \frac{1}{a^2}R_{ij}, \tag{1.69}$$

where we have used the fact that we have the FLRW metric (equation 1.50). Then:

$$R = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2\right]. \tag{1.70}$$

Therefore, if we consider the time-time component of equation (1.17), we obtain:

$$R_{00} - \frac{1}{2} g_{00} R = 8\pi G T_{00}. \tag{1.71}$$

Knowing that the energy-momentum tensor is defined in equation (1.49), this becomes:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \tag{1.72}$$

which is the first Friedmann equation.

Similarly, we can derive the second Friedmann equation by considering the space-space component of equation (1.17):

$$R_{ij} - \frac{1}{2} g_{ij} R = 8\pi G T_{ij}, \tag{1.73}$$

which becomes:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + 3p\right). \tag{1.74}$$

1.6 ΛCDM model

The ΛCDM model is the concordance cosmological model and it describes the composition and the evolution of the Universe. It provides a theoretical framework that matches a wide range of cosmological observations and experiments, making it a cornerstone of modern cosmology. Estimates of cosmological parameters, such as the one performed in this work (see chapter 6), try to improve our constraints and to test the validity of the ΛCDM model.

As the name suggests, according to this model there are two main constituents of the Universe, which completely dominate its energy content: dark energy, represented by the cosmological constant Λ, which is the component of the Universe responsible for its accelerated expansion, and cold dark matter (CDM), which stands for non-relativistic matter that is not visible since it does not interact with the electromagnetic force.

With the aim of having the energy densities of the constituents of the Universe in the same units, we define the density parameter of the species s as:

$$\Omega_s \equiv \frac{\rho_s(t_0)}{\rho_c}, \tag{1.75}$$

where the critical density ρ_c is defined as:

$$\rho_c \equiv \frac{3H_0^2}{8\pi G}. \tag{1.76}$$

As a consequence, we can write the evolution of the energy density of the species s as:

$$\rho_s(a) = \Omega_s \rho_c a^{-3(1+w_s)}. \tag{1.77}$$

The species, referred to with the subscript s, which we consider are radiation (γ), referring to relativistic particles like photons and neutrinos, baryons (b), CDM (c) and dark energy (Λ). In addition to these, we indicate all the non-relativistic matter components with the subscript m, so that we have $\Omega_m = \Omega_b + \Omega_c$; moreover we denote all the relativistic components with the subscript r, which include photons, relativistic neutrinos and relativistic matter. To be independent of the value of the Hubble constant H_0, when we make a measurement we are constraining $\omega_s \equiv \Omega_s h^2$, rather than Ω_s only, since the h^2 term contains our ignorance regarding H_0.

Observations suggest that today $\Omega_{\rm tot} \simeq 1$, which means that the energy density of the Universe today is consistent to the critical density ρ_c and that, consequently, the Universe is flat. The contributions of the various components of the Universe can be summarized as:

- $\Omega_\gamma \approx 10^{-5}$;
- $\Omega_b \approx 0.05$;
- $\Omega_c \approx 0.26$;
- $\Omega_\Lambda \approx 0.69$.

1.7 Jeans theory

In this section, we consider the gravitational instabilities which lead to the formation of cosmic structures, described by the so-called Jeans theory. Understanding the formation of these structures is fundamental to analyze their clustering properties, described by the 2PCF and 3PCF (see sections 2.1 and 2.2), which are central concepts in this work.

Jeans instabilities are thought to be driven by density fluctuations, tiny oscillations in the density of the primordial fluid. The Universe at Mpc scales appears quite inhomogeneous today, since we can observe, for instance, galaxies, galaxy clusters and other astrophysical objects, implying that the evolution today is highly non-linear. However, the early Universe was extremely homogeneous at such scales, with oscillations in density which can be derived from the oscillations in temperature observed in the CMB, of the order of:

$$\frac{\delta T}{T} \approx 10^{-5}. \tag{1.78}$$

Jeans theory provides an analytic explanation for the growth of density perturbations as long as they remain linear, while their non-linear evolution can only be described by numerical simulations.

1.7.1 Jeans theory in a static Universe

Let us consider a homogeneous, isotropic and static, hence non-expanding, Universe, filled with a perfect fluid with a constant matter density $\rho(\mathbf{x}, t)$. Adopting

a Newtonian approach, we can write the following equations of motion of such fluid:

$$\begin{cases} \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \\ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \cdot \mathbf{v} = -\frac{1}{\rho} \nabla p - \nabla \Phi \\ \nabla^2 \Phi = 4\pi G \rho \\ p = p(\rho, S) = p(\rho) \\ \frac{\mathrm{d}S}{\mathrm{d}t} = 0 \end{cases} , \tag{1.79}$$

where $\mathbf{v}$ is the velocity of a fluid element, Φ is the gravitational potential and S is the entropy. The first equation of the system (1.79) is the continuity equation, the second one is the Euler equation, the third one is the Poisson equation, the fourth one is the equation of state and the last one is the adiabatic condition, which is assumed since observations suggest that the majority of density fluctuations are adiabatic. Let us assume to know the exact solutions for the background, which, since the Universe is static, can be written as:

$$\begin{cases} \rho = \rho_B = \text{const} \\ p = p_B = \text{const} \\ \mathbf{v} = 0 \\ \Phi = \Phi_B = \text{const} \end{cases} . \tag{1.80}$$

Before proceeding it is worth mentioning that the latter set of equations is not correct since the Poisson equation implies that if $\Phi_B = \text{const}$, then $\rho_B = 0$. Despite this inconsistency here, we will not encounter any problem, since it will not be present in the expanding Universe case, considered in section 1.7.2. Let us introduce now some small perturbations to the background solution:

$$\begin{cases} \rho = \rho_B + \delta\rho \\ p = p_B + \delta p \\ \mathbf{v} = \delta \mathbf{v} \\ \Phi = \Phi_B + \delta\Phi \end{cases} , \tag{1.81}$$

where $\delta\rho$, δp, $\delta\mathbf{v}$, $\delta\Phi \ll 1$, so that we can neglect all the terms beyond the linear ones. We can now define the density contrast $\delta(\mathbf{x}, t)$ as:

$$\delta(\mathbf{x}, t) \equiv \frac{\delta\rho(\mathbf{x}, t)}{\rho_B} \equiv \frac{\rho(\mathbf{x}, t) - \rho_B}{\rho_B}, \tag{1.82}$$

implying that the first equation of the system (1.81) can be written as:

$$\rho = \rho_B(1 + \delta). \tag{1.83}$$

Now we can substitute the solutions in (1.81) into the system (1.79) and, moving to Fourier space considering plane waves as solutions, we obtain a set of three equations in three variables:

$$\begin{cases} \omega\delta_k + k\delta v_k = 0 \\ \omega\delta v_k = -v_s^2 k\delta_k - k\delta\Phi_k \\ -k^2\delta\Phi_k = 4\pi G\rho_B\delta_k \end{cases} , \tag{1.84}$$

where ω is the angular frequency of the wave, $k \equiv |\mathbf{k}|$ is the wave vector, δ_k, δv_k, $\delta\Phi_k$ are the Fourier transforms of δ, $\mathbf{v}$, Φ respectively and $v_s \equiv \sqrt{\partial p/\partial\rho}$ is the sound velocity. By requiring that the Jacobian of the system is equal to zero, we get the dispersion relation:

$$\omega^2 = k^2 v_s^2 - 4\pi G\rho_B. \tag{1.85}$$

By imposing that $\omega^2 = 0$, we find:

$$k_J = \sqrt{\frac{4\pi G\rho_B}{v_s^2}}, \tag{1.86}$$

which is the wave vector corresponding to the so-called Jeans scale λ_J:

$$\lambda_J = \frac{2\pi}{k_J} = v_s\sqrt{\frac{\pi}{G\rho_B}}. \tag{1.87}$$

We can discriminate between two cases:

- $\lambda < \lambda_J$: equation (1.85) can be rewritten as:

$$\omega^2 = k^2 v_s^2 \left[1 - \left(\frac{\lambda}{\lambda_J}\right)^2\right]. \tag{1.88}$$

  In this case we have $\omega^2 > 0$, implying that we have 2 real solutions:

$$\omega_\pm = \mp k v_s \sqrt{\left[1 - \left(\frac{\lambda}{\lambda_J}\right)^2\right]}. \tag{1.89}$$

  If we substitute these solutions into the wave equation:

$$\rho(\mathbf{r}, t) = \delta\rho_k e^{i\omega t + i\mathbf{kr}}, \tag{1.90}$$

  we can see that the temporal part remains imaginary, therefore the amplitude does not change with time. This implies that these solutions represent plane waves which propagate, but do not grow with time.

- $\lambda > \lambda_J$: equation (1.85) can be rewritten as:

$$\omega^2 = -4\pi G\rho_B \left[1 - \left(\frac{\lambda_J}{\lambda}\right)^2\right]. \tag{1.91}$$

  In this case we have $\omega^2 < 0$, implying that we have 2 imaginary solutions:

$$\omega_\pm = \mp i\sqrt{4\pi G\rho_B \left[1 - \left(\frac{\lambda_J}{\lambda}\right)^2\right]} = \mp|\omega|. \tag{1.92}$$

If we substitute these solutions into equation (1.90) we obtain:

$$\delta(\mathbf{r}, t) = \delta\rho_k e^{\mp|\omega|t} e^{i\mathbf{kr}}, \tag{1.93}$$

implying that the amplitude changes with time. Therefore, we have one exponentially growing solution and one exponentially decaying solution.

Thus, we can conclude that in the case of a static Universe, the perturbations with a scale larger than the Jeans scale λ_J grow exponentially.

1.7.2 Jeans theory in an expanding Universe

Let us now consider a homogeneous, isotropic and expanding Universe. This implies that in this case the energy density of the background depends on time, $\rho_B = \rho_B(t)$. Now that the Universe is expanding, it becomes necessary to distinguish between the proper distance, denoted as r, and the comoving distance, represented by the variable x. Their relation is the following:

$$r = ax. \tag{1.94}$$

If we derive the proper distance r with respect to time, we get the velocity u:

$$u = \frac{\mathrm{d}r}{\mathrm{d}t} = \dot{a}x + a\dot{x} = Hr + a\dot{x} = Hr + v, \tag{1.95}$$

where Hr is the Hubble flow, due to the expansion of the Universe, and v is the peculiar velocity, due to gravity. We can write the exact solutions of the system (1.79) for the background as:

$$\begin{cases} \rho = \rho_B(t) \\ p = p_B \\ \mathbf{u} = H\mathbf{r} \\ \Phi = \Phi_B \end{cases} . \tag{1.96}$$

Here we no longer have the problem of incoherency we have encountered in section 1.7.1 when we have wrote equation (1.80). Now we proceed using the same strategy we adopted in section 1.7.1, therefore we introduce small perturbations to these background solutions, and then we substitute them into equation (1.79). Then, if we move to Fourier space and we search for plane waves, we can obtain the following dispersion relation:

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k + \delta_k \left[k^2 v_s^2 - 4\pi G\rho_B\right] = 0, \tag{1.97}$$

where the second term is related to the Hubble friction, due to the expansion of the Universe, and the $\delta_k k^2 v_s^2$ term accounts for the characteristic velocity field of the considered fluid. Both these terms have the effect of damping the growth of the density fluctuations. Equation (1.97) is valid for both baryons and DM,

with the only difference being that when we are considering DM we substitute the sound velocity v_s with the velocity dispersion, since DM is assumed to be non-collisional. We define the Jeans scale as:

$$\lambda_J \simeq v_s \sqrt{\frac{\pi}{G\rho_B}}. \tag{1.98}$$

Similarly to the static Universe case, for $\lambda < \lambda_J$ the perturbation does not grow and it propagates as a stationary wave. On the other hand, for $\lambda > \lambda_J$ the dispersion relation has a growing and a decaying solution:

$$\delta(\mathbf{x}, t) = A(\mathbf{x})\delta_+(t) + B(\mathbf{x})\delta_-(t), \tag{1.99}$$

where A and B are two functions depending on the comoving coordinate x, while δ_+ and δ_- are the growing and decaying solutions, respectively. We are interested to the growing solution, since it is the one which gives rise to the gravitational instabilities, and its expression for a generic Universe can be written as:

$$\delta_+(t) = H(t) \int \frac{\mathrm{d}t}{a^2 H^2(t)}. \tag{1.100}$$

The integral in equation 1.100 does not have an analytic solution, but we can write an approximated parametric solution:

$$f \equiv \frac{\mathrm{d}\log\delta_+}{\mathrm{d}\log a} = \Omega_M^{0.55} + \frac{\Omega_\Lambda}{70}\left(1 + \frac{1}{2}\Omega_{0,M}\right). \tag{1.101}$$

This equation suggests that the growth rate f has a strong dependence on the matter density parameter Ω_M and a relatively small dependence on the density parameter of Λ, meaning that Λ does not play a relevant role in the growth of density fluctuations. Moreover, the exponent of Ω_M, 0.55, is a prediction of GR, consequently its measurement is extremely useful to put understand whether GR is still valid at cosmological scales.

Chapter 2

Statistical properties of the large-scale structure

Statistical tools are crucial in cosmology, since they provide the mathematical and computational framework necessary to analyze, interpret, and make predictions about the Universe based on observational data, while accounting for uncertainties and the inherent complexity of cosmological phenomena. Over the last few decades, galaxy clustering emerged as a pivotal cosmological probe, becoming fundamental for deriving cosmological constraints. This chapter begins with a description of the Bayesian analysis, a fundamental approach to infer model parameters, and of the Monte Carlo Markov Chain method, which is an extensively exploited approach for sampling probability distributions. Then, we delineate the concepts of two-point statistics, namely the 2PCF in configuration space and the power spectrum in Fourier space, and three-point statistics, that is the 3PCF in configuration space and the bispectrum in Fourier space. After that, we introduce the dynamical distortions, generated when neglecting the peculiar velocities of the sources, and geometrical distortions, produced when considering a cosmology different from the true one. These clustering distortions are crucial for the definition of the new model for the 3PCF presented in chapter 4 and for the cosmological parameter estimation discussed in chapter 6.

Assuming the cosmological principle, let us consider the density contrast field of the Universe $\delta(\mathbf{x})$, defined in equation (1.82), to be homogeneous and isotropic at cosmological scales. This implies that $\delta(\mathbf{x})$ has the same statistical properties in all points of the Universe and at all directions. Since the density contrast field is produced at the end of the inflationary phase by statistical fluctuations of the metric, we do not expect to have phase correlations between different points, implying that the field is assumed to be stochastic. Therefore, the probability field $P(\delta)$ of finding a perturbation δ at a certain point is assumed to be, initially, almost Gaussian. Since, by definition, the Universe is unique, the so-called ergodic hypothesis has a crucial importance. It states that the statistical average over many volumes is equivalent to the spatial average over

sufficiently large subregions of the same volume. This implies that we can average over different regions of the Universe, given that they are sufficiently large, so that they can be considered as statistically independent. Today, a fair sample in which we have homogeneity and isotropy is reached at scales $\gtrsim 100$ Mpc. Moreover, in the case of a Gaussian distribution, the ergodic hypothesis becomes a theorem.

We can write the expression of the $\delta(\mathbf{x})$ in configuration space through the Fourier anti-trasform:

$$\delta(\mathbf{x}) = \frac{1}{(2\pi)^3} \int \delta(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mathbf{k}, \tag{2.1}$$

while the expression of the $\delta(\mathbf{k})$ in Fourier space can be written using the Fourier transform of (2.1):

$$\delta(\mathbf{k}) = \int \delta(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mathbf{x}, \tag{2.2}$$

where $\delta(\mathbf{x})$ is adimensional, whereas $\delta(\mathbf{k})$ has the dimensions of a cubic length.

2.1 Two-point correlation function and power spectrum

The 2PCF, $\xi(\mathbf{r})$, is defined as follows:

$$\xi(r) \equiv \langle \delta(\mathbf{x}) \delta(\mathbf{x}+\mathbf{r}) \rangle, \tag{2.3}$$

where $r = |\mathbf{r}|$ and the angular brackets indicate a spatial average. In particular, here there is a double averaging: one over all the points $\mathbf{x}$ in the Universe, the other over all the points at a distance $\mathbf{r}$ from the fixed point $\mathbf{x}$. The 2PCF measures the degree of correlation of the density field δ at different points of the Universe. If we substitute the expression of $\delta(\mathbf{x})$ in equation (2.1) into equation (2.3) we obtain:

$$\xi(r) = \frac{1}{(2\pi)^6} \int \mathrm{d}\mathbf{k} \int \mathrm{d}\mathbf{k}' \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle e^{i\mathbf{k}\cdot(\mathbf{x}+\mathbf{r})} e^{i\mathbf{k}'\cdot\mathbf{x}}. \tag{2.4}$$

Let us introduce the 3-dimensional Dirac delta, defined as follows:

$$\delta_D^{(3)} \equiv \frac{1}{(2\pi)^3} \int e^{i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}^3 x, \tag{2.5}$$

with the dimensions of a cubic length. We can also provide an operative definition:

$$\int \mathrm{d}^3 z' F(z') \delta_D^{(3)}(z - z') = F(z). \tag{2.6}$$

Given this definition, we can now introduce the power spectrum $P(k)$ as:

$$\langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle \equiv (2\pi)^3 P(k) \delta_D^{(3)}(\mathbf{k} + \mathbf{k}'), \tag{2.7}$$

where $k = |\mathbf{k}|$. Therefore, by substituting it into equation (2.4) we derive:

$$\xi(r) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3 k P(k) e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.8}$$

which implies that the power spectrum, defined in Fourier space, is the Fourier transform of the 2PCF, defined in configuration space. Thus, these two quantities provide the same information, but in two different spaces. Moreover, it is easy to prove that the 2PCF is adimensional, whereas the power spectrum has the dimensions of a cubic length. The density field $\delta(\mathbf{x})$ is real, meaning that $\delta(\mathbf{k})$ is imaginary. As a consequence, we have that $\delta^*(k) = \delta(-k)$, where the superscript $*$ indicated the complex conjugate. From equation (2.8) we can deduce that:

$$P(k) \propto \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle \propto \langle \delta^2(\mathbf{k}) \rangle. \tag{2.9}$$

In order to be rigorous, it is worth mentioning that $P(k)$ is not properly a power spectrum, which should be adimensional. Instead, it is a sort of power density, since the true power spectrum is represented by $P(k)\mathrm{d}^3k$. The definition of the power spectrum can be also written in the following form:

$$\langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle \equiv (2\pi)^3 P(k) \delta_D^{(3)}(\mathbf{k} - \mathbf{k}'). \tag{2.10}$$

In the case of $\mathbf{k} = \mathbf{k}'$, the 3-dimensional Dirac delta becomes:

$$\delta_D^{(3)}(0) = \frac{1}{(2\pi)^3} \int \mathrm{d}^3 x = \frac{V_U}{(2\pi)^3}, \tag{2.11}$$

where $V_U \equiv \int \mathrm{d}^3 x$ is the volume of the Universe. This implies that:

$$\langle \delta(\mathbf{k}) \delta^*(\mathbf{k}') \rangle = \langle |\delta^2(\mathbf{k})| \rangle = (2\pi)^3 P(k) \delta_D^{(3)}(0) = V_U P(k), \tag{2.12}$$

which implies that, as we mentioned, the power spectrum is proportional to the square mean value of the amplitude of fluctuations in Fourier space $\delta(\mathbf{k})$, with the normalization factor represented by the volume of the Universe V_U.

Since the amplitudes of fluctuations are Gaussian distributed, their mean value is equal to zero by definition, while their variance is defined as:

$$\sigma^2 \equiv \langle |\delta^2(\mathbf{x})| \rangle. \tag{2.13}$$

By applying the ergodic hypothesis, we can divide the Universe into independent regions large enough to be fair samples, we can compute the variance σ^2 with a double average: the statistical average over all the volumes of the spatial average of the quadratic value of δ:

$$\sigma^2 = \frac{1}{V_U} \int_{-\infty}^{\infty} \langle \delta(\mathbf{x}) \rangle \mathrm{d}\mathbf{x}. \tag{2.14}$$

The Parseval theorem in 3-dimensions can be written as:

$$\int_{-\infty}^{\infty} f(\mathbf{x}) g^*(\mathbf{x}) \mathrm{d}\mathbf{x} = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \hat{f}(\mathbf{k}) \hat{g}^*(\mathbf{k}) \mathrm{d}\mathbf{k}, \tag{2.15}$$

which implies that the integral in configuration space of the product of the two complex functions $f(\mathbf{x})$ and $g^*(\mathbf{x})$ is proportional to the integral in Fourier space of the product of the Fourier transforms of the two functions $\hat{f}(\mathbf{k})$ and $\hat{g}^*(\mathbf{k})$, with a normalization factor equal to $1/(2\pi)^3$. We can use this result in order to rewrite equation (2.14) in Fourier space as:

$$\sigma^2 = \frac{1}{(2\pi)^3}\frac{1}{V_U}\int_{-\infty}^{\infty}\langle\delta(\mathbf{k})\delta^*(\mathbf{k})\rangle \mathrm{d}\mathbf{k}, \tag{2.16}$$

where we can recognize the power spectrum $P(k)$, so that it becomes:

$$\sigma^2 = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}P(k)\mathrm{d}^3k, \tag{2.17}$$

which is also called punctual variance.

If the density field δ is perfectly Gaussian, meaning that we have:

$$P(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\delta^2}{2\sigma^2}}, \tag{2.18}$$

then Wick's theorem ensures that all the statistical information is encoded in two-point statistics, namely the 2PCF in configuration space and the power spetrum in Fourier space. Indeed, in Fourier space we have:

$$\langle\delta(\mathbf{k}_1)\dots\delta(\mathbf{k}_{2p+1})\rangle = 0, \tag{2.19}$$

$$\langle\delta(\mathbf{k}_1)\dots\delta(\mathbf{k}_{2p})\rangle = \sum_{\text{all pair associations}}\prod_{p\text{ pairs }(i,j)}\langle\delta(\mathbf{k}_i)\delta(\mathbf{k}_j)\rangle. \tag{2.20}$$

This means that all the odd moments are equal to zero, whereas all the even moments can be written in terms of the power spectrum.

Let us now go back to the 2PCF, which is strictly related to the multi-point probability function. Indeed, if we consider a discrete distribution of objects with mean number density $\bar{n}$, the probability of finding an object within the volume element $\mathrm{d}V_1$ is given by $\mathrm{d}P_1 = \bar{n}\mathrm{d}V_1$. Then, the joint probability of finding on object within the volume element $\mathrm{d}V_1$ and another one within the volume element $\mathrm{d}V_2$ at a distance r from the first in the case of independent probabilities is:

$$\mathrm{d}P_{12} = \bar{n}^2\mathrm{d}V_1\mathrm{d}V_2. \tag{2.21}$$

In general, these probabilities are not independent and we can have some correlations. Specifically, the 2PCF is the quantity that measures the excess probability that two particles at volume elements $\mathrm{d}V_1$ and $\mathrm{d}V_2$ are separated by a distance r with respect to a random uniform distribution, that is:

$$\mathrm{d}P_{12} = \bar{n}^2\mathrm{d}V_1\mathrm{d}V_2\left[1+\xi(r)\right]. \tag{2.22}$$

We can have three cases:

- $\xi(r) = 0$: the objects are not correlated and the distribution is completely random and hence uniform;
- $\xi(r) > 0$: the objects are correlated, meaning that the probability of finding the two objects within the two volume elements is larger than the one of a random distribution;
- $\xi(r) < 0$: the objects are anti-correlated, meaning that the probability of finding the two objects within the two volume elements is smaller than the one of a random distribution.

Since the definition of probability implies that $0 \leq \mathrm{d}P_{12} \leq 1$, we deduce that:

$$\xi(r) \geq -1 \tag{2.23}$$

for all r. Operatively, we can obtain an estimator for the 2PCF by counting pairs of objects. Let us assume to have a data catalogueue with N_D objects and a random catalogue with N_R objects uniformly distributed. We can define a grid whose cells have a size r and count, in every cell, the pairs of objects coming from the data catalogue, $DD(r)$, the pairs of objects coming from the random catalogue, $RR(r)$, and the pairs of objects in which one comes from the data catalogue and the other comes from the random catalogue, $DR(r)$. This allows us to write the estimator of the 2PCF introduced by Landy and Szalay, 1993:

$$\hat{\xi}_{LS}(r) = 1 + \frac{N_{RR}}{N_{DD}}\frac{DD(R)}{RR(r)} - 2\frac{N_{RR}}{N_{DR}}\frac{DR(r)}{RR(r)}, \tag{2.24}$$

where N_{DD}, N_{RR} and N_{DR} are the total number of data-data, random-random and data-random pairs, respectively, and their values are:

$$N_{DD} = \frac{N_D(N_D - 1)}{2}, \tag{2.25}$$

$$N_{RR} = \frac{N_R(N_R - 1)}{2}, \tag{2.26}$$

$$N_{DR} = N_D N_R. \tag{2.27}$$

The estimator expressed in equation (2.24) is one of the most used, mainly because its dependence on the uncertainty on the mean density, due to the discretization, is of the second-order, but also because it can be proven to be almost of minimal, hence Poissonian, variance (Labatie et al., 2012).

2.2 Three-point correlation function and bispectrum

In previous section 2.1, we have mentioned that a perfectly Gaussian random field encodes the entirety of its statistical information in two-point statistics. Nevertheless, it is unattainable to have a perfectly Gaussian random field in

real-world scenarios, because of the concurrence of a variety of factors. These include the nonlinearity of the equations regulating the gravitational evolution of cosmic structures, but also effects due to spatial correlations and measurement limitations (see Desjacques and Seljak, 2010). Consequently, n-points statistics provide significant information and become fundamental for clustering analysis (e.g. Szapudi, 2005).

In general, correlation functions, of any order, are defined as the connected part of the joint ensemble average of the density in an arbitrary number of positions. They can be written as:

$$\begin{aligned}\xi_N(\mathbf{x}_1,...,\mathbf{x}_N) =& \langle\delta(\mathbf{x}_1)...\delta(\mathbf{x}_N)\rangle_c \\ & \langle\delta(\mathbf{x}_1)...\delta(\mathbf{x}_N)\rangle - \\ & \sum_{\mathcal{S}\in\mathcal{P}(\{\mathbf{x}_1,...,\mathbf{x}_N\})} \prod_{s_i\in\mathcal{S}} \xi_{\#s_i}(\mathbf{x}_{s_i(1)},...,\mathbf{x}_{s_i(\#s_i)}),\end{aligned} \tag{2.28}$$

where the subscript c stands for connected, the sum is made over the proper partitions (any partition except the set itself) of $\{\mathbf{x}_1,...,\mathbf{x}_N\}$, referred to as $\mathcal{P}(\{\mathbf{x}_1,...,\mathbf{x}_N\})$ and s_i is a subset of $\{\mathbf{x}_1,...,\mathbf{x}_N\}$ contained in the partition $\mathcal{S}$.

Let us now move to Fourier space. Due to the homogeneity of space, $\langle\delta(\mathbf{k}_1)...\delta(\mathbf{k}_N)\rangle_c$ is always proportional to $\delta_D^{(N)}(\mathbf{k}_1+...+\mathbf{k}_N)$. Therefore, we can define $P_N(\mathbf{k}_1,...,\mathbf{k}_N)$ as:

$$\langle\delta(\mathbf{k}_1)...\delta(\mathbf{k}_N)\rangle_c = (2\pi)^N P_N(\mathbf{k}_1,...,\mathbf{k}_N)\delta_D^{(N)}(\mathbf{k}_1+...+\mathbf{k}_N). \tag{2.29}$$

The case of $N=1$ is not interesting, since we have defined the density field $\delta(\mathbf{x})$ to have zero mean. We have already addressed the case $N=2$ in section 2.1 when we discussed about the 2PCF and the power spectrum. Then, the most simple statistics beyond the second-order is the three-point statistics, that is the case of $N=3$. The 3PCF $\zeta(r_{12},r_{13},r_{23})$ is defined as:

$$\zeta(r_{12},r_{13},r_{23}) \equiv \langle\delta(\mathbf{r}_1)\delta(\mathbf{r}_2)\delta(\mathbf{r}_3)\rangle, \tag{2.30}$$

where $r_{12}\equiv|\mathbf{r}_1-\mathbf{r}_2|$, $r_{13}\equiv|\mathbf{r}_1-\mathbf{r}_3|$ and $r_{23}\equiv|\mathbf{r}_2-\mathbf{r}_3|$. Let us consider a discrete distribution of objects with mean number density $\bar{n}$. The joint probability of finding three objects within the independent volume elements $\mathrm{d}V_1$, $\mathrm{d}V_2$ and $\mathrm{d}V_3$, centered at the positions $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$, respectively, is given by:

$$\mathrm{d}P_{123} = \bar{n}^3\mathrm{d}V_1\mathrm{d}V_2\mathrm{d}V_3\left[1+\xi_3(r_{12},r_{13},r_{23})\right], \tag{2.31}$$

where the function $\xi_3(r_{12},r_{13},r_{23})$ contains redundant information, since the probability of finding triplets of objects is artificially increased by the presence of pairs, then we can write:

$$\xi_3(r_{12},r_{13},r_{23}) = \xi(r_{12})+\xi(r_{13})+\xi(r_{23})+\zeta(r_{12},r_{13},r_{23}), \tag{2.32}$$

where $\zeta(r_{12},r_{13},r_{23})$ is the connected 3PCF. Fig. 2.1 shows the $\xi_3(r_{12},r_{13},r_{23})$ function, in which we can discriminate between the different terms we have also

in equation (2.32): the first one is equal to zero, since it is made of singlets only and we have defined the average of $\delta(\mathbf{x})$ to be zero, the second, third and fourth terms are the 2PCF terms $\xi(r_{12})$, $\xi(r_{13})$ and $\xi(r_{23})$, and the last term is the connected 3PCF $\zeta(r_{12}, r_{13}, r_{23})$. If we consider equation (2.29) and apply the

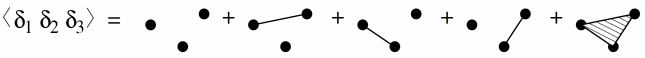

Figure 2.1: Graphic representation of the $\xi_3(r_{12}, r_{13}, r_{23})$ function. The last term represents the connected part, which is the $\zeta(r_{12}, r_{13}, r_{23})$ function. Image of Bernardeau et al., 2002.

case of $N = 3$, we obtain the definition of the bispectrum $B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$:

$$\langle \delta(\mathbf{k}_1)\delta(\mathbf{k}_2)\delta(\mathbf{k}_3)\rangle_c = (2\pi)^3 B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)\delta_D^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \tag{2.33}$$

The bispectrum is the Fourier transform of the 3PCF, and the 3PCF is the Fourier anti-transform of the bispectrum. In chapter 4, we present a new model for both the bispectrum and the 3PCF.

There are several estimators of the 3PCF which consist of counting the triplets of galaxies contained in a certain portion of the total volume, similarly to what we have discussed for pairs of galaxies with the 2PCF (see e.g. Szapudi and Szalay, 1998). Indeed, if we want to measure the 3PCF up to some scale $R_{\max}$, there are $N\left(nV_{R_{\max}}\right)^2$ possible triangles, where N is the total number of galaxies, n is the galaxy number density and $V_{R_{\max}} = 4/3\pi R_{\max}^3$ is the considered volume. As a consequence, such algorithms scale as the number of possible triangles, i.e. $N\left(nV_{R_{\max}}\right)^2$. In the past years, alternative techniques have emerged, such as those utilizing multiple kd-trees, which prove to be faster than simply enumerating all triangles (see e.g. Gardner et al., 2007; Gray et al., 2004). However, even these methods fall short in terms of speed when dealing with extensive surveys like the Baryon Oscillation Spectroscopic Survey (BOSS), which has an average separation of $13\,h^{-1}$Mpc. Recently, Slepian and Eisenstein, 2015b proposed a new algorithm for the measurement of the 3PCF based on a spherical harmonic decomposition (SHD method). Explicitly, the 3PCF is parametrized through two triangle sides r_{12} and r_{13} and the angle θ between them. We can expand it onto Legendre polynomials:

$$\zeta(r_{12}, r_{13}, \theta) = \sum \zeta_\ell(r_{12}, r_{13}) P_\ell(\theta), \tag{2.34}$$

where $\zeta_\ell(r_{12}, r_{13})$ are the coefficients of the expansion, or multipole moments, and $P_\ell(\theta)$ are the Legendre polynomials. Such decomposition has the great advantage that the multipole moments of the 3PCF can be obtained much faster than other 3PCF decompositions without sacrificing accuracy, since this method is exact in angles. Specifically, using the spherical harmonic addition theorem we can decompose the Legendre polyomials into factors depending on a single angle each. As a result, this algorithm scales as $N\left(nV_{R_{\max}}\right)$, that is a factor $\left(nV_{R_{\max}}\right)$ faster than the direct triplet counting approach.

2.3 Clustering distortions

In the previous section we have analyzed one of the main statistical tools used to constrain cosmology, that is the 2PCF and the 3PCF, and their Fourier counterparts. We can now direct our study to analyze the strategies of exploitation of such statistics, some of which are used in chapter 6 to estimate some cosmological parameters.

Let us assume to have a redshift survey, which maps the spatial distribution of of galaxies, or other cosmic tracers. Let us consider a galaxy at a comoving distance $\chi(z)$ (1.34) from us. Its observed position can be written as:

$$\mathbf{x}_{\rm obs}(z,\theta,\phi) = \chi(z)\hat{n}(\theta,\phi), \tag{2.35}$$

where $\hat{n}(\theta,\phi) = \mathbf{x}_{\rm obs}/|\mathbf{x}_{\rm obs}|$ is the unit vector which identifies the direction of the galaxy in the sky, as Fig. 2.2 shows. In cosmology, we usually express the

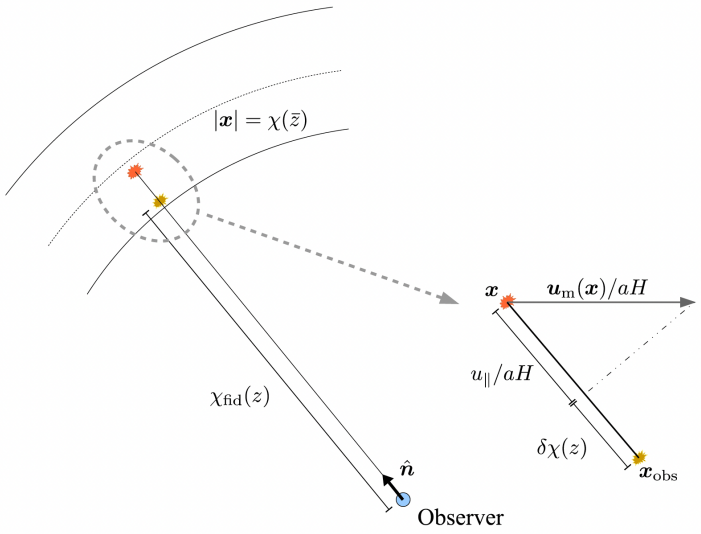

Figure 2.2: Graphic representation of the process of observation of a galaxy, whose direction in the sky is identified by the unit vector $\hat{n}(\theta,\phi)$. The galaxy is selected within the redshift range identified by the solid arcs centered at redshift $\bar{z}$ identified by the dashed arc. The observed position of the galaxy $\mathbf{x}_{\rm obs}$ is displayed in yellow, whereas its true position $\mathbf{x}$ is shown in orange. On the right of the diagram there is a zoom in to show the relation between these two positions. Image of Dodelson, 2003.

distance of an object in terms of redshift. However, in the context of clustering, this introduces two problems. The first one is due to the fact that in order to relate distances to redshifts, we need to assume a cosmology, which is in general

different from the true one, so that:

$$\chi_{\rm fid}(z) = \chi(z) + \delta\chi(z), \tag{2.36}$$

where $\chi_{\rm fid}(z)$ is the fiducial distance-redshift relation, $\chi(z)$ is the true one and $\delta\chi(z)$ is the difference between the two. Assuming a cosmology that deviates from the true one leads to the emergence of what we term as geometrical distortions in the clustering of cosmic tracers. These distortions can be used to get cosmological constraints by exploiting the peak of the baryonic acoustic oscillations (BAO). The second problem consists on the fact that $\mathbf{x}_{\rm obs}$ does not correspond to the true position of the galaxy if it is not at rest with respect to the background Universe, but it has a peculiar velocity $\mathbf{u}$. In the latter case, the observed redshift of the galaxy would be:

$$1 + z = \frac{1}{a_{\rm em}} \left[1 + u_\parallel\right], \tag{2.37}$$

where $a_{\rm em}$ is the scale factor at the moment of the emission of the radiation from the galaxy and $u_\parallel \equiv \mathbf{u} \cdot \hat{n}$ is the component of the peculiar velocity of the galaxy parallel to the unit vector $\hat{n}$. The factor $1/a_{\rm em}$ is the cosmological redshift and it takes into account the expansion of the Universe, whereas the $u_\parallel/a_{\rm em}$ term represents the linear-order Doppler shift due to the fact that the galaxy has a peculiar velocity with respect to the observer. We are allowed to use the linear-order Doppler shift since the peculiar velocity $\mathbf{u}$ is much smaller than the speed of light c. We can compute the error that we make in equation (2.35) in this case:

$$\Delta x = \left.\frac{\partial x_{\rm obs}}{\partial u_\parallel}\right|_{u_\parallel = 0} u_\parallel = \frac{1}{aH} u_\parallel \hat{n} \tag{2.38}$$

Since only the component of the velocity parallel to line of sight $u_\parallel$ enters in the equation, the direction which identifies the galaxy, that is the unit vector $\hat{n}$, remains unchanged. Moreover, as we can deduce from equation (2.38), if the peculiar velocity is positive, it means that the galaxy is receding from us, resulting in an increase of the observed distance, hence a positive $\Delta\mathbf{x}$. Peculiar velocities of galaxies give rise to another class of distortions known as redshift space distortions (RSD), sometimes referred to as dynamical distortions. In numerical simulations, we can deal with galaxies with vanishing peculiar velocities, resulting in the absence of RSD. In such cases, we commonly refer to the space as real space. Conversely, when galaxies have non-vanishing peculiar velocities, hence in all real scenarios, we have RSD, and we denote the space as redshift space. Combining the effects of equations (2.36) and (2.38), we obtain:

$$\mathbf{x}_{\rm obs} = \mathbf{x} + \left[\delta\chi(z) + \frac{1}{aH} u_\parallel(\mathbf{x})\right] \hat{n}, \tag{2.39}$$

where $\mathbf{x}$ is the true position of the galaxy, that is the position of the galaxy in absence of geometrical and RSD.

The first studies regarding the relation between the redshift-space and the real-space power spectra were carried out by Kaiser, 1987, whose starting point was the fact that the number of galaxies in a specific region does not change if we use the observed position $\mathbf{x}_{\rm obs}$ or the true one $\mathbf{x}$. If we consider the galaxy number density field n_g, we can write:

$$n_{g,\rm obs}(\mathbf{x}_{\rm obs})\mathrm{d}^3x_{\rm obs} = n_g(\mathbf{x})\mathrm{d}^3x, \tag{2.40}$$

where $n_{g,\rm obs}(\mathbf{x}_{\rm obs})$ is the number density of galaxies in redshift space, whereas $n_g(\mathbf{x})$ is the number density of galaxies in real space, $x_{\rm obs} = |\mathbf{x}_{\rm obs}|$ and $x = |\mathbf{x}|$. The volume elements $\mathrm{d}^3x_{\rm obs}$ and d^3x can be written as:

$$\mathrm{d}^3x_{\rm obs} = x_{\rm obs}^2\mathrm{d}x_{\rm obs}\mathrm{d}\Omega, \tag{2.41}$$

$$\mathrm{d}^3x = x^2\mathrm{d}x\mathrm{d}\Omega, \tag{2.42}$$

where $\mathrm{d}\Omega$ is the angular volume element, which does not change whether we are in redshift space or in real space. For this reason, we can write:

$$n_{g,\rm obs}(\mathbf{x}_{\rm obs}) = n_g(\mathbf{x})J, \tag{2.43}$$

where J is the Jacobian of the change of coordinates:

$$J \equiv \left|\frac{\mathrm{d}^3x}{\mathrm{d}^3x_{\rm obs}}\right| = \left|\frac{\mathrm{d}x}{\mathrm{d}x_{\rm obs}}\right| \frac{x^2}{x_{\rm obs}^2}. \tag{2.44}$$

By considering equation (2.39), the Jacobian becomes:

$$J = \left(1 + \frac{\delta\chi}{x} + \frac{u_\parallel}{aHx}\right)^{-2} \left|1 + \frac{\mathrm{d}}{\mathrm{d}x}\delta\chi + \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel\right|^{-1}. \tag{2.45}$$

The derivative of $\delta\chi$ with respect to x can be computed as:

$$\frac{\mathrm{d}}{\mathrm{d}x}\delta\chi = \frac{\mathrm{d}z}{\mathrm{d}x}\frac{\mathrm{d}\delta\chi}{\mathrm{d}z} = H\delta\left(H^{-1}\right) = -H^{-1}\delta H, \tag{2.46}$$

where we have used $\mathrm{d}z/\mathrm{d}x = \mathrm{d}z/\mathrm{d}\chi = H$ and

$$\delta H(z) = H(z) - H_{\rm fid}(z) \tag{2.47}$$

is the difference between the true Hubble parameter $H(z)$ and the fiducial one $H_{\rm fid}(z)$. Therefore we obtain:

$$\begin{aligned} J &= \left(1 + \frac{\delta\chi}{x} + \frac{u_\parallel}{aHx}\right)^{-2} \left|1 - H^{-1}\delta H + \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel\right|^{-1} \\ &\simeq \left(1 - 2\frac{\delta\chi}{x} + H^{-1}\delta H - 2\frac{u_\parallel}{aHx}\right)\left(1 - \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel\right), \end{aligned} \tag{2.48}$$

where in the second line we have removed all second-order terms. Let us now consider the terms in the first parentheses. The ones with $\delta\chi$ and δH depend on

x only through the redshift z. Assuming that we are observing sources within a narrow redshift range centered at $\bar{z}$, we can set x to $\bar{\chi} = \chi(\bar{z})$, which is the distance to the mean redshift $\bar{z}$. Moreover, we can evaluate both $\delta\chi$, δH and H at $\bar{z}$. By doing so, the first three terms of the first parentheses can be considered as constant. Let us focus on the fourth term, $2u_\parallel/aHx$. The ratio $u_\parallel/aH$ represents the apparent displacement of galaxies due to the component of their peculiar velocity parallel to the line of sight. Using linear theory, it can be found that this displacement is typically $\lesssim 10h^{-1}$ Mpc. Conversely, $x \sim \bar{\chi}$ is typically of the order of hundreds of Mpc in current redshift surveys, implying that this term can be neglected. Consequently, the first parentheses can be considered as a constant prefactor, which we denote with $\bar{J}$:

$$J \simeq \bar{J}\left(1 - \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel\right), \tag{2.49}$$

where:

$$\bar{J} = 1 - 2\frac{\delta\chi(\bar{z})}{\chi(\bar{z})} + H^{-1}(\bar{z})\delta H(\bar{z}). \tag{2.50}$$

If we refer to the average number density of galaxies with $\bar{n}_g$, we can write the number density of galaxies in true and observed coordinates as, respectively:

$$n_g = \bar{n}_g\left(1 + \delta_g\right), \tag{2.51}$$

$$n_{g,\mathrm{obs}} = \bar{n}_g\left(1 + \delta_{g,\mathrm{obs}}\right). \tag{2.52}$$

Considering equation (2.43), substituting the Jacobian in equation (2.49) and expanding to first-order in perturbations we obtain:

$$1 + \delta_{g,\mathrm{obs}}(\mathbf{x}_{\mathrm{obs}}) = \bar{J}\left[1 + \delta_g\left(\mathbf{x}[\mathbf{x}_{\mathrm{obs}}]\right) - \frac{1}{aH}\frac{\partial}{\partial x}u_\parallel\left(\mathbf{x}[\mathbf{x}_{\mathrm{obs}}]\right)\right], \tag{2.53}$$

where both the galaxy density δ_g and the galaxy peculiar velocity $u_\parallel$ in the right-hand side are evaluated at the true position $\mathbf{x}$. In the next two sections we will analyze separately the effects of RSD and geometrical distortions.

2.3.1 Redshift-space distortions

Fig. 2.3 graphically shows the distortions introduced by neglecting the peculiar velocity of cosmic tracers when computing comoving coordinates from redshifts, both in the linear, hence at large-scales, and in the nonlinear, hence at small-scales, regimes.

Let us first consider the linear RSD, meaning the RSD at large scales, that is of the order of tens to hundreds of Mpc. The overdense region, represented as an orange circle, is situated at the center. In a scenario where galaxies are at rest with respect to the observer, hence their peculiar velocities are zero, then the shape of a contour of constant density would be circular and we would be in real space. On the other hand, if galaxies have non-zero peculiar velocities,

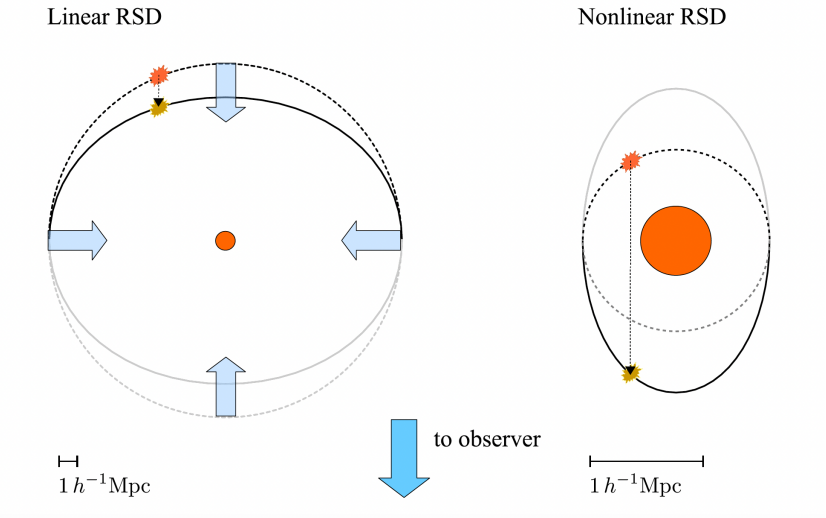

Figure 2.3: Graphic representation of linear (left panel) and nonlinear (right panel) RSD. In both cases, there is a central overdensity displayed with the orange circle. The observer is assumed to be far below the figure, so that the line of sight $\hat{n}$ can be considered vertical. In both cases, the circular dashed line indicates a contour of constant density in real space, while the solid line denotes its distortion in redshift space. The observed position of the galaxy is shown in yellow, while the true one is represented in orange. Image of Dodelson, 2003.

then they tend to fall towards the central region, due to its gravitational attraction. Consequently, galaxies beyond the overdensity are moving towards us and will appear closer to us than they actually are, whereas galaxies preceding the overdensity are moving away from us and will appear farther to us than they actually are. The resulting effect is an apparent anisotropy in the contours of constant density, which in redshift space are flattened in the direction perpendicular to the line of sight. Galaxies moving towards each other result in an increase of their number density, consequently we have a stronger clustering in redshift space than in real space.

Let us now examine the nonlinear RSD, hence the RSD at small scales, of the order of $1\,h^{-1}$Mpc. At these scales, the velocities are a bit larger and, more significantly, the displacement of galaxies due to the component of their peculiar velocity parallel to the line of sight, the $u_{\parallel}/aH$ term, becomes much larger than the distance x separating the two galaxies. As a result, the contours of constant density appear to be elongated along the line of sight. Furthermore, as the right panel in Fig. 2.3 displays, a galaxy located beyond the overdensity is mapped onto a point which precedes the overdensity.

The effects introduced by the RSD can be efficiently exploited to constrain cosmology. Specifically, we can measure the power spectrum and the bispec-

trum, or, analogously, the 2PCF and the 3PCF, in redshift space and then perform a fit in order to constrain the values of the linear growth rate of perturbations. This strategy will be thoroughly described in chapter 3.

2.3.2 Alcock-Paczynski effect

Let us now turn to the distortions introduced when a cosmology different from the true one is assumed, resulting in the assignment of comoving distances using a distance-redshift relation different from the true one. These geometrical distortions, also called Alcock-Paczynski effect (Alcock and Paczynski, 1979), induce the apparent displacement of tracers from their true position by a quantity $\delta\chi(z)$, as Fig. 2.4 shows. The displacement $\delta\chi(z)$ depends on the redshift

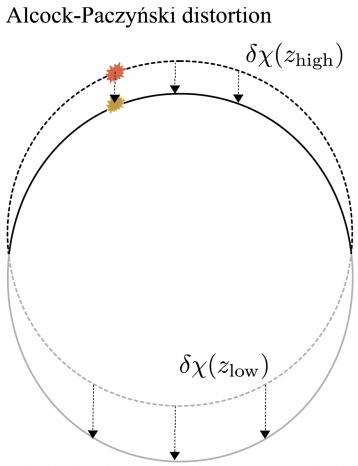

Figure 2.4: Graphic representation of the Alcock-Paczynski effect. The true positions of galaxies (in orange) are shifted by a quantity $\delta\chi(z)$ and mapped onto their observed positions (in yellow). In absence of this effect the contours of constant density would be circular (dashed line), but they become elliptical (solid line) when it is present. Image of Dodelson, 2003.

z, specifically the higher is the redshift, the lower is the displacement, implying that sources closer to us are subject to a larger displacement compared to farther ones. The resulting effect is that the contours of constant density, which would be circular in absence of these distortions, appear to be elliptical.

Practically, we assume a fiducial distance-redshift relation $\chi_{\rm fid}$ and we assign the 3-dimensional observed position $\mathbf{x}_{\rm obs}$ to a galaxy by using the redshift z and the two angles (θ, ϕ) which identify the galaxy in the sky. We assume the flat-sky approximation, which allows us to consider the position of the galaxy in the sky to be a 2-dimensional vector θ. We choose the origin such that:

$$\mathbf{x}_{\rm obs} = 0 \;\Leftrightarrow\; \theta = 0, z = \bar{z}, \tag{2.54}$$

which means that the origin of the reference frame corresponds to an angular position in the sky of $\theta = 0$, that is a point at the center of the survey, and a redshift $z = \bar{z}$, which is the central value of the considered redshift range. The transverse components, namely the components perpendicular to the line of sight, that we assign to the galaxy are:

$$(x^1_{\rm obs}, x^2_{\rm obs}) = \chi_{\rm fid}(z)(\theta^1, \theta^2). \tag{2.55}$$

The true transverse components that we should assign the the galaxy are instead:

$$(x^1, x^2) = \chi(z)(\theta^1, \theta^2) = \left[1 - \frac{\delta\chi(z)}{\chi_{\rm fid}(z)}\right](x^1_{\rm obs}, x^2_{\rm obs}), \tag{2.56}$$

which implies that if $\delta\chi(z) > 0$, then we assign galaxies a comoving distance larger than the true one, resulting to the fact that the assigned transverse components $(x^1_{\rm obs}, x^2_{\rm obs})$ are farther form the origin compared to the true transverse components (x^1, x^2). Conversely, if $\delta\chi(z) < 0$, then the comoving distance assigned to the galaxies is smaller than the true one, meaning that the assigned transverse components $(x^1_{\rm obs}, x^2_{\rm obs})$ are closer to the origin with respect to the true ones (x^1, x^2).

While the transverse components are identified by the angular position of the galaxy in the sky, the line of sight component $x^3_{\rm obs}$ is determined by the redshift z. In our reference frame $z = \bar{z}$ corresponds to $x^3 = 0$, then:

$$x^3_{\rm obs}(z) = \chi_{\rm fid}(z) - \chi_{\rm fid}(\bar{z}). \tag{2.57}$$

If we assume to have a narrow redshift range, so that $(z - \bar{z}) \ll 1$, we can expand around $(z - \bar{z})$ and, up to linear order, we obtain:

$$x^3_{\rm obs}(z) \simeq \frac{1}{H_{\rm fid}(\bar{z})}(z - \bar{z}), \tag{2.58}$$

where we have used that $\mathrm{d}\chi/\mathrm{d}z = 1/H$. The true line of sight component is:

$$x^3(z) \simeq \frac{1}{H(\bar{z})}(z - \bar{z}) = \frac{H_{\rm fid}(\bar{z})}{H(\bar{z})} x^3_{\rm obs}(z). \tag{2.59}$$

Knowing that $\delta H(z) = H(z) - H_{\rm fid}(z)$ and keeping only linear order terms in δH, we can write:

$$x^3(z) = \left[1 - \frac{\delta H(\bar{z})}{H_{\rm fid}(\bar{z})}\right] x^3_{\rm obs}(z). \tag{2.60}$$

Therefore, considering all the three components, we have:

$$\mathbf{x}(\mathbf{x}_{\rm obs}) = \left([1 - \alpha_\perp]x^1_{\rm obs}, [1 - \alpha_\perp]x^2_{\rm obs}, [1 - \alpha_\parallel]x^3_{\rm obs}\right), \tag{2.61}$$

where:

$$\alpha_\perp = \left.\frac{\delta\chi}{\chi_{\rm fid}}\right|_{\bar{z}}, \tag{2.62}$$

$$\alpha_{\parallel} = \left.\frac{\delta H}{H_{\rm fid}}\right|_{\bar{z}} . \tag{2.63}$$

By measuring the power spectrum or the bispectrum, we can get constraints on the parameters $\alpha_{\perp}$ and $\alpha_{\parallel}$, which are extremely important in cosmology. Indeed, from equation (2.61) we can derive:

$$\chi(\bar{z}) = \chi_{\rm fid}(\bar{z})\,[1+\alpha_{\perp}]\,, \tag{2.64}$$

$$H(\bar{z}) = H_{\rm fid}(\bar{z})\,\left[1+\alpha_{\parallel}\right], \tag{2.65}$$

implying that constraints on $\alpha_{\perp}$ and $\alpha_{\parallel}$ impose constraints on the comoving distances χ of cosmic tracers and on the Hubble rate H at a specific redshift $\bar{z}$.

2.3.3 Baryonic acoustic oscillations

The Alcock-Paczynski effect can be effectively quantified using the peak of the BAO, which represents a standard ruler in cosmology (refer to Bassett and Hlozek, 2009 for an in-depth review). These oscillations are fluctuations in the density of baryonic matter, arising from acoustic waves in the early Universe. Indeed, during the hot and dense phase of the Universe, photons and baryons formed a tightly interlinked medium known as the photon-baryon fluid, primarily interacting through Thomson scattering. The complex interplay between photons and baryons resulted in a competition between the radiation pressure, which tended to avoid the collapse, and gravity, which tended to favour the collapse. Consequently, spherical acoustic waves were generated, propagating at the speed of sound. Then, the Universe underwent cooling and H recombination occurred: photons were liberated to free stream reaching us today, producing the CMB. At this point, the radiation pressure offered by photons was removed, causing the acoustic waves to freeze-out. These waves left an imprint in the CMB, which can be observed bystudying its anisotropies, but also in the clustering properties of the large-scale structure (LSS) of the Universe at a specific scale. This scale, corresponding to the sound horizon at the epoch of H recombination, is, as Planck Collaboration et al., 2020 measured:

$$r_* = (144.57 \pm 0.22)\,\text{Mpc}. \tag{2.66}$$

Indeed, as it is shown in Fig. 2.5, we can observe the BAO feature in the 2PCF at a scale around $100\,h^{-1}$Mpc. Furthermore, the BAO feature can be detected also in the power spectrum, where, instead of the peak we observe in the 2PCF, it appears as a series of oscillations, as Fig. 2.6 displays. The first peak of these oscillations is located at a wave number $k_* = 2\pi/r_*$.

The scale r_s is known with high accuracy from the CMB, therefore measuring this feature with clustering analysis at redshift $\bar{z}$ provides an efficient method to measure $\alpha_{\perp}$ and $\alpha_{\parallel}$. Indeed, the BAO peak is a feature of well-known size imprinted in the distribution pattern of galaxies and other cosmic tracers, thus measuring its observed size at a given redshift $\bar{z}$ allows us to measure the distance

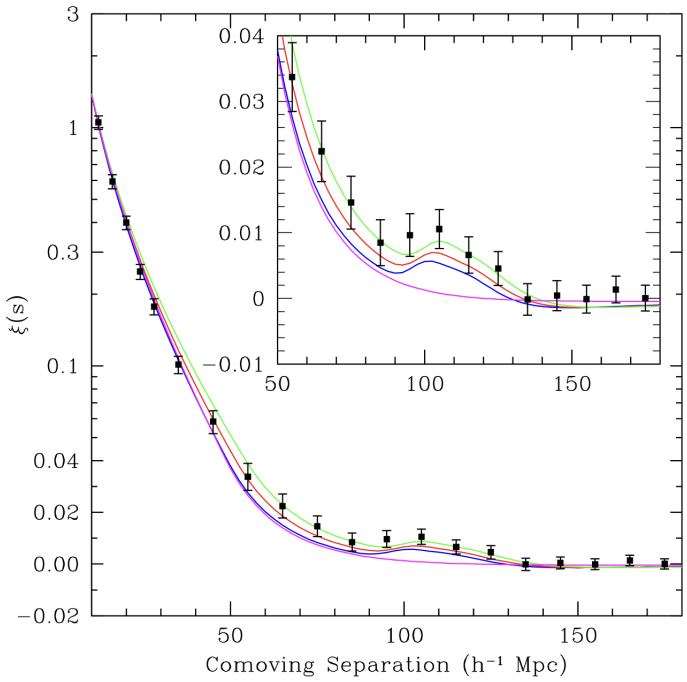

Figure 2.5: BAO peak in the 2PCF $\xi(s)$ of a sample of Luminous Red Galaxies (LRGs) from the Sloan Digital Sky Survey. The green, red and blue lines represent models with $\Omega_m h^2 = 0.12, 0.13, 0.14$, respectively. The purple line represents a pure CDM model with $\Omega_b = 0$. Image from Eisenstein et al., 2005.

to that redshift. Specifically, through $\alpha_\perp$ we can obtain a direct measurement of the angular diameter distance $d_A(\bar{z})$, defined as:

$$d_A(\bar{z}) = \frac{r(\bar{z})}{1+\bar{z}}, \tag{2.67}$$

which in an Euclidean Universe is equal to the comoving distance $\chi(\bar{z})$.

To summarize, the BAO feature is a well determined scale imprinted in the clustering properties of the LSS of the Universe and it can be exploited to measure the angular diameter distance $d_A(\bar{z})$, via $\alpha_\perp$, and the Hubble rate $H(\bar{z})$, via $\alpha_\parallel$.

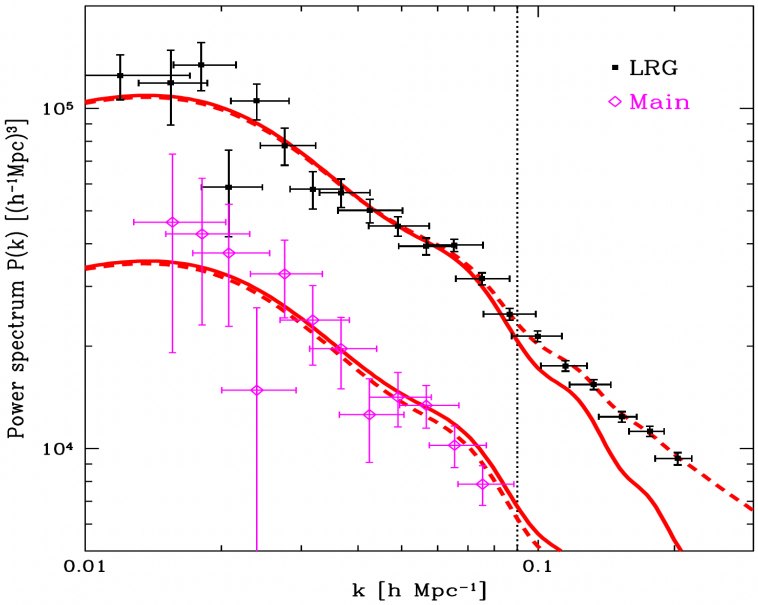

Figure 2.6: BAO feature in the power spectrum of the main sample (bottom curve) and the LRG sample (top curve) from the Sloan Digital Sky Survey. The solid lines show the ΛCDM fits and the dashed lines include also nonlinear corrections. Image from Tegmark et al., 2006.

Chapter 3

Modelling the two-point correlation function: the eTNS model

In this chapter, we describe the theoretical model used for the analysis of the 2PCF, for the cosmological parameter estimation presented in chapter 6. First, we discuss the galaxy bias model, which represents the relation between the dark matter and the galaxy density fields. Then, we introduce the model we adopt for the anisotropic power spectrum, that is the one proposed by Beutler et al., 2014, which we call extended-TNS (eTNS) model since it is an extension of the model proposed by Taruya et al., 2010 (TNS). We finally present the 2PCF model, obtained through the Fourier anti-transformation of the power spectrum model.

3.1 Galaxy bias model

As we have discussed in chapter 2, the matter power spectrum and bispectrum encode extremely rich information for cosmology. However, we do not have a direct way of measuring them, since the majority of matter is non-baryonic, and much of the baryonic matter is not easily visible (e.g. it is in the form of diluted hot gas). Instead, we can directly observe galaxies, or any other astrophysical objects, as tracers of the large-scale matter distribution. The galaxy distribution only indirectly reflects the underlying matter distribution. Thus, the galaxy density field is not the same as the matter density field. Therefore, we need a way to relate them. In principle, this relation should be predicted by a given cosmological model. However, the formation and evolution of galaxies, despite the significant progress made in recent years (e.g. Cen and Ostriker, 2000; Katz et al., 1999), is still poorly understood, since the models rely on simplified descriptions of the main physical processes, such as star formation, supernova

feedback, and active galactic nuclei (AGN) feedback, all of which continue to elude a complete understanding. Consequently, the parameters of the relation between the matter density field and that of galaxies, called bias parameters, must be determined from the data themselves. In general, we can write the biasing relation as:

$$\delta_g(x) = \mathcal{B}\delta(x), \tag{3.1}$$

where $\delta_g(x)$ is the galaxy density field, $\delta(x)$ is the matter density field, and $\mathcal{B}[\delta(x)]$ is the galaxy bias. Since there is not a universally accepted model for the galaxy bias yet, we can choose the parametrization which best suites our needs. Indeed, in literature many different approaches have been proposed.

A widely used model for the galaxy bias $\mathcal{B}$ consists of a Taylor expansion in δ of δ_g (Fry and Gaztañaga, 1993):

$$\delta_g(x) = \sum_{m=0} \frac{b_m}{m!}\delta^m(x). \tag{3.2}$$

The simplest possible galaxy bias model can be obtained by truncating this Taylor expansion at first-order, resulting in a linear relation between the galaxy density field and the matter density field:

$$\delta_g(x) = b_1\delta(x), \tag{3.3}$$

where b_1 is called the linear bias parameter. This model does not account for nonlinearities in the evolution of structures. For this reason, it is more accurate to include also the nonlinear bias parameter in our description. Then, truncating the Taylor expansion in equation (3.2) at second-order, we obtain:

$$\delta_g(x) = b_1\delta(x) + \frac{b_2}{2}\delta^2(x), \tag{3.4}$$

where b_2 is called the nonlinear bias parameter. Furthermore, recent works have shown that we must include in the model also non-local effects, due to the fact that the evolution of the matter density field is influenced by large-scale tidal fields, even in the case of local initial conditions (Catelan et al., 1998). Consequently, neglecting these contributions can lead to errors in the estimations of the other bias parameters, even when considering linear regimes. As a result, we introduce the tidal bias parameter b_t, which accounts for these effects, into our galaxy bias model[1]:

$$\delta_g(x) = b_1\delta(x) + \frac{b_2}{2}\delta^2(x) + b_t s^2(x), \tag{3.5}$$

where $s^2(x)$ is the square of the tidal field, defined as $s^2(x) = s_{ij}(x)s_{ij}(x)$, with $s_{ij}(x) = \partial_i\partial_j\Phi(x) - \delta^{Kr}_{ij}\delta(x)$, where $\Phi(x)$ is the gravitational potential, related to the density field through $\nabla^2\Phi(x) = \delta(x)$, and δ^{Kr}_{ij} is the Kronecker delta.

[1]In the literature, also other notations can be found, like the one of McDonald and Roy, 2009, who has an additional factor 1/2 in the definition of the tidal bias.

3.2 Power spectrum model

The TNS model is defined as follows:

$$P_g(k,\mu) = e^{-(fk\mu\sigma_v)^2}\Big[P_{g,\delta\delta}(k) + 2f\mu^2 P_{g,\delta\theta}(k) + f^2\mu^4 P_{g,\theta\theta}(k) + A(k,\mu) + B(k,\mu)\Big], \tag{3.6}$$

where μ is the cosine of the angle between the wavenumber vector k and the line of sight, f is the growth rate defined in equation (1.101), σ_v is the 1-dimensional velocity dispersion, whose estimate at linear order can be obtained from:

$$\sigma_v^2 = \frac{1}{3}\int \frac{\mathrm{d}\mathbf{q}}{(2\pi)^3}\frac{P_{\rm lin}(q,z)}{q^2}, \tag{3.7}$$

where $P_{\rm lin}(q,z)$ is the linear power spectrum calculated at q, where q is the modulus of the wave vector $|\mathbf{q}|$, and redshift z. The exponential factor in equation (3.6) represents the damping function accounting for the Finger-of-God effect, while the first three terms in the square brackets describe an extension of the Kaiser factor (see Kaiser, 1987) and the A and B terms are corrections deriving from the higher-order correlation between Kaiser terms and velocity fields in mapping to redshift space (we refer to Taruya et al., 2010 for their definitions). The velocity divergence θ is defined as:

$$\theta \equiv -\frac{\nabla u}{aHf}. \tag{3.8}$$

The spectra $P_{g,\delta\delta}$ and $P_{g,\delta\theta}$ denote, respectively, the auto power spectrum of density and velocity divergence, while $P_{g,\theta\theta}$ refers to their cross power spectrum. Let us define the velocity bias as the parameter describing the tendency for galaxies to have a different velocity distribution compared to that of the overall matter distribution. That is, it is the relation between the velocity field of galaxies, θ_g, and the velocity field of matter, θ. Beutler et al., 2014 assumes no velocity bias, hence $\theta_g = \theta$, and its extension to the TNS model consists in the fact that it includes five galaxy bias parameters: the linear bias b_1, the nonlinear bias b_2, the tidal bias b_t, the 3rd-order non-local bias $b_{\rm 3nl}$, which is important to explain the large-scale power spectrum (Saito et al., 2014), and the constant stochasticity term N, which accounts for any random or poorly understood process that affect the distribution of galaxies, but is not directly tied to the underlying matter density field. Indeed, the spectra $P_{g,\delta\delta}$ and $P_{g,\delta\theta}$ are written as:

$$P_{g,\delta\delta}(k) = b_1^2 P_{\delta\delta}(k) + 2b_2 b_1 P_{b2,\delta}(k) + 2b_t b_1 P_{bs2,\delta}(k) + 2b_{\rm 3nl} b_1 \sigma_3^2(k) P_{\rm lin}(k) + b_2^2 P_{b22}(k) + 2b_2 b_t P_{b2s2}(k) + b_t^2 P_{bs22}(k) + N, \tag{3.9}$$

$$P_{g,\delta\theta}(k) = b_1 P_{\delta\theta}(k) + b_2 P_{b2,\theta}(k) + b_t P_{bs2,\theta}(k) + b_{\rm 3nl}\sigma_3^2(k) P_{\rm lin}(k), \tag{3.10}$$

where $P_{\rm lin}$ is the linear matter power spectrum. The power spectra $P_{g,\delta\delta}$, $P_{g,\delta\theta}$ and $P_{g,\theta,\theta}$ are evaluated with the publicly available `RegPT` code[2] (Taruya et al., 2012) at 2-loop order. The expressions for the other power spectra in equations (3.9) and (3.10) can be found in equations (42)-(48) of Beutler et al., 2014, as well as the definition of σ_3^2, which is shown in equation (53) of the same paper.

Now that we have obtained the model for the power spectrum in Fourier space, we have to move to configuration space and derive the model for the 2PCF. This is a straightforward process, since it only consists of computing the inverse-Fourier transform of the eTNS model, expressed in equations (3.6), (3.9) and (3.10). To do so, we exploit FFTLog[3] (Hamilton, 2000), a code to take the fast Fourier or Hankel transform of a discrete periodic sequence of logarithmically spaced points, to take the inverse Fourier transform of the power spectrum model and obtain the 2PCF model.

[2]The code is available at https://www2.yukawa.kyoto-u.ac.jp/~atsushi.taruya/regpt_code.html.

[3]The code is available at http://jila.colorado.edu/~ajsh/FFTLog/.

Chapter 4

Modelling the three-point correlation function: a new approach

In this Thesis work, we propose an extension of the bispectrum model proposed by Scoccimarro et al., 1999 (hereafter SCF99), including an additional bias parameter, that is the tidal bias. Furthermore, we propose a revision of the model proposed by Slepian and Eisenstein, 2017 (hereafter SE17), which includes some calculation errors. For this reason, starting from our model for the bispectrum, we theoretically derived our model for the 3PCF. In this chapter, we present our results regarding the models for both the bispectrum and the 3PCF.

4.1 Bispectrum model

We extend the bispectrum model proposed by SCF99 by including in our analysis also the contribution from the tidal bias parameter b_t. Specifically, our model contains an additional term, with respect to SCF99, in the definition of the second-order perturbation theory redshift-space kernel Z_2. This term, $b_t S_2(\mathbf{k}_1, \mathbf{k}_2)$, accounts for the effect of the tidal field, whose kernel is S_2, which affects the evolution of the matter density field (see section 3.1). In this section, we present the theoretical derivation of our bispectrum model.

4.1.1 Perturbation theory

We define the first- and second-order perturbation theory (PT) redshift-space kernels Z_n for the galaxy density field as:

$$Z_1(\mathbf{k}_i) = (b_1 + f\mu_i^2), \tag{4.1}$$

$$Z_2(\mathbf{k}_1,\mathbf{k}_2) = b_1 F_2(\mathbf{k}_1,\mathbf{k}_2) + f\mu^2 G_2(\mathbf{k}_1,\mathbf{k}_2) + \frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}(b_1 + f\mu_2^2) + \frac{\mu_2}{k_2}(b_1 + f\mu_1^2)\right] + \frac{b_2}{2} + b_t S_2(\mathbf{k}_1,\mathbf{k}_2), \tag{4.2}$$

where $\mu \equiv \mathbf{k}\cdot\hat{z}/k$, with $\mathbf{k} \equiv \mathbf{k}_1 + ... + \mathbf{k}_n$, and $\mu_i \equiv \mathbf{k}_i \cdot \hat{z}/k_i$. F_2 and G_2 are the second-order kernels for the real-space density and velocity-divergence fields and they are defined as:

$$F_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{5}{7} + \frac{x}{2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2x^2}{7}, \tag{4.3}$$

$$G_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{3}{7} + \frac{x}{2}\left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{4x^2}{7}, \tag{4.4}$$

where $x \equiv \frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1 k_2} \equiv \cos\theta$. Finally, the second-order kernel S_2 for the tidal tensor s_{ij} is defined as[1]:

$$S_2(\mathbf{k}_1,\mathbf{k}_2) = \frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1 k_2} - 1 = x^2 - 1. \tag{4.5}$$

4.1.2 Derivation of the model

We define the tree-level redshift-space galaxy bispectrum as follows:

$$B_s(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3) = 2Z_2(\mathbf{k}_1,\mathbf{k}_2)Z_1(\mathbf{k}_1)Z_1(\mathbf{k}_2)P(k_1)P(k_2) + \text{cyc.}, \tag{4.6}$$

where "cyc." refers to a sum over cyclic permutations of $\{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3\}$, since we have to account for the indistinguishability of the vertices of the triangles. Indeed, when computing the bispectrum, we deal with the expectation value of $\delta_g(\mathbf{k}_1)\delta_g(\mathbf{k}_2)\delta_g(\mathbf{k}_3)$, where δ_g is defined in (3.5). If we only keep terms up to the fourth (leading) order in δ, we will have two linear terms and one second-order term. For simplicity, we assume that the second-order density field is always contributed by the galaxy at $\mathbf{k}_1$, which is also the origin of the reference frame. By doing so we obtain the so-called pre-cyclic solution, that is the solution prior to the cyclic summing over $\{\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3\}$, which assures that all the galaxies contribute to all the density fields.

The variables which characterize the bispectrum are five: three of them are needed to define the triangles' shape (the two sides $\mathbf{k}_1$ and $\mathbf{k}_2$ and the angle between them, θ), while the other two are required to describe the triangles' orientation with respect to the line-of-sight (the polar angle ω and the azimuthal angle ϕ). Following the choice of SFC99, we define:

$$\mu_1 = \mu = \cos\omega = \hat{\mathbf{k}_1}\cdot\hat{z}, \tag{4.7}$$

$$\mu_2 = \mu\cos\theta - \sqrt{1-\mu^2}\sin\theta\cos\phi, \tag{4.8}$$

$$\mu_3 = -\frac{k_1}{k_3}\mu - \frac{k_2}{k_3}\mu_2. \tag{4.9}$$

[1] In the literature it is common to find also the alternative definition $S_2(\mathbf{k}_1,\mathbf{k}_2) = x^2 - \frac{1}{3}$.

We can decompose the tree-level redshift-space galaxy bispectrum B_s using the spherical harmonics $Y_{\ell m}$:

$$B_s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} B_s^{(\ell m)}(k_1, k_2, \theta) Y_{\ell m}(\omega, \phi). \tag{4.10}$$

We consider for simplicity the $m = 0$ terms only, which is equivalent to average over the azimuthal angle ϕ. In this case, it is convenient to decompose the bispectrum B_s using the Legendre polynomials P_ℓ:

$$B_s(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \sum_{\ell=0}^{\infty} B_s^{(\ell)}(k_1, k_2, \theta) P_\ell(\mu). \tag{4.11}$$

By substituting the Z_n kernels into the (4.6), and then performing the Legendre expansion of the result, we obtain an expression for the multipoles of the tree-level redshift-space galaxy bispectrum, which can be arranged as:

$$\begin{aligned} B_s^{(\ell)}(k_1, k_2, \theta) = P(k_1) P(k_2) b_1^4 \Big[& F_2(\mathbf{k}_1, \mathbf{k}_2) D_{SQ1}^{(\ell)} + G_2(\mathbf{k}_1, \mathbf{k}_2) D_{SQ2}^{(\ell)} \\ & + \Big(\gamma + 2\gamma_t \, S_2(\mathbf{k}_1, \mathbf{k}_2) \Big) D_{NLB}^{(\ell)} + D_{FOG}^{(\ell)} \Big] + \text{cyc.}, \end{aligned} \tag{4.12}$$

where $\gamma = b_2/b_1$ and $\gamma_t = b_t/b_1$. $D_{SQ1}^{(\ell)}$ and $D_{SQ2}^{(\ell)}$ represent the first- and second-order contributions for the large-scale squashing (also called Kaiser effect or pancakes-of-God), $D_{NLB}^{(\ell)}$ accounts for the nonlinear bias contribution and $D_{FOG}^{(\ell)}$ refers to the damping effect due to the velocity dispersion (also called fingers-of-God). Here we consider the monopole only, hence the $\ell = 0$ terms. Knowing that $\beta = f/b_1$, we can write the expressions for the different contributions:

$$D_{SQ1}^{(0)} = \frac{2(15 + 10\beta + \beta^2 + 2\beta^2 x^2)}{15 b_1}, \tag{4.13}$$

$$\begin{aligned} D_{SQ2}^{(0)} = & 2\beta(35k_1^2 + 28\beta k_2^2 + 3\beta^2 k_1^2 + 35k_2^2 + 28\beta k_1^2 + 3\beta^2 k_2^2 + 70 k_1 k_2 x \\ & + 84\beta k_1 k_2 x + 18\beta^2 k_1 k_2 x + 14\beta k_1^2 x^2 + 12\beta^2 k_1^2 x^2 + 14\beta k_2^2 x^2 \\ & + 12\beta^2 k_2^2 x^2 + 12\beta^2 k_1 k_2 x^3)/(105 k_3^2 b_1), \end{aligned} \tag{4.14}$$

$$D_{NLB}^{(0)} = \frac{(15 + 10\beta + \beta^2 + 2\beta^2 x^2)}{15 b_1}, \tag{4.15}$$

$$
\begin{aligned}
D^{(0)}_{FOG} = \beta(&210k_1k_2 + 210\beta k_1k_2 + 54\beta^2k_1k_2 + 6\beta^3k_1k_2 + 105k_1^2x \\
&+ 189\beta k_1^2x + 99\beta^2k_1^2x + 15\beta^3k_1^2x + 105k_2^2x + 189\beta k_2^2x \\
&+ 99\beta^2k_2^2x + 15\beta^3k_2^2x + 168\beta k_1k_2x^2 + 216\beta^2k_1k_2x^2 \\
&+ 48\beta^3k_1k_2x^2 + 36\beta^2k_1^2x^3 + 20\beta^3k_1^2x^3 + 36\beta^2k_2^2x^3 \\
&+ 20\beta^3k_2^2x^3 + 16\beta^3k_1k_2x^4)/(315k_1k_2).
\end{aligned} \tag{4.16}
$$

4.2 3PCF model

SE17 developed a model for the redshift-space 3PCF by converting the bispectrum model of SCF99 to configuration space. However, their model includes some calculation errors. For this reason, we propose a revision of the model of SE17. In this section, we present the theoretical derivation of our 3PCF model, which is based on the procedure of SE17, but applied to our bispectrum model, discussed in section 4.1.

Since it will be used later in the calculations, let us consider the Legendre expansion of the arbitrary function $f(r, \theta)$ onto the Legendre polynomials P_ℓ:

$$
f(r, \theta) = \sum_\ell f_\ell(r) P_\ell(x), \tag{4.17}
$$

where r is a spatial variable and $x = \cos\theta$, where θ is an angular variable. After the expansion, the multipole moments $f_\ell(r)$ do not depend on the angular variable anymore, since the angular dependence is included in the Legendre polynomials. If we want an expression for the multipole moments $f_\ell(r)$, we need to solve the following integral:

$$
f_\ell(r) = \frac{2\ell + 1}{2} \int_{-1}^{1} f(r, \theta) P_\ell(x) \mathrm{d}x. \tag{4.18}
$$

When we will discuss the Legendre expansion of a term in the following sections, we will refer to equations (4.17) and (4.18) as an example. For the whole mathematical derivation we are about to discuss, hence to solve all the integrals, we used the Python library for symbolic mathematics SymPy[2] (Meurer et al., 2017).

4.2.1 Legendre decomposition

First of all, we can write a Legendre expansion of the 3PCF as follows (see equation 4.17):

$$
\zeta(r_1, r_2, \hat{r_1} \cdot \hat{r_2}) = \sum_\ell \zeta_\ell(r_1, r_2) P_\ell(\hat{r_1} \cdot \hat{r_2}), \tag{4.19}
$$

[2]The full documentation of SymPy is available at https://docs.sympy.org/latest/citing.html.

where P_ℓ are the Legendre polynomials and the coefficients $\zeta_\ell(r_1, r_2)$ can be obtained through (see equation 4.18):

$$\zeta_\ell(r_1, r_2) = \frac{2\ell+1}{2} \int_{-1}^{1} \mathrm{d}x_{12} P_\ell(x_{12}) \Big[\zeta_{\rm pc}(r_1, r_2, x_{12}) + \zeta_{\rm pc}(r_1, r_3, x_{13}) + \zeta_{\rm pc}(r_2, r_3, x_{23}) \Big], \tag{4.20}$$

where $x_{12} \equiv \cos\theta_{12}$ is the cosine of the angle between r_1 and r_2, $x_{13} \equiv \cos\theta_{13}$ is the cosine of the angle between r_1 and r_3, $x_{23} \equiv \cos\theta_{23}$ is the cosine of the angle between r_2 and r_3 and $\zeta_{\rm pc}$ refers to the pre-cyclic 3PCF.

4.2.2 Pre-cyclic 3PCF

In order to get an expression for $\zeta_{\rm pc}(r_1, r_2, x)$, where we use $x \equiv x_{12}$, we have to cast the bispectrum monopole[3], derived in (4.12), to configuration space. We begin by projecting it on a basis of Legendre polynomials (see equation 4.17):

$$B_s(k_1, k_2, x) = \sum_\ell B_{s,\ell}(k_1, k_2) P_\ell(x). \tag{4.21}$$

This form is convenient, since, as Slepian and Eisenstein, 2015a proved, the inverse Fourier transform of (4.21) is:

$$\zeta_{\rm pc}(r_1, r_2, x) = \sum_\ell \zeta_{{\rm pc},\ell}(r_1, r_2) P_\ell(x), \tag{4.22}$$

with:

$$\zeta_{{\rm pc},\ell}(r_1, r_2) = (-1)^\ell \int \frac{k_1^2 k_2^2 \mathrm{d}k_1 \mathrm{d}k_2}{(2\pi^2)^2} B_{s,\ell}(k_1, k_2) j_\ell(k_1 r_1) j_\ell(k_2 r_2). \tag{4.23}$$

The terms in (4.13), (4.15) and (4.16) can be easily put in this form, while the term in (4.14) causes some difficulties due to its dependence on $1/k_3^2$, with $|\mathbf{k}_3| = |\mathbf{k}_1 + \mathbf{k}_2|$. As noted by SE17, we could expand $1/k_3$ in a Legendre series and then square it, but this would lead to the squaring of an infinite Legendre series with coefficients of the form $(k_1/k_2)^\ell$ for $k_1 < k_2$ and $(k_2/k_1)^\ell$ for $k_1 > k_2$. This poses a significant challenge, since moving to configuration space requires integrals of arbitrary powers of k_1 and k_2, which, if computed independently, may not generally converge. Nevertheless, it has been observed that these integrals converge when calculated as a 2-D integral depending on the ratio k_1/k_2. Therefore, we aim to express the inverse Fourier transform of the terms including k_3 as a product of separable 1-D integrals. For this reason, we can rewrite (4.21) as:

$$B_s(k_1, k_2, x) = \sum_\ell \Big[B_{s,\ell,{\rm no}\,k_3}(k_1, k_2) + B_{s,\ell,{\rm with}\,k_3}(k_1, k_2) \Big] P_\ell(x). \tag{4.24}$$

[3] In order to simplify the notation, we drop the superscript (0) in the bispectrum monopole and write $B_s \equiv B_s^{(0)}$.

4.2.3 Expansion of the k_3 independent terms

We can now proceed to expand the k_3 independent terms in eq.(4.24). For the purpose of clarity and to make a direct comparison with the work of SE17 easier, we address the expansion of each term in (4.12) separately.

Let us consider the $F_2(\mathbf{k}_1, \mathbf{k}_2) D^{(0)}_{SQ1}$ term. Its multipole moments can be computed by solving the integral in equation 4.18 with $F_2(\mathbf{k}_1, \mathbf{k}_2) D^{(0)}_{SQ1}$ instead of $f(r, \theta)$. In particular, we find:

$$\begin{aligned}
\ell = 0: &\quad \frac{34}{21\, b_1}\left[1 + \frac{2}{3}\beta + \frac{49}{425}\beta^2\right] \\
\ell = 1: &\quad \frac{\mathcal{G}(k_1, k_2)}{b_1}\left[1 + \frac{2}{3}\beta + \frac{11}{75}\beta^2\right] \\
\ell = 2: &\quad \frac{8}{21\, b_1}\left[1 + \frac{2}{3}\beta + \frac{18}{35}\beta^2\right] \\
\ell = 3: &\quad \frac{\mathcal{G}(k_1, k_2)}{b_1}\frac{4}{75}\beta^2 \\
\ell = 4: &\quad \frac{64\beta^2}{3675\, b_1},
\end{aligned} \tag{4.25}$$

where:

$$\mathcal{G}(k_1, k_2) \equiv \left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right). \tag{4.26}$$

We have non-zero terms up to $\ell = 4$ because of angular momentum addition, since both $F_2(\mathbf{k}_1, \mathbf{k}_2)$ and $D^{(0)}_{SQ1}$ have x^2 terms, then $\ell = 2$ each.

Let us now take into account the $\gamma D^{(0)}_{NLB}$ term. Its multipole moments can be computed by solving the integral in equation 4.18 with $\gamma D^{(0)}_{NLB}$ instead of $f(r, \theta)$. In particular, we find:

$$\begin{aligned}
\ell = 0: &\quad \frac{\gamma}{b_1}\left[1 + \frac{2}{3}\beta + \frac{1}{9}\beta^2\right] \\
\ell = 2: &\quad \gamma\frac{4\beta^2}{45\, b_1}.
\end{aligned} \tag{4.27}$$

Given that $D^{(0)}_{NLB}$ solely contains terms that are either independent of x or proportional to x^2, we find that only $\ell = 0$ and $\ell = 2$ moments are non-zero.

Let us analyze the $2\gamma_t\, S_2(\mathbf{k}_1, \mathbf{k}_2) D^{(0)}_{NLB}$ term. Its multipole moments can be computed by solving the integral in equation 4.18 with $2\gamma_t\, S_2(\mathbf{k}_1, \mathbf{k}_2) D^{(0)}_{NLB}$ instead of $f(r, \theta)$. In particular, we find:

$$\begin{aligned}
\ell = 0: &\quad -\frac{4}{3}\frac{\gamma_t}{b_1}\left[1 + \frac{2}{3}\beta + \frac{7}{75}\beta^2\right] \\
\ell = 2: &\quad \frac{4}{3}\frac{\gamma_t}{b_1}\left[1 + \frac{2}{3}\beta + \frac{1}{21}\beta^2\right] \\
\ell = 4: &\quad \frac{32}{525\, b_1}\beta^2\gamma_t.
\end{aligned} \tag{4.28}$$

Since here $D^{(0)}_{NLB}$ is multiplied by $S_2(\mathbf{k}_1, \mathbf{k}_2)$, which has a term in x^2, we have moments up to $\ell = 4$ because of angular momentum addition.

Let us now study the $D^{(0)}_{FOG}$ term. Its multipole moments can be computed by solving the integral in equation 4.18 with $D^{(0)}_{FOG}$ instead of $f(r, \theta)$. In particular, we find:

$$\begin{aligned}
\ell = 0: &\quad \frac{2}{3}\beta + \frac{38}{45}\beta^2 + \frac{2}{5}\beta^3 + \frac{2}{25}\beta^4 \\
\ell = 1: &\quad \mathcal{G}(k_1, k_2)\left[\frac{1}{3}\beta + \frac{3}{5}\beta^2 + \frac{67}{175}\beta^3 + \frac{3}{35}\beta^4\right] \\
\ell = 2: &\quad \frac{16}{45}\beta^2 + \frac{16}{35}\beta^3 + \frac{32}{245}\beta^4 \\
\ell = 3: &\quad \mathcal{G}(k_1, k_2)\left[\frac{8}{175}\beta^3 + \frac{8}{315}\beta^4\right] \\
\ell = 4: &\quad \frac{128}{11025}\beta^4.
\end{aligned} \tag{4.29}$$

The multipole moments here reach $\ell = 4$ since $D^{(0)}_{FOG}$ has terms in x^4.

4.2.4 Expansion of the k_3 dependent term

Now we can expand the k_3 dependent term in (4.24), which is $G_2(\mathbf{k}_1, \mathbf{k}_2)D^{(0)}_{SQ2}$, since k_3 only enters through $D^{(0)}_{SQ2}$. In particular, we can rewrite $D^{(0)}_{SQ2}$ as follows:

$$\begin{aligned}
D^{(0)}_{SQ2} = \frac{2}{105\, b_1}\Bigg[&35\beta + \Big(28 + 3\beta + 2x^2\,(7 + 6\beta)\Big) \\
&- \frac{\beta^2}{k_3^2}\, 4\beta^2 k_1 k_2 x\,(7 + 3\beta)\left(x^2 - 1\right)\Bigg],
\end{aligned} \tag{4.30}$$

where we have used the relation:

$$k_1^2 + k_2^2 = k_3^2 - 2k_1k_2x. \tag{4.31}$$

This form is convenient since we have isolated the part of $D^{(0)}_{SQ2}$ which contains k_3.

Let us consider the first term in the square brackets of (4.30), and let us multiply it by $G_2(\mathbf{k}_1, \mathbf{k}_2)$. Let us now compute the multipole moments of the resulting term by solving the integral in equation (4.18) with the result of the

multiplication instead of $f(r,\theta)$:

$$\begin{aligned}
\ell = 0: &\quad \frac{26}{63 b_1}\beta + \frac{628}{1575 b_1}\beta^2 + \frac{346}{3675 b_1}\beta^3 \\
\ell = 1: &\quad \frac{\mathcal{G}(k_1,k_2)}{b_1}\left[\frac{1}{3}\beta + \frac{26}{75}\beta^2 + \frac{17}{175}\beta^3\right] \\
\ell = 2: &\quad \frac{16}{63\, b_1}\beta + \frac{808}{2205\, b_1}\beta^2 + \frac{832}{5145\, b_1}\beta^3 \\
\ell = 3: &\quad \frac{\mathcal{G}(k_1,k_2)}{b_1}\left[\frac{4}{75}\beta^2 + \frac{8}{175}\beta^3\right] \\
\ell = 4: &\quad \frac{128}{3675\, b_1}\beta^2 + \frac{256}{8575\, b_1}\beta^3.
\end{aligned} \tag{4.32}$$

The second term in the square brackets of (4.30) depends on $1/k_3^2$ and, as earlier addressed, we use the strategy proposed by SE17 to compute its inverse Fourier transform. First, let us multiply it by $G_2(\mathbf{k}_1, \mathbf{k}_2)$. Then, let us compute the multipole moments of the resulting term by substituting in the integral of equation (4.18) the result of the multiplication instead of $f(r,\theta)$:

$$\begin{aligned}
\ell = 0: &\quad \frac{8}{1575\, b_1 k_3^2}\beta^2\left(k_1^2 + k_2^2\right)(7 + 3\beta) \\
\ell = 1: &\quad \frac{176}{8575\, b_1 k_3^2}\beta^2 k_1 k_2\,(7 + 3\beta) \\
\ell = 2: &\quad \frac{8}{2205\, b_1 k_3^2}\beta^2\left(k_1^2 + k_2^2\right)(7 + 3\beta) \\
\ell = 3: &\quad -\frac{496}{33075\, b_1 k_3^2}\beta^2 k_1 k_2\,(7 + 3\beta) \\
\ell = 4: &\quad -\frac{32}{3675\, b_1 k_3^2}\beta^2\left(k_1^2 + k_2^2\right)(7 + 3\beta) \\
\ell = 5: &\quad -\frac{256}{46305\, b_1 k_3^2}\beta^2 k_1 k_2\,(7 + 3\beta)\,.
\end{aligned} \tag{4.33}$$

We now use the relation (4.31) in the even multipoles, so that they become dependent of $k_1 k_2$ only. This reintroduces an x-dependence, therefore we need to reproject the result of the substitution, again, onto the Legendre polynomials, computing the integral of equation (4.18). In particular, we need to reproject each ℓ term in equation (4.33) separately. Thus, we compute the integral of equation (4.18) with each ℓ term in equation (4.33) instead of the function $f(r,\theta)$, and we only consider the first moment of the expansion (hence the $\ell = 0$ term resulting from the integral). Therefore, for the k_3 dependent part of the

$G_2(\mathbf{k}_1, \mathbf{k}_2) D^{(0)}_{SQ2}$ term, we find:

$$\begin{aligned}
\ell = 0: &\quad \frac{8}{1575\, b_1} \beta^2 (7 + 3\beta) \\
\ell = 1: &\quad -\frac{16}{1225\, b_1} \frac{k_1 k_2}{k_3^2} \beta^2 (7 + 3\beta) \\
\ell = 2: &\quad \frac{8}{2205\, b_1} \beta^2 \cdot (7 + 3\beta) \\
\ell = 3: &\quad \frac{16}{4725\, b_1} \frac{k_1 k_2}{k_3^2} \beta^2 (7 + 3\beta) \\
\ell = 4: &\quad -\frac{32}{3675\, b_1} \beta^2 (7 + 3\beta) \\
\ell = 5: &\quad \frac{64}{6615\, b_1} \frac{k_1 k_2}{k_3^2} \beta^2 (7 + 3\beta)\,.
\end{aligned} \tag{4.34}$$

4.2.5 Adding all the terms together

We can now sum all the terms in (4.25), (4.27), (4.28), (4.29), (4.32) and the even moments only of (4.34), since the odd ones depend on k_3. By including

also the pre-factor $P(k_1)P(k_2)b_1^4$ in (4.12), we find:

$$
\begin{aligned}
\ell = 0: \; & P(k_1)P(k_2)\Bigg\{ b_1^4 \left[\frac{2}{3}\beta + \frac{38}{45}\beta^2 + \frac{2}{5}\beta^3 + \frac{2}{25}\beta^4 \right] \\
& + b_1^3 \Bigg[\frac{34}{21}\left(1 + \frac{47}{51}\beta + \frac{163}{425}\beta^2 + \frac{201}{2975}\beta^3 \right) \\
& + \gamma \left(1 + \frac{2}{3}\beta + \frac{1}{9}\beta^2 \right) + -\frac{4}{3}\gamma_t \left(1 + \frac{2}{3}\beta + \frac{7}{75}\beta^2 \right) \Bigg] \Bigg\} \\
\ell = 1: \; & P(k_1)P(k_2)\Bigg\{ b_1^4 \left[\frac{1}{3}\beta + \frac{3}{5}\beta^2 + \frac{67}{175}\beta^3 + \frac{3}{35}\beta^4 \right] \\
& + b_1^3 \left[1 + \beta + \frac{37}{75}\beta^2 + \frac{17}{175}\beta^3 \right] \Bigg\} \mathcal{G}\left(k_1, k_2\right) \\
\ell = 2: \; & P(k_1)P(k_2)\Bigg\{ b_1^4 \left[\frac{16}{45}\beta^2 + \frac{16}{35}\beta^3 + \frac{32}{245}\beta^4 \right] \\
& + b_1^3 \Bigg[\frac{8}{21}\left(1 + \frac{4}{3}\beta + \frac{54}{35}\beta^2 + \frac{111}{245}\beta^3 \right) \\
& + \frac{4\beta^2\gamma}{45} + \frac{4}{3}\gamma_t \left(1 + \frac{2}{3}\beta + \frac{1}{21}\beta^2 \right) \Bigg] \Bigg\} \\
\ell = 3: \; & P(k_1)P(k_2)\Bigg\{ b_1^4 \left[\frac{8}{175}\beta^3 + \frac{8}{315}\beta^4 \right] \\
& + b_1^3 \left[\frac{8}{75}\beta^2 + \frac{8}{175}\beta^3 \right] \Bigg\} \mathcal{G}\left(k_1, k_2\right) \\
\ell = 4: \; & P(k_1)P(k_2)\Bigg\{ \frac{128}{11025}\beta^4 b_1^4 \\
& + b_1^3 \left[-\frac{32}{3675}\beta^2 + \frac{32}{8575}\beta^3 + \frac{32}{525}\beta^2\gamma_t \right] \Bigg\}.
\end{aligned}
\tag{4.35}
$$

The k_3-dependent terms, instead, are the odd multipoles of equation (4.34):

$$
\begin{aligned}
\ell = 1: \; & -\frac{16}{1225\, b_1}\frac{k_1 k_2}{k_3^2}\beta^2\left(7 + 3\beta\right) \\
\ell = 3: \; & \frac{16}{4725\, b_1}\frac{k_1 k_2}{k_3^2}\beta^2\left(7 + 3\beta\right) \\
\ell = 5: \; & \frac{64}{6615\, b_1}\frac{k_1 k_2}{k_3^2}\beta^2\left(7 + 3\beta\right).
\end{aligned}
\tag{4.36}
$$

The sum of (4.35) and (4.36) describes the full Legendre decomposition of the line-of-sight averaged bispectrum model. Therefore, the next step consists of the computation of the inverse Fourier transform of these multipole moments.

First, we assess the inverse Fourier transform of the coefficients in (4.35), exploiting (4.22) and (4.23). We define the following integrals of the power spectrum:

$$\begin{aligned}\xi_i^{[n]} &= \int \frac{k^2 \mathrm{d}k}{2\pi^2} P(k) j_n(kr_i) \\ \xi_i^{[n\pm]} &= \int \frac{k^2 \mathrm{d}k}{2\pi^2} k^{\pm 1} P(k) j_n(kr_i),\end{aligned} \tag{4.37}$$

where $j_n(kr_i)$ is the Bessel function of the first kind of order n and the subscript i refers to one of the three sides of the triangles. These expressions will be useful in presenting the result more concisely.

Then, we consider the inverse Fourier transform of the coefficients in (4.36). As it is shown in Appendix A of SE17, the k_3-dependent terms, which contribute to the odd moments only in Fourier space, enter in all the ℓ-terms in configuration space, as follows:

$$0 \leq \ell < \infty : \quad b_1^3 \beta^2 \, (7 + 3\beta) \, \kappa_\ell(r_1, r_2), \tag{4.38}$$

with:

$$\kappa_\ell(r_1, r_2) = \frac{64}{77175} \Big[9 I_{1\ell}(r_1, r_2) - 14 I_{3\ell}(r_1, r_2) + 5 I_{5\ell}(r_1, r_2) \Big], \tag{4.39}$$

$$\begin{aligned}I_{\mathcal{L}\ell}(r_1, r_2) = &\sum_{l_1} (-1)^{l_1+\ell} (2l_1 + 1)(2\ell + 1) \begin{pmatrix} l_1 & \ell & \mathcal{L} \\ 0 & 0 & 0 \end{pmatrix}^2 \\ &\times \int r \mathrm{d}r \, f_{\ell l_1}(r_1; r) f_{\ell l_1}(r_2; r),\end{aligned} \tag{4.40}$$

$$f_{\ell l_1}(r_1; r) = \int \frac{k^2 \mathrm{d}k}{2\pi^2} j_\ell(kr_2) j_{l_1}(kr) P(k). \tag{4.41}$$

Thus, the inverse Fourier transform of the line-of-sight averaged bispectrum

model, which is the line-of-sight averaged 3PCF model, is:

$$
\begin{aligned}
\ell = 0: \quad & \xi_1^{[0]}\xi_2^{[0]}\Bigg\{ b_1^4\left[\frac{2}{3}\beta + \frac{38}{45}\beta^2 + \frac{2}{5}\beta^3 + \frac{2}{25}\beta^4\right] \\
& + b_1^3\Bigg[\frac{34}{21}\left(1 + \frac{47}{51}\beta + \frac{163}{425}\beta^2 + \frac{201}{2975}\beta^3\right) \\
& + \gamma\left(1 + \frac{2}{3}\beta + \frac{1}{9}\beta^2\right) - \frac{4}{3}\gamma_t\left(1 + \frac{2}{3}\beta + \frac{7}{75}\beta^2\right)\Bigg]\Bigg\} \\
& + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_0(r_1, r_2) \\
\ell = 1: \quad & - \left[\xi_1^{[1+]}\xi_2^{[1-]} + \xi_2^{[1+]}\xi_1^{[1-]}\right]\Bigg\{ b_1^4\left[\frac{1}{3}\beta + \frac{3}{5}\beta^2 + \frac{67}{175}\beta^3 + \frac{3}{35}\beta^4\right] \\
& + b_1^3\left[1 + \beta + \frac{37}{75}\beta^2 + \frac{17}{175}\beta^3\right]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_1(r_1, r_2) \\
\ell = 2: \quad & \xi_1^{[2]}\xi_2^{[2]}\Bigg\{ b_1^4\left[\frac{16}{45}\beta^2 + \frac{16}{35}\beta^3 + \frac{32}{245}\beta^4\right] \\
& + b_1^3\Bigg[\frac{8}{21}\left(1 + \frac{4}{3}\beta + \frac{54}{35}\beta^2 + \frac{111}{245}\beta^3\right) \\
& + \frac{4\beta^2\gamma}{45} + \frac{4}{3}\gamma_t\left(1 + \frac{2}{3}\beta + \frac{1}{21}\beta^2\right)\Bigg]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_2(r_1, r_2) \\
\ell = 3: \quad & - \left[\xi_1^{[3+]}\xi_2^{[3-]} + \xi_2^{[3+]}\xi_1^{[3-]}\right]\Bigg\{ b_1^4\left[\frac{8}{175}\beta^3 + \frac{8}{315}\beta^4\right] \\
& + b_1^3\left[\frac{8}{75}\beta^2 + \frac{8}{175}\beta^3\right]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_3(r_1, r_2) \\
\ell = 4: \quad & \xi_1^{[4]}\xi_2^{[4]}\left\{\frac{128}{11025}\beta^4 b_1^4 + b_1^3\left[-\frac{32}{3675}\beta^2 + \frac{32}{8575}\beta^3 + \frac{32}{525}\beta^2\gamma_t\right]\right\} \\
& + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_4(r_1, r_2) \\
\ell \geq 5: \quad & b_1^3\beta^2\left(7 + 3\beta\right)\kappa_\ell(r_1, r_2).
\end{aligned}
\tag{4.42}
$$

4.3 Correction of the SE17 3PCF model

In the previous section, we derived our 3PCF model, expressed in equation (4.42). Our 3PCF model is different from the 3PCF model of SE17, which includes some errors. To make the comparison between the two 3PCF models easier, our specific corrections of the SE17 model are:

- we added a missing b_1^{-1} factor in the expansions of the $F_2(\mathbf{k}_1, \mathbf{k}_2)D^{(0)}_{SQ1}$ (equation (4.25) in this work and equation (6) in SE17), $\gamma D^{(0)}_{NLB}$ (equation (4.27) in this work and equation (8) in SE17) and $G_2(\mathbf{k}_1, \mathbf{k}_2)D^{(0)}_{SQ2}$ (equations (4.32) and (4.34) in this work and equations (11) and (13) in SE17) terms;

- we completely changed the term corresponding to the non-local bias parameter, whose expansion is shown in equation (4.28), while in SE17 this contribution is given by equation (22).

These corrections result in our corrected 3PCF model presented in equation (4.42), in opposition to equation (21) of SE17. The difference between the two 3PCF models can be visualized in Fig. 4.1, where the 3PCF $\zeta(r_{12}, r_{13}, \theta)$ is shown as a function of the angle θ, rescaled as θ/π. The two models have been computed using the model parameters obtained by Veropalumbo et al., 2022, which are $b_1 = 2.7$, $b_2 = 0.75$, $b_t = -0.75$, and they are represented along with the real-space MINERVA 3PCF at the fixed illustrative scales $r_{12} = 25\,h^{-1}$Mpc and $r_{13} = 45\,h^{-1}$Mpc. It is evident that the new 3PCF model presented in this work (red curve) better approximates the data (black circles), compared to the SE17 model (blue curve).

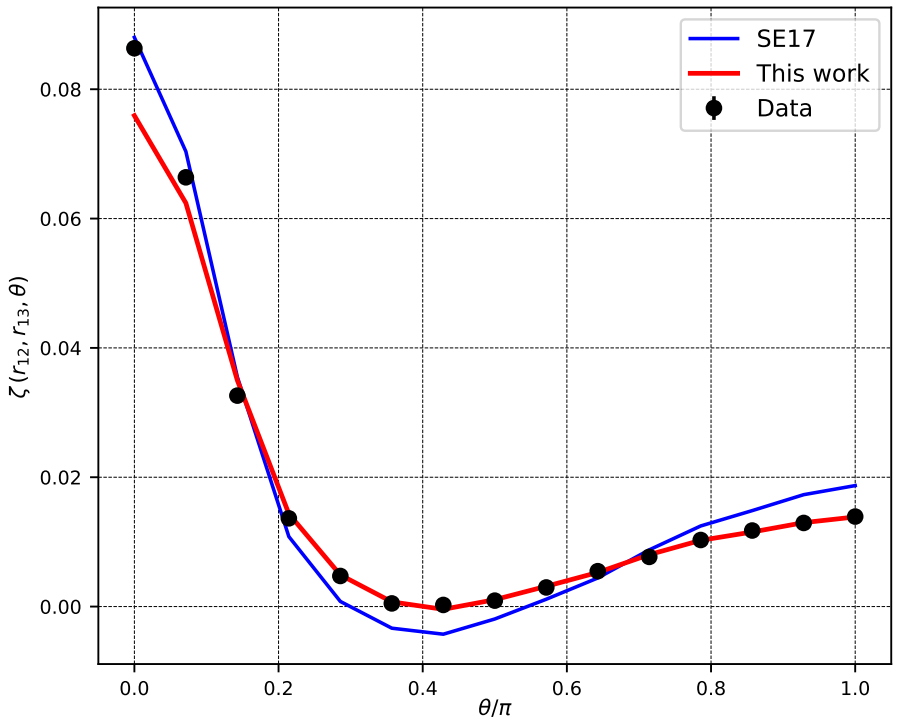

Figure 4.1: Comparison between the corrected 3PCF model presented in this work (red curve) and the 3PCF model of SE17 (blue curve) with the model parameters obtained by Veropalumbo et al., 2022. The black circles represent the data coming from the MINERVA simulations for the scales $r_{12} = 25\,h^{-1}$Mpc and $r_{13} = 45\,h^{-1}$Mpc. The error bars are the square root of the diagonal elements of the covariance matrix, estimated from the mock catalogues, divided by the square root of the number of mock catalogues, and they are not visible since they are very small.

Chapter 5

Analysis framework and methodology

One of the primary objectives of this Thesis work is to develop analytical and numerical methods for the combined analysis of 2PCF and 3PCF. Achieving this purpose necessitated the development of new theoretical models for the bispectrum and the 3PCF (see chapter 4), and the implementation of new methods to estimate the covariance matrix, which has to take into account the correlation between the 2PCF and 3PCF. In this chapter, we present and discuss the codes produced in this regard as part of this work, implemented within the *CosmoBolognaLib* (CBL)[1] framework, using C++ programming language. First, we consider the improvements we introduced to the handling of covariance matrices, whose computation is made more user-friendly. Then, we deal with the modifications regarding the 2PCF modelling, focusing on the 2PCF multipoles. In particular, we implemented a new function for the model in which the tidal bias parameter b_t is free to vary in the fit and the parametrization is consistent with the one of the 3PCF model. Furthermore, we describe the developments concerning the 3PCF modelling, which are one of the major results of this Thesis project. Specifically, we implemented our new model for the 3PCF (see section 4.2). Finally, we show four tests with which we validated our new 3PCF model, comparing it with the 3PCF model of Veropalumbo et al., 2022.

5.1 CosmoBolognaLib

The CBL[2] are a large set of *free software* C++/Python libraries for cosmological calculations. These libraries provide a wide computational framework that can be exploited, among many other things, to measure the 2PCF and 3PCF and to model RSD and BAO. Indeed, the CBL include different estimators and

[1] For more information visit https://gitlab.com/federicomarulli/CosmoBolognaLib.

[2] The full documentation of the CosmoBolognaLib can be found at https://federicomarulli.github.io/CosmoBolognaLib/Doc/html/index.html.

models for two-point and three-point statistics. Furthermore, they can be used to handle catalogues of different populations of objects, such as galaxies, galaxy clusters or cosmic voids.

5.2 Implementation of the full 2PCF+3PCF covariance matrix

Let us consider two variables x_i and x_j. The covariance matrix C_{ij} between these two variables is defined as:

$$C_{ij} \equiv E\left[(x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle)\right], \tag{5.1}$$

where $E(x)$ is the expectation value of the variable x, and $\langle x \rangle$ is the mean value of the variable x.

Inside the CBL, there is a class, `cbl::data::CovarianceMatrix`, developed to compute and manage the covariance matrices for various data sets. This class offers the possibility to either read the covariance matrix from a file or compute it from the provided data. It is able to manage multiple sets of data, providing their covariance and cross-covariance, as well as other related quantities, such as the correlation matrix, precision matrix, variance, and standard deviation.

With the purpose of making the method used to measure the covariance matrix of a collection of data sets (see appendix A) more user-friendly, mainly for Python users, we implemented two new functions for the estimate of the covariance matrix from mock catalogues, which are then used to compute the covariance matrices of the data analyzed in chapter 6. Specifically, they are exploited to estimate the covariance matrix of the 2PCF and of the 3PCF, and the total covariance matrix of the two statistics together.

5.3 Two-point correlation function

The modelling of the 2PCF implemented in the CBL allows the user to exploit the eTNS model, described in equations (3.6), (3.9) and (3.10) in chapter 3. It is possible to model the multipoles of the 2PCF with the class `cbl::modelling::twopt::Modelling_TwoPointCorrelation_multipoles`. This class provides a range of functions, such as e.g. to define the fit range and store the model in an output file. It also offers the possibility to select among different models to be utilized for the fitting process, such as e.g. the Scoccimarro model (Scoccimarro, 2004), the de-wiggled model (Vargas-Magaña et al., 2018), the TNS model, and, as said, the eTNS model, the latter two described in section 3.2 (see Garcıéa-Farieta et al., 2020 for a description of the implementations).

In this work, we adopt the eTNS model for the 2PCF multipoles, since it is the model providing the most accurate estimate of $f\sigma_8$ (Beutler et al., 2014; Di Benedetto, 2023). Previously, the tidal bias parameter b_t was fixed at the

value given by local Lagrangian biasing (Baldauf et al., 2012; Chan et al., 2012):

$$b_t = -\frac{4}{7}(b_1 - 1). \tag{5.2}$$

However, since b_t is a free parameter in the 3PCF analysis, we need it to be a free parameter of the fit also in the analysis of the 2PCF, so that we can constrain it by combining these statistics. For this reason, the model functions were modified in order to let the parameter b_t free to vary (see appendix B).

5.4 Three-point correlation

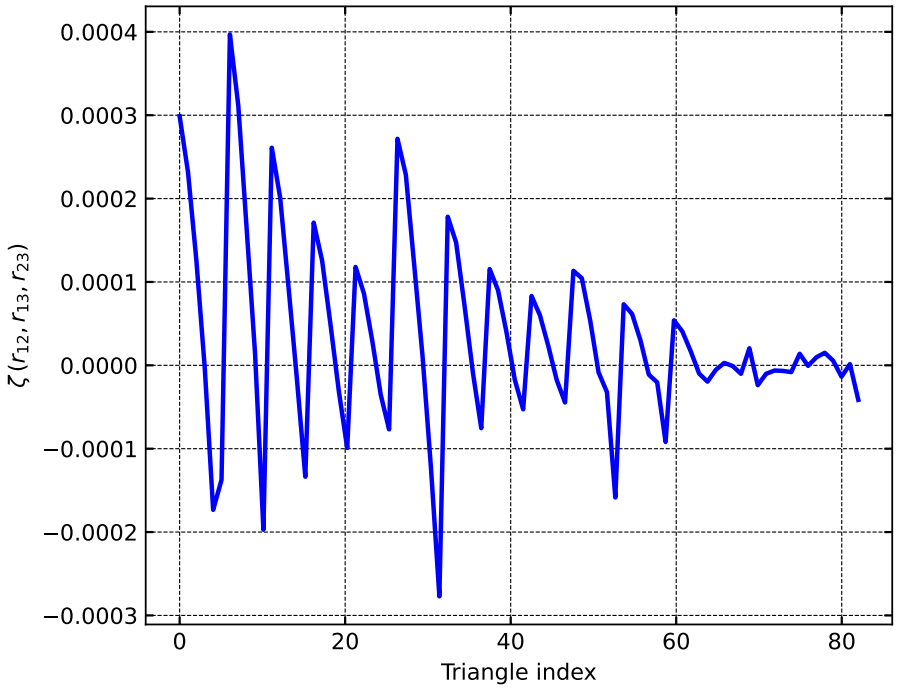

Figure 5.1: 3PCF model at all scales obtained with the best-fit parameters and the cut in scales of the analysis carried out in section 6.6. The $\zeta(r_{12}, r_{13}, r_{23})$ is shown as a function of the triangle configuration $\{r_{12}, r_{13}, r_{23}\}$, identified by an index.

Within the CBL framework, it is possible to model the comoving connected 3PCF using the class `cbl::modelling::threept::Modelling_ThreePointCorrelation_comoving_connected`. In particular, the CBL contained different models to analyze the 3PCF. There were two models for the real space 3PCF, based on Slepian and Eisenstein, 2015a and Barriga and Gaztañaga, 2002 respectively, and a model for the redshift space 3PCF, based on SE17.

However, as we have discussed in chapter 4, the SE17 model contains some calculation errors, implying that the model for the redshift space 3PCF implemented in the CBL was not correct. Therefore, we implemented our new model for the 3PCF (see section 4.2) from scratch, organizing it into multiple functions, crafted to optimize both the computational efficiency and the clarity of the implementation.

Furthermore, the existing parametrization only allowed the study of the 3PCF at fixed scales, varying the angles between two sides of the triangles. Thus, we implemented new functions which allows to consider a different parametrization, that is expressing the 3PCF as a function of the three sides of the triangles (see appendix C for a more detailed description). This enables to analyze the 3PCF at all scales, as done in other works (Veropalumbo et al., 2021; Veropalumbo et al., 2022). A representative plot of the new 3PCF model at all scales is shown in Fig. 5.1. The 3PCF is represented as a function of the triangle configuration $\{r_{12}, r_{13}, r_{23}\}$, identified by an index. The indices are organized in a way such that the configurations change through a cascading iteration. We fix the first side r_{12} while sequentially varying the second side r_{13}, and subsequently, when the second side is held constant, the third side r_{23} is iterated through. The range of scales is between $5\,h^{-1}$Mpc and $145\,h^{-1}$Mpc, with a binning size of $10\,h^{-1}$Mpc.

5.4.1 Testing our model

To validate our new model for the 3PCF, we carried out four tests, comparing it to the 3PCF model used by Veropalumbo et al., 2022 for their analysis. In particular, we considered the multipoles moments of the two models after having expanded them onto Legendre polynomials. The model parameters used in the four tests are shown in Tab. 5.1.

	b_1	b_2	b_t	β
Test 1	1	0	0	0
Test 2	1	1	0	0
Test 3	1	0	1	0
Test 4	1	0	0	1

Table 5.1: Model parameters of the four tests we carried out to validate our 3PCF model.

We tested the two models at various different scales, but we show only the result of the comparison for the scales $r_{12} = 45\,h^{-1}$Mpc and $r_{13} = 105\,h^{-1}$Mpc as a representative example, since it includes the BAO scales as a function of the angle θ. As Figs. 5.2-5.5 display, for all tests the models agree, with an average difference on ζ_ℓ smaller than 3×10^{-6}. The small discrepancies that we find are due to the different numerical methods used in the integration. In fact, Veropalumbo et al., 2022 uses a fast Fourier transform (FFT) algorithm,

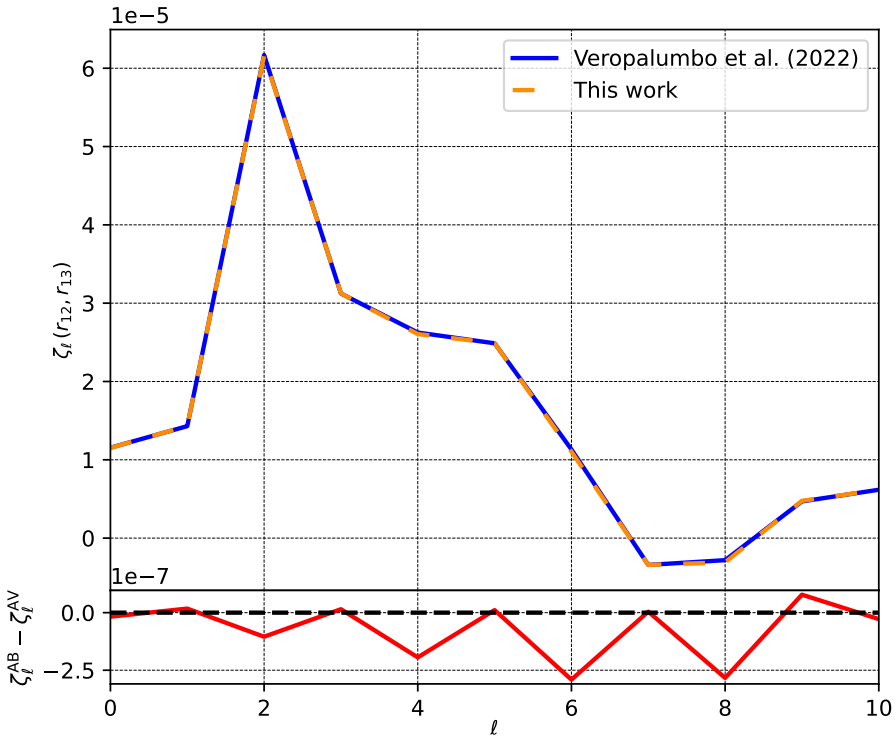

Figure 5.2: *Top panel*: comparison of the multipoles moments of our new 3PCF model (AB, orange) and the one of Veropalumbo et al., 2022 (AV, blue) for test 1 ($b_1 = 1$, $b_2 = 0$, $b_t = 0$, $\beta = 0$). *Bottom panel*: difference between the two models.

whereas in this work we use the quadrature method. Thus, we can conclude that the two models are compatible, implying that our corrected model works properly.

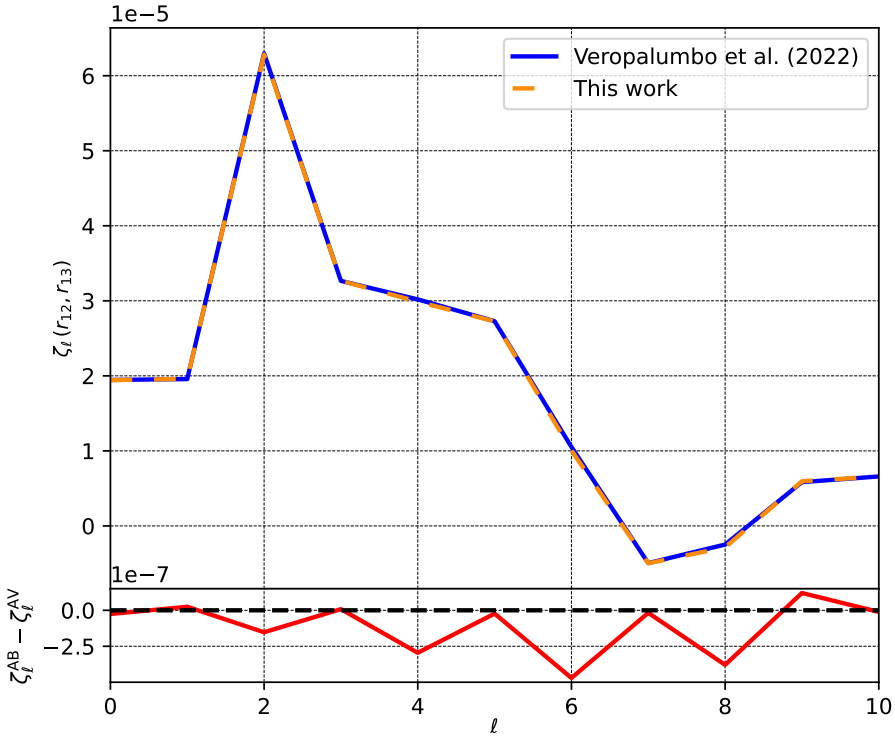

Figure 5.3: As in Fig. 5.2, but for test 3 ($b_1 = 1$, $b_2 = 1$, $b_t = 0$, $\beta = 0$).

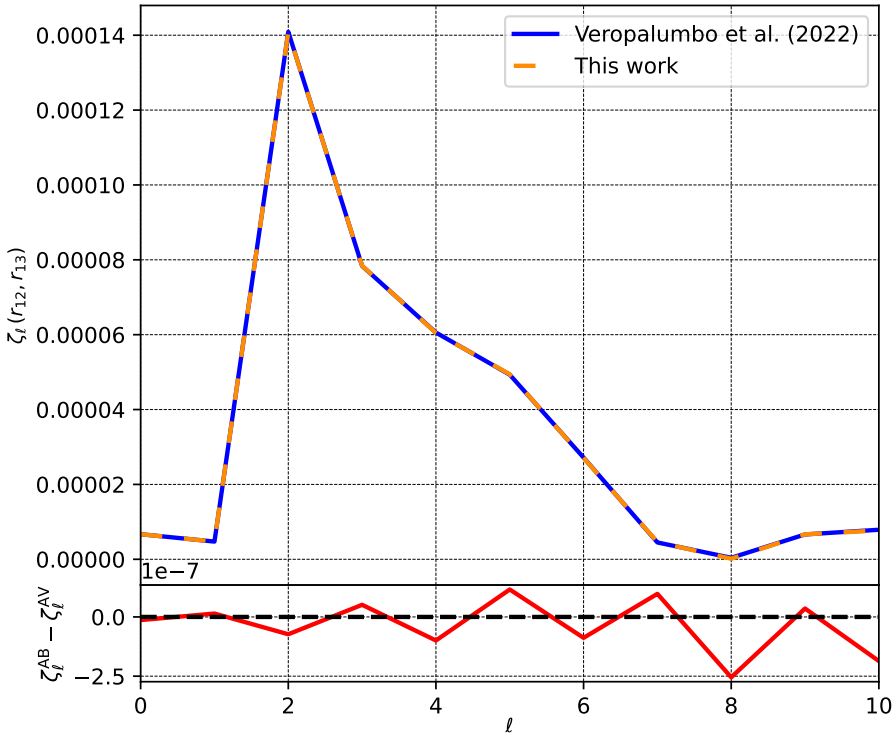

Figure 5.4: As in Fig. 5.2, but for test 3 ($b_1 = 1$, $b_2 = 0$, $b_t = 1$, $\beta = 0$).

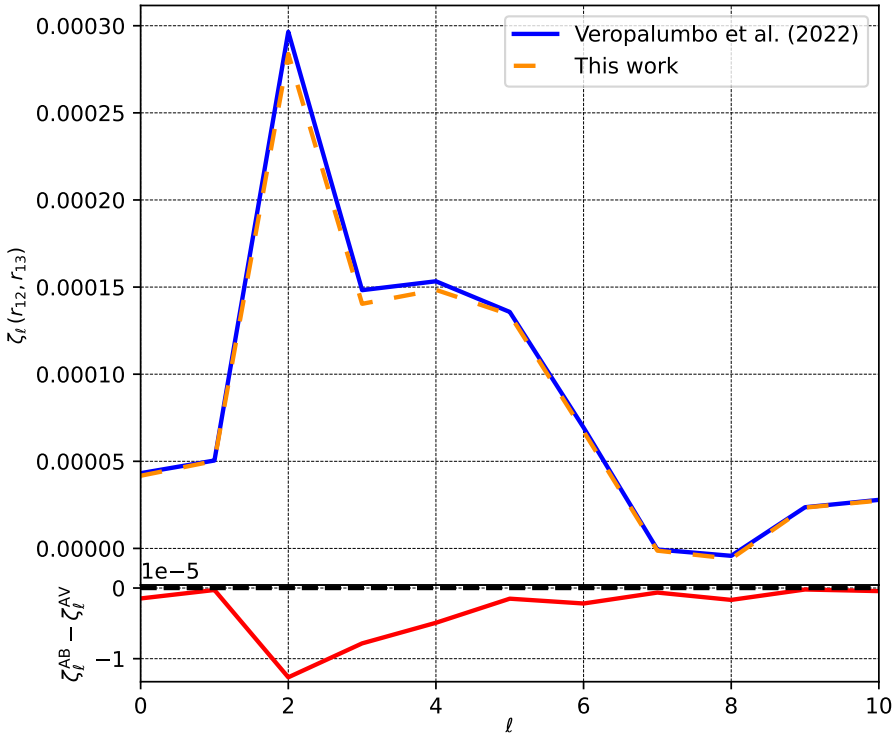

Figure 5.5: As in Fig. 5.2, but for test 3 ($b_1 = 1$, $b_2 = 0$, $b_t = 0$, $\beta = 1$).

Chapter 6

Cosmological parameter estimation

One of the primary goals of this Thesis work is to prepare the necessary framework for a joint likelihood analysis of the 2PCF and 3PCF to estimate cosmological parameters. In this chapter, we present our results based on the MINERVA 2PCF and 3PCF, and the developed statistical method and analysis for the joint analysis of these statistics. First, we introduce the concepts of Bayesian analysis and Monte Carlo Markov Chain method. Then, we describe our dataset, represented by the 298 MINERVA mock catalogues. After that, we illustrate how we derived the covariance matrices we used in our analyses, by estimating it from the MINERVA mocks. Then, we show and discuss the results of the analyses regarding the 2PCF and 3PCF. In the first one, we model the multipoles of the 2PCF using the modified eTNS model (see section 5.3) and sample the posterior distribution using an MCMC method. From this fit, we obtain estimates of the parameters of the eTNS model. In the second analysis, we model the 3PCF both at fixed scales and at all scales using our new model (see sections 4.2 and 5.4), again, using an MCMC method to sample the posterior distribution. As a result, we derive estimates of the parameters of our new model. Finally, we present the developed environment for the joint analysis of 2PCF and 3PCF, and discuss the obtained results.

6.1 Bayesian analysis

Bayesian analysis is a statistical approach for making probabilistic inferences about unknown quantities based on available data and prior knowledge. The starting point of this method is the Bayes' theorem. Let us consider two events A and B. The Bayes' theorem asserts that:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{6.1}$$

where $P(A|B)$ represents the conditional probability of the event A given the occurrence of the event B, and similarly, $P(B|A)$ represents the conditional probability of the event B given the occurrence of the event A. Let us assume to have a set of data D and a model $M(\theta)$, where θ is the vector of model parameters. We define the posterior distribution $P(M(\theta)|D)$ as the conditional probability of obtaining the model $M(\theta)$ given the data D. Therefore, using Bayes' theorem, we can write:

$$P(M(\theta)|D) = \frac{P(D|M(\theta))P(M(\theta))}{P(D)}, \tag{6.2}$$

where:

- $P(D|M(\theta))$ is the likelihood function, which quantifies the plausibility of the observed data given a specific set of model parameters. It is usually denoted as $\mathcal{L}(D|M(\theta))$ and, typically, in cosmology we assume Gaussian likelihoods.
- $P(M(\theta))$ is the prior distribution, which is the probability of the model prior to the observed data. It represents our prior knowledge about the process and it could be, for example, the posterior distribution of a previous experiment. It is usually written as $\pi(M(\theta))$.
- $P(D)$ is the evidence and it can be expressed as:

$$P(D) = \int \mathcal{L}(D|M(\theta))\pi(M(\theta))\mathrm{d}\theta. \tag{6.3}$$

  Since the data do not change, the evidence is a constant.

Thus, we can rewrite equation (6.2) as:

$$P(M(\theta)|D) \propto \mathcal{L}(D|M(\theta))\pi(M(\theta)), \tag{6.4}$$

where the evidence has been omitted since it is only a constant factor in which we are not interested here, and it does not depend on the model.

6.2 Monte Carlo Markov Chain method

The Monte Carlo Markov Chain (MCMC) method is a computational technique used for sampling from complex probability distributions, such as the posterior distribution. Specifically, this method returns in output points, which we call samples, in parameter space, that are statistically independent and whose distribution follows the posterior distribution. In particular, one of the most common MCMC method is the Metropolis-Hastings algorithm, which is based on the detailed balance, which can be expressed as:

$$K(\mathbf{x}_{n+1}|\mathbf{x}_n)f(\mathbf{x}_n) = K(\mathbf{x}_n|\mathbf{x}_{n+1})f(\mathbf{x}_{n+1}), \tag{6.5}$$

where $\mathbf{x}_n$ and $\mathbf{x}_{n+1}$ are consecutive elements of the Markov chain, the probability K is the Markov chain's transition kernel and f is a stationary distribution. We can think of the detailed balance as the requirement that the flow of points out of state $\mathbf{x}_n$ into state $\mathbf{x}_{n+1}$ is equal to the counter flow of points out of state $\mathbf{x}_{n+1}$ into state $\mathbf{x}_n$. If we start at the point $\mathbf{x}_n$ of parameter space, the Metropolis-Hastings algorithm can be summarized in the following steps:

- we choose a trial point $\mathbf{x}_t$ from a proposal distribution $q(\mathbf{x}_t|\mathbf{x}_n)$, which can be, for example, a Gaussian distribution;
- we calculate the value of α, which is defined as:

$$\alpha(\mathbf{x}_t, \mathbf{x}_n) = \min\left\{1, \frac{q(\mathbf{x}_n|\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_n)} \frac{f(\mathbf{x}_t)}{f(\mathbf{x}_n)}\right\}. \tag{6.6}$$

- if $\alpha < 1$ we draw a number r from a uniform distribution between 0 and 1. If $\alpha > r$ we accept the trial point and we set $\mathbf{x}_{n+1} = \mathbf{x}_t$. Otherwise, we refuse it and we set $\mathbf{x}_{n+1} = \mathbf{x}_n$.

Then, we repeat these steps until the MCMC algorithm converges. In the CBL, the Stretch Move algorithm, described in Goodman and Weare, 2010 and Foreman-Mackey et al., 2013, is implemented.

6.3 The dataset: MINERVA simulations

The MINERVA simulations are a set of 298 N-body simulations performed with the GADGET-2 code (Springel, 2005) with $N = 1000^3$ DM particles per realization in a cubic box of side length $L = 1500\,h^{-1}$Mpc with periodic boundary conditions. The initial redshift of the simulation is $z_{\rm ini} = 63$ and the input linear power spectrum for the initial conditions was calculated using CAMB (Lewis et al., 2000) for the cosmological parameters reported in tab. 6.1, which are the parameters matching the best-fitting ΛCDM model of the WMAP9 + BOSS DR9 $\xi(r)$ analysis (Sánchez et al., 2013). DM halos are identified with a Friends-of-

h	Ω_m	Ω_b	n_s	σ_8
0.695	0.285	0.046	0.9632	0.828

Table 6.1: Cosmological parameters of the MINERVA simulations.

Friends algorithm with a linking length equal to 0.2 times the mean separation between particles. The positions and velocities of particles were stored for five different redshifts: $z \in \{2.0, 1.0, 0.57, 0.3, 0\}$. In this work, however, we consider only the $z = 1$ output and DM halos with a mass $M > 1.12 \times 10^{13}\,h^{-1}M_\odot$, which corresponds to a mean number density $\bar{n} = 2.13 \times 10^{-4}\,h^3{\rm Mpc}^{-3}$, since this is approximately the number density of H-α line emission galaxies which will be the target of forthcoming redshift surveys, such as Euclid.

6.4 Covariance matrices

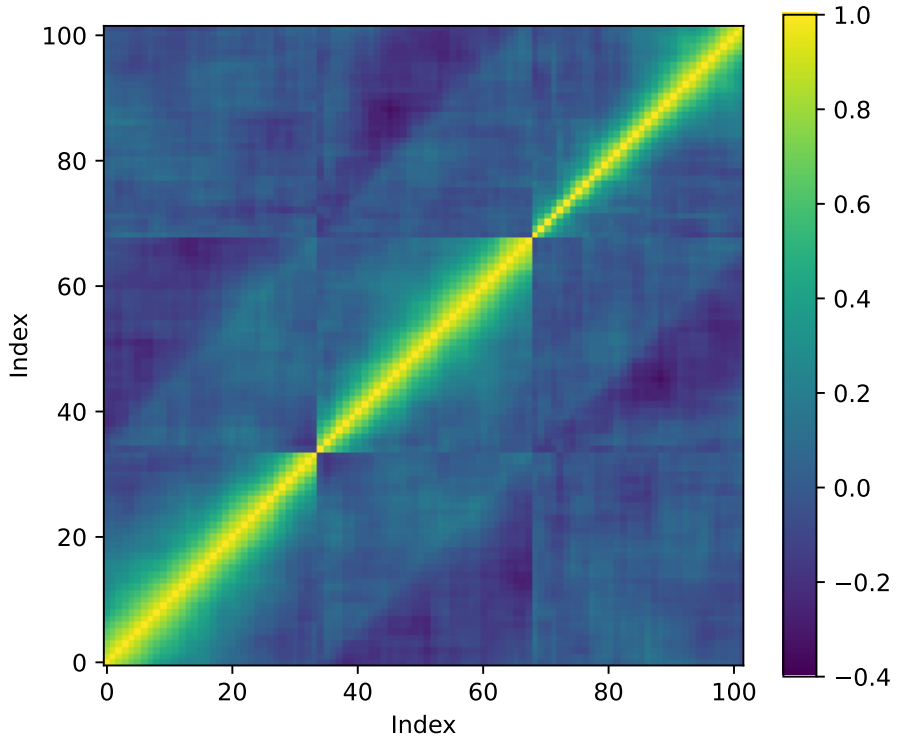

Figure 6.1: Correlation matrix used in the MINERVA 2PCF only analysis, measured from the 298 MINERVA mock catalogues. The matrix is composed by nine square sub-blocks. The three diagonal sub-blocks correspond to the 2PCF monopole, quadrupole and hexadecapole, respectively in the bottom-left, central and top-right regions. The other sub-blocks represent the cross-correlation between the different moments.

For all the three analyses, we built the likelihood function using the covariance matrix measured from the 298 MINERVA mock catalogues, exploiting the new implemented CBL function `CovarianceMatrix::measure()` (see section 5.2). The covariance matrix C_{ij} has been computed using the following maximum likelihood estimator (Barlow, 1991):

$$\hat{C}_{ij}^{\mathrm{ML}} = \frac{1}{N_m - 1} \sum_{k=1}^{N_m} (d_i^{(k)} - \bar{d}_i)(d_j^{(k)} - \bar{d}_j), \tag{6.7}$$

where N_s is the number of mock catalogues, $d_i^{(k)}$ is the data in the i-th bin and k-th mock catalogue, and $\bar{d}_i$ is the mean value of the data in the i-th bin

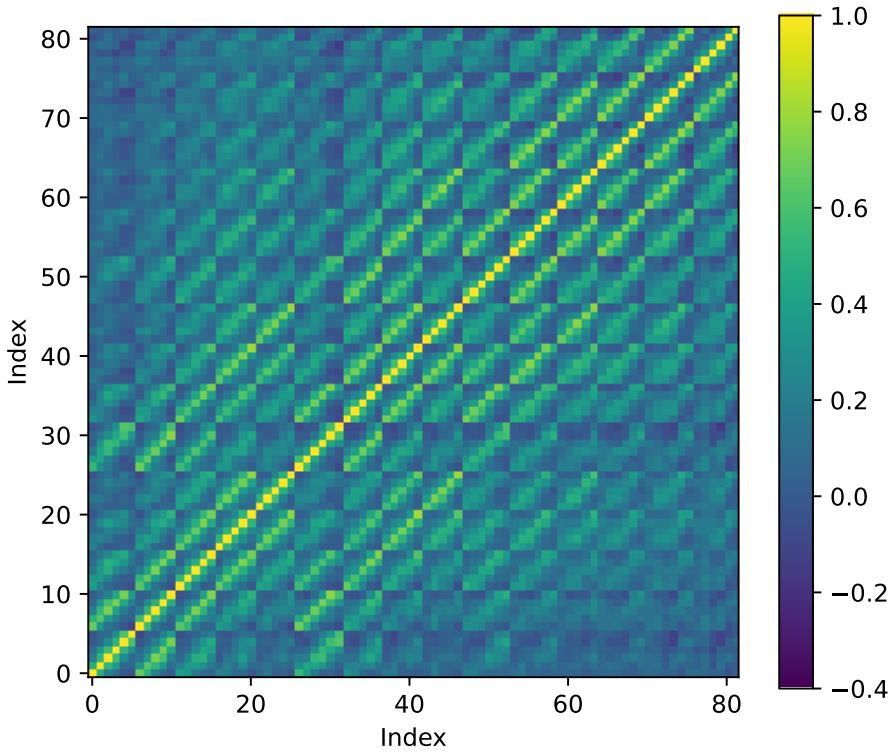

Figure 6.2: Correlation matrix used in the MINERVA 3PCF only analysis, measured from the 298 MINERVA mock catalogues.

(averaged over all the mock catalogues). The size of the covariance matrix depends on the number of bins used in the data vector. Specifically, if N_b is the number of bins, the covariance matrix will be a $N_b \times N_b$ matrix. It is important that the number of bins N_b does not exceed the number of mocks N_m, because the covariance matrix estimator in equation (6.7) yields a singular, hence non-invertible, matrix in the case of $N_b > N_m$.

To compute the likelihood function $\mathcal{L}$, we need the inverse of the covariance matrix, also called precision matrix. However, if we invert the estimator in equation (6.7), we obtain an estimator of the precision matrix, that we call $\hat{C}_*^{-1} = (\hat{C}^{\mathrm{ML}})^{-1}$, which is consistent, but not unbiased. In order to make it unbiased, we need to add a normalization factor (Hartlap et al., 2007), so that the unbiased estimator of the precision matrix becomes:

$$\hat{C}^{-1} = 1 - \frac{N_b + 1}{N_m - 1}\hat{C}_*^{-1}, \quad \text{for } N_b < N_m - 2. \tag{6.8}$$

Neglecting this normalization factor results in an underestimation of the size of the confidence regions, making the log-likelihood function steeper. However, the

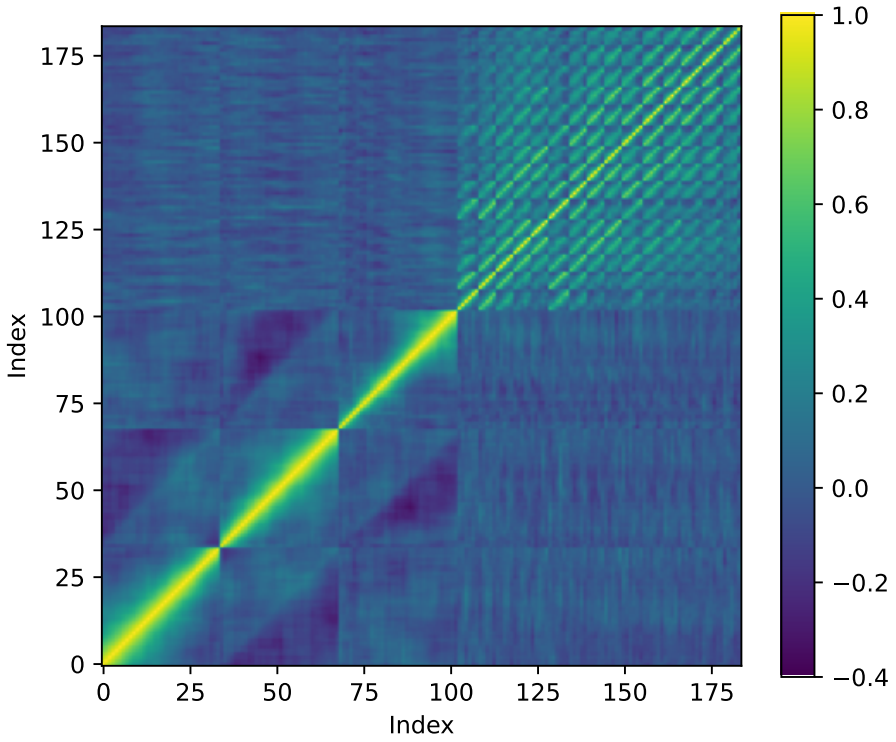

Figure 6.3: Correlation matrix used in the joint analysis of the MINERVA 2PCF and 3PCF, measured from the 298 MINERVA mock catalogues. We can identify the two sub-blocks corresponding to the 2PCF multipoles (bottom-left) and 3PCF (top-right). The other sub-blocks represent the cross-correlation between the two probes.

introduction of this factor does not change the point of maximum likelihood, implying that it does not modify the estimate of the parameters, while it only affects their errors.

Furthermore, the estimate of the covariance matrix expressed in equation (6.7) has a numerical error itself. Therefore, we have to take this error into account when we compute the errors on the parameters. Specifically, let us consider a parameter P, whose posterior distribution has a standard deviation σ_P, which can be considered the error on the estimate of P. The correction that we have to include is the following (Percival et al., 2014):

$$\sigma_{P,\mathrm{corr}} = \sigma_P \sqrt{\frac{1 + B(N_b - N_p)}{1 + A + B(N_p + 1)}} \tag{6.9}$$

where N_p is the number of parameters, and A and B are defined as:

$$A = \frac{2}{(N_m - N_b - 1)(N_m - N_b - 4)}, \tag{6.10}$$

$$B = \frac{N_m - N_b - 2}{(N_m - N_b - 1)(N_m - N_b - 4)}. \tag{6.11}$$

Since it is easier to visualize, instead of showing the covariance matrix, we show the correlation matrix of our data, which is defined as:

$$\mathrm{Corr}_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \tag{6.12}$$

In Fig. 6.1, we show the correlation matrix of the multipoles of the MINERVA 2PCF. Since we have 34 bins for each multipole moment, we have 102 bins in total, implying that the covariance matrix is a 102×102 matrix.

In Fig. 6.2, we show the correlation matrix of the MINERVA 3PCF, computed using all the available scales. After the cut in the scales (see section 6.6), we are left with 82 bins, therefore the covariance matrix is a 82×82 matrix. We can appreciate here that the cross-correlation terms in the 3PCF are much more relevant, compared with the 2PCF multipoles.

In Fig. 6.3, we show the total correlation matrix of the MINERVA 2PCF and 3PCF.

Once we have found the covariance matrix, computing the likelihood function is a straightforward process. We assume a multi-variate Gaussian likelihood, so that it can be written as:

$$\mathcal{L} = \frac{1}{(2\pi)^{N_b/2}|C|^{1/2}} e^{-\frac{1}{2}\sum_{ij}[d_i - \mu_i(\theta)]C_{ij}^{-1}[d_j - \mu_j(\theta)]}, \tag{6.13}$$

where $|C_{ij}| \equiv \det C_{ij}$, d_i is the data at the i-th bin, and $\mu_i(\theta)$ is the model prediction at the i-th bin. The posterior distribution of each parameter is then obtained by multiplying the likelihood function to the prior distribution.

6.5 The analysis of 2PCF in the MINERVA simulations

In the first analysis, we performed the fit of the MINERVA 2PCF multipoles using the eTNS model (see section 3.2). We sampled the posterior distribution using an MCMC method with 64 walkers and the chain size parameter set to 3000. The model parameters are $f\sigma_8$, $b_1\,\sigma_8$, $b_2\,\sigma_8$, $b_t\,\sigma_8$, σ_v, $\alpha_\perp$ and $\alpha_\parallel$, where f is the growth rate of cosmic structures, σ_8 is the amplitude of matter density fluctuations, b_1 is the linear bias, b_2 is the nonlinear bias, σ_v is the velocity dispersion, $\alpha_\perp$ and $\alpha_\parallel$ are the parameters related to the Alcock-Paczynski effect defined in equations (2.62) and (2.63). In this analysis, we fixed the b_t parameter to the value expected from the local Lagrangian bias model (equation 5.2), since

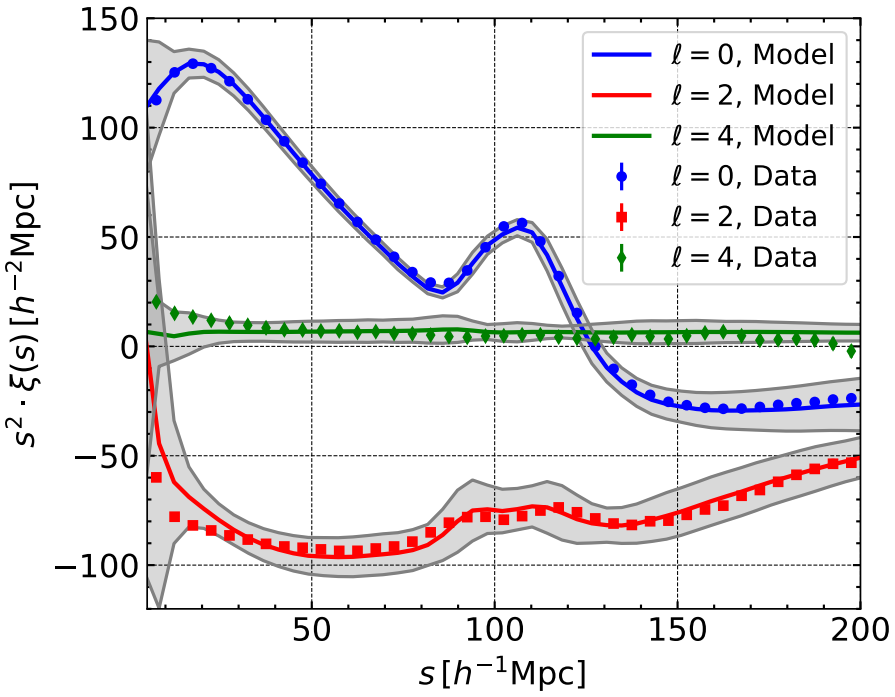

Figure 6.4: Multipole moments (monopole in blue, quadrupole in red, hexadecapole in green) of the MINERVA 2PCF. Circles, squares and diamonds represent the data. The error bars are the square root of the diagonal elements of the covariance matrix divided by the square root of the number of mock catalogues, and they are not visible since they are very small. The solid curves show the best-fit model, while the grey shaded areas display the 1-σ error on the model.

with the 2PCF only we are not able to constrain it and allowing it vary worsened the fit of the other parameters. Therefore, we assume a uniform prior for the parameter $b_t\sigma_8$ equal to the value of b_t derived from equation (5.2) times the value of σ_8 of the MINERVA simulations (see table 6.1) consistently rescaled at the considered redshift. All the other parameters of the model are free to vary, and their priors are assumed to be uniform, as table 6.2 shows. The

$f\sigma_8$	$b_1\,\sigma_8$	$b_2\,\sigma_8$	σ_v	$\alpha_\perp$	$\alpha_\parallel$
$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$	$\mathcal{U}(0,10)$	$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$

Table 6.2: Prior distributions for the free parameters used in the MINERVA 2PCF analysis, where $\mathcal{U}(a,b)$ indicates a uniform distribution between a and b.

data and the best-fit model are shown in Fig. 6.4. It appears evident that

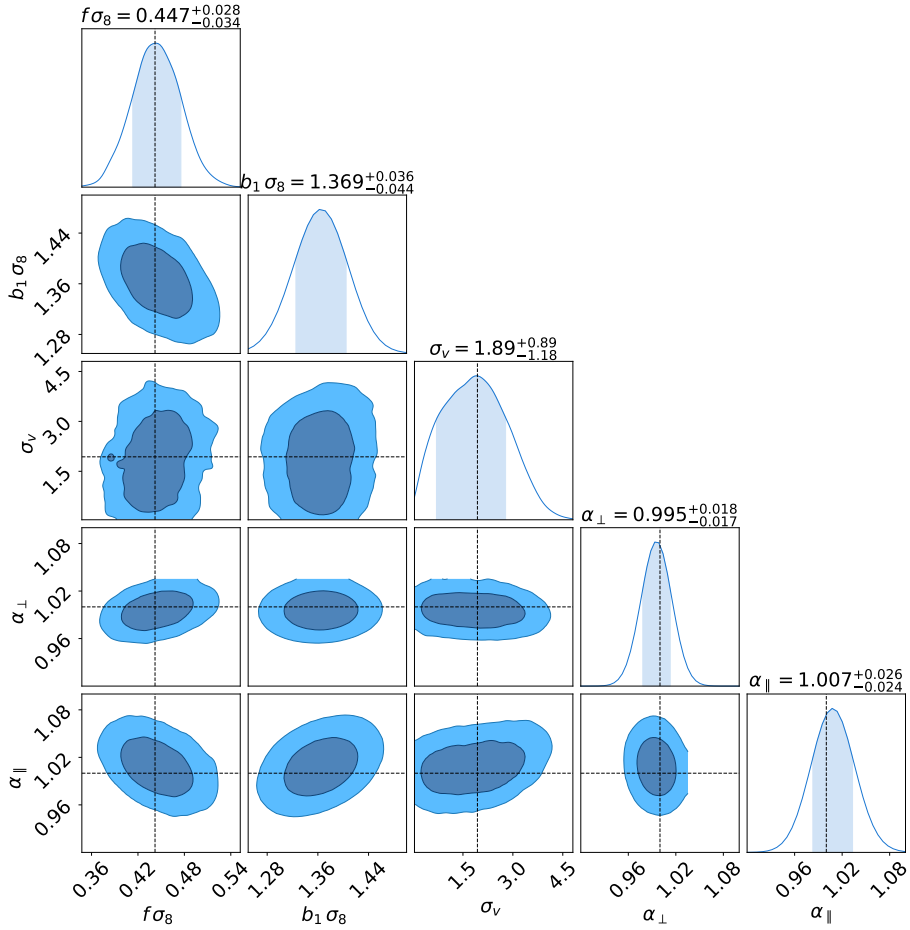

Figure 6.5: 1D and 2D posterior distributions for the free parameters $f\sigma_8$, $b_1\sigma_8$, σ_v, $\alpha_\perp$ and $\alpha_\parallel$ of the eTNS model, obtained from the analysis of the MINERVA 2PCF multipoles. The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed lines represent the theoretical predictions for the parameters.

the model accurately approximates the data in the range $30\,h^{-1}$Mpc $< s <$ $200\,h^{-1}$Mpc, which is the range of the fit. In Fig. 6.5 we display the 1D and 2D posterior distributions for the free parameters of the fit $f\sigma_8$, $b_1\sigma_8$ σ_v, $\alpha_\perp$ and $\alpha_\parallel$. The parameter $b_2\sigma_8$ is not shown since it is not constrained by this analysis. The dashed lines in Fig. 6.5 represent the expected theoretical values for the parameters. Specifically, the values of f and σ_8 have been computed from the cosmological parameters of the MINERVA simulations (see Tab. 6.1) and consistently rescaled to the considered redshift. Moreover, the reference value for σ_v is estimated using equation (3.7), and the reference value for both $\alpha_\perp$

and $\alpha_\parallel$ is 1, due to the fact that we know the true cosmology of the simulations we are analyzing.

6.6 The analysis of 3PCF in the MINERVA simulations

In the second analysis, we performed the fit of the 3PCF of the MINERVA mock catalogues using our new 3PCF model (see section 4.2), whose parameters are β, b_1, b_2 and b_t, where $\beta \equiv f/b_1$ with f the linear growth rate, b_1 is the linear bias, b_2 is the nonlinear bias, and b_t is the tidal bias.

First of all, the MINERVA 3PCF data set was given in the form of the coefficients of the Legendre expansion of the 3PCF, namely the ζ_ℓ in the following equation:

$$\zeta(\mathbf{r_{12}}, \mathbf{r_{13}}, \theta) = \sum_{\ell=0}^{\ell_{\max}} \zeta_\ell(\mathbf{r_{12}}, \mathbf{r_{13}})\, P_\ell(\cos\theta), \tag{6.14}$$

where ζ is the 3PCF, parametrized as a function of the first side of the triangles, $\mathbf{r_{12}}$, the second side of the triangles, $\mathbf{r_{13}}$, and the angle between them, θ. P_ℓ are the Legendre polynomials. Therefore, we proceeded to reconstruct the 3PCF, ζ, by multiplying the coefficients ζ_ℓ by the Legendre polynomials, up to $\ell_{\max} = 10$, which is the value of the provided data set. The result model of the MINERVA 3PCF is showed in Fig.6.6 for the illustrative scales of the $\{r_{12} = 25\,h^{-1}\text{Mpc}, r_{13} = 105\,h^{-1}\text{Mpc}\}$ and $\{r_{12} = 45\,h^{-1}\text{Mpc}, r_{13} = 115\,h^{-1}\text{Mpc}\}$, which were chosen since they include the BAO feature, indicated, in each panel, by the red arrow.

After that, we performed a fit of the real-space MINERVA 3PCF at fixed scales, in order to compare our results with the one obtained by Binetti, 2020 in configuration space and Oddo et al., 2020 in Fourier space (both these analyses have been carried out on the same real-space data from the MINERVA simulations). We sampled the posterior distribution using an MCMC method with 64 walkers and the chain size parameter set to 3000. In particular, the β parameter is fixed to zero, since we are dealing with real-space data, while the other parameters are free to vary, their prior distributions being represented in Tab. 6.3. In Fig. 6.7 we show the data and the best-fit model for the illustrative scales $r_{12} = 25\,h^{-1}$Mpc and $r_{13} = 45\,h^{-1}$Mpc, which are the same scales considered by Binetti, 2020. Fig. 6.8 displays the 1D and 2D posterior distributions for

b_1	b_2	b_t
$\mathcal{U}(0.5, 5)$	$\mathcal{U}(-10, 10)$	$\mathcal{U}(-5, 5)$

Table 6.3: Prior distributions for the free parameters used in the analysis of the real-space MINERVA 3PCF at fixed scales, where $\mathcal{U}(a, b)$ indicates a uniform distribution between a and b.

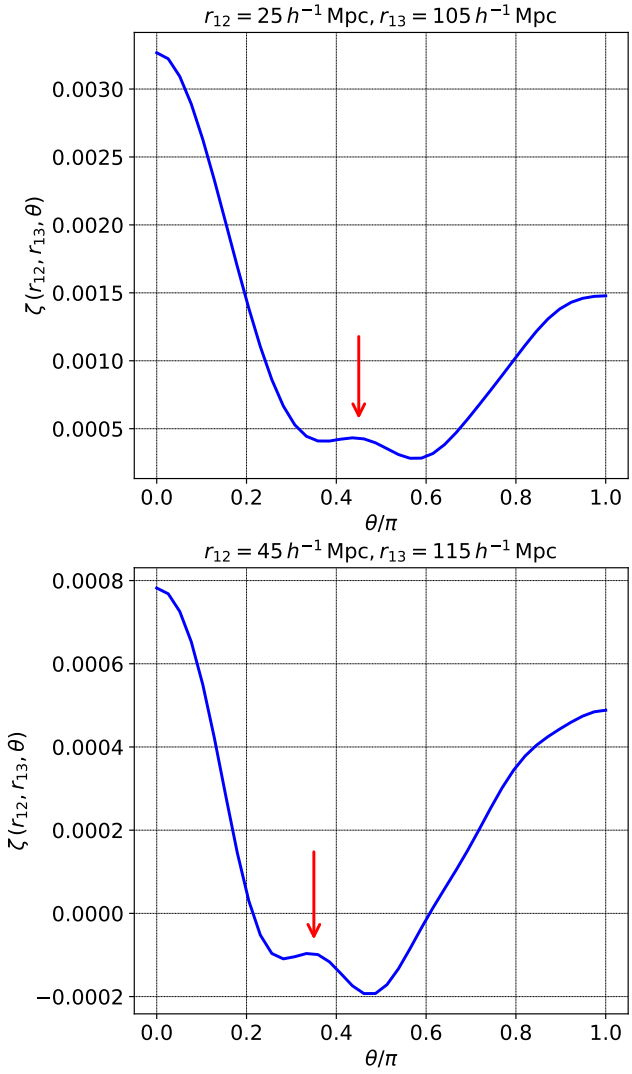

Figure 6.6: The model 3PCF, ζ, as a function of the angle θ between the two sides $\mathbf{r_{12}}$ and $\mathbf{r_{13}}$ of the triangles for two different configurations, indicated in the title of each box. In each panel, the red arrow displays the position of the BAO peak.

the free parameters of the fit. The results show a perfect agreement with the results of both the configuration space analysis of Binetti, 2020 and the Fourier space analysis of Oddo et al., 2020.

After that, we performed a fit of the redshift-space MINERVA 3PCF at all scales. Again, we set the number of walkers in the MCMC method to 64 and the

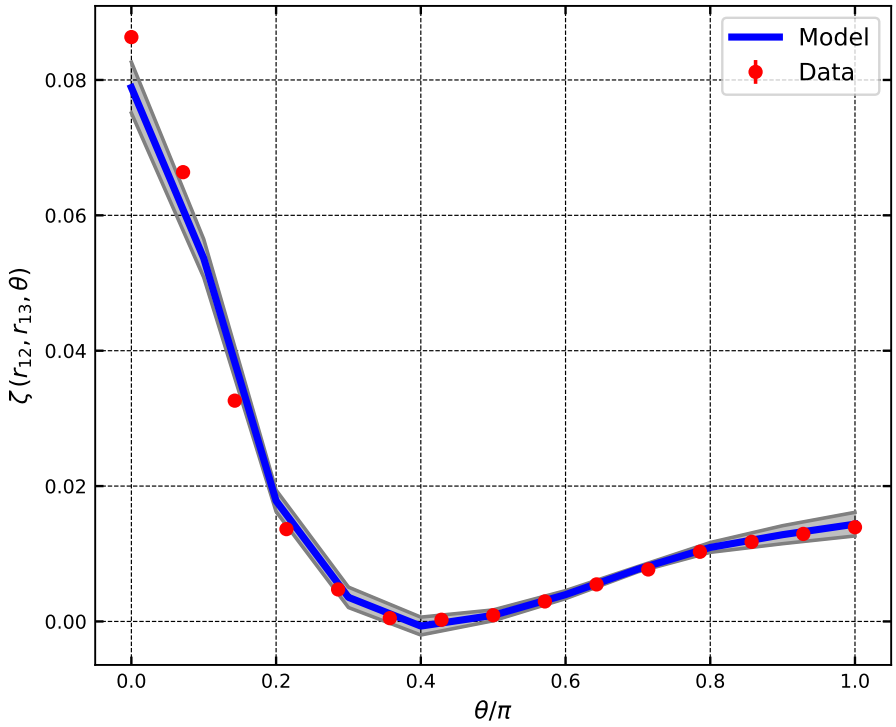

Figure 6.7: Real-space 3PCF of the MINERVA mock catalogues, for the fixed scales $r_{12} = 25\,h^{-1}$Mpc and $r_{13} = 45\,h^{-1}$Mpc. The red circles represent the data, whose error bars are the square root of the diagonal elements of the covariance matrix divided by the square root of the number of mock catalogues, and they are not visible since they are very small. The blue curve show the best-fit model, while the grey shaded areas display the 1-σ error on the model.

chain size parameter to 3000. In this case, β is considered as a free parameter of the model, as b_1, b_2 and b_t. We did not fit all the configurations, but we performed two cuts in the scales. The first one excludes the configurations which do not satisfy the triangle inequality. The second one excludes all the scales smaller than $r_{\rm min} = 60\,h^{-1}$Mpc, since the model is expected to be less accurate at small scales (see Veropalumbo et al., 2022), larger than $r_{\rm max} = 250\,h^{-1}$Mpc, and with an η parameter larger than $\eta_{\rm min}$. The η parameter is defined as:

$$\eta \equiv \frac{r_{13} - r_{12}}{\Delta r}, \tag{6.15}$$

where $\Delta r = 10\,h^{-1}$Mpc is the binning size. Then, if we set $\eta_{\rm min} = 0$, it corresponds to consider all the available triangles with sides within the range of scales $[r_{\rm min}, r_{\rm max}]$. If we set $\eta_{\rm min} = 1$, it corresponds to the exclusion of isosceles configurations, hence with $r_{12} = r_{13}$, and higher values of $\eta_{\rm min}$ would further

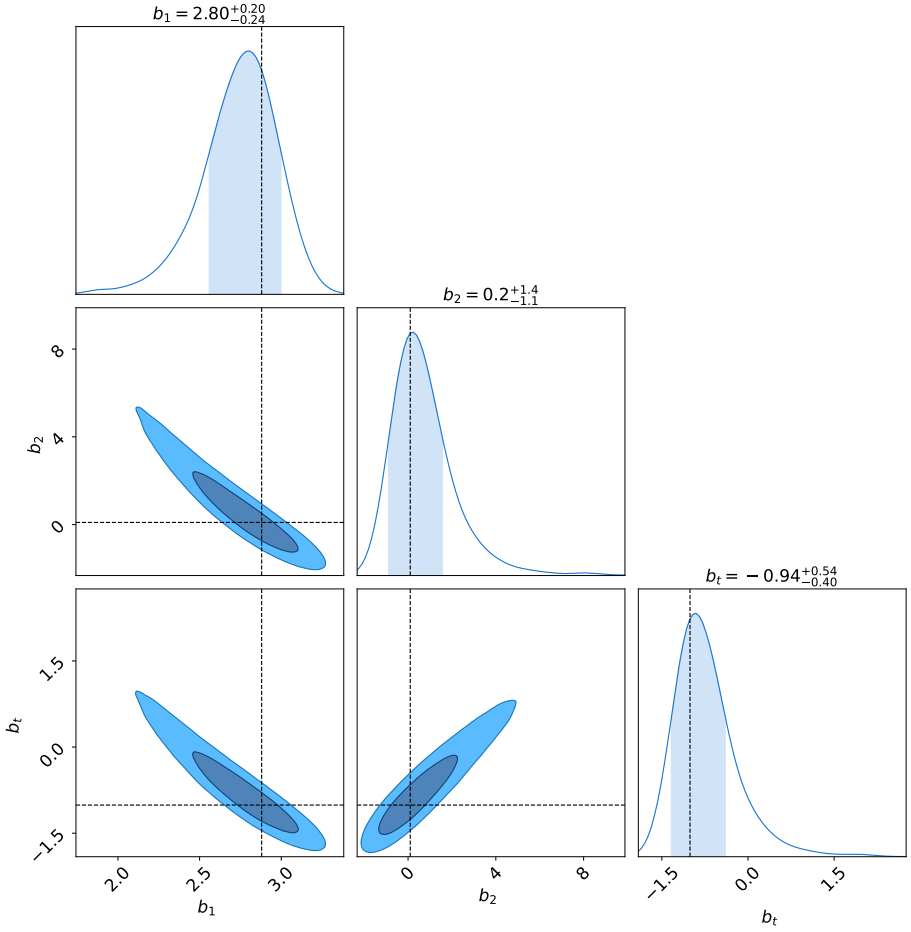

Figure 6.8: 1D and 2D posterior distributions for the free parameters b_1, b_2 and b_t of our new 3PCF model, obtained from the analysis of the real-space MINERVA 3PCF for the fixed scales $r_{12} = 25\,h^{-1}$Mpc and $r_{13} = 45\,h^{-1}$Mpc. The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively. The dashed lines represent the results of the analysis of Binetti, 2020 at the same scales.

reduce the triangle set. Since the estimator based on the SHD method (see section 2.2) cannot properly recover the 3PCF corresponding to isosceles configurations (Slepian and Eisenstein, 2015b; Veropalumbo et al., 2022), we set $\eta_{\rm min} = 4$. Our choice of $r_{\rm min}$ and $r_{\rm max}$ was made with the aim of excluding the scales with a reduced chi-square significantly larger than 1. Fig. 6.9 displays the data along with the best-fit model, showing that the model is able to reproduce the expected behaviour in the data. The 1D and 2D posterior distributions for the free parameters of the fit are represented in Fig. 6.10. The dashed lines indi-

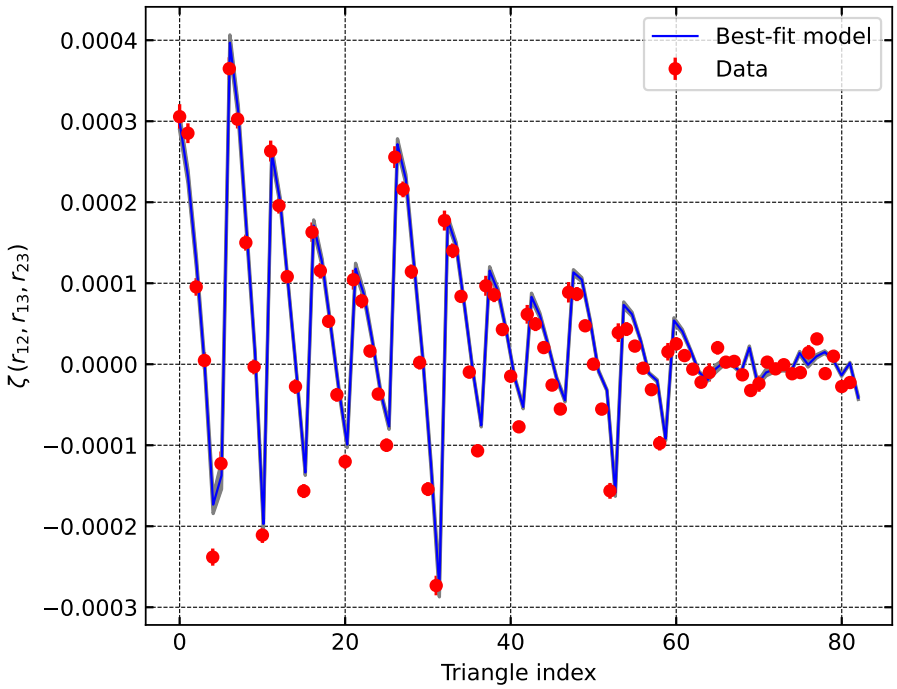

Figure 6.9: Redshift-space MINERVA 3PCF at all scales. The red circles represent the data, whose error bars are the square root of the diagonal elements of the covariance matrix divided by the square root of the number of mock catalogues, and they are not visible since they are very small. The blue curve show the best-fit model, while the grey shaded areas, which are really small, display the 1-σ error on the model.

cate the results obtained by Veropalumbo et al., 2022 with a different cut in the scales, since they used $r_{\rm min} = 40\,h^{-1}$Mpc, $r_{\rm max} = 130\,h^{-1}$Mpc and $\eta = 3$. The differences in the estimates of the parameters can be attributed to the different scales considered, and to the fact that Veropalumbo et al., 2022 took advantage of a binned model, computing the multipoles as:

$$\tilde{\zeta}_\ell(r_{12}, r_{13}) = \frac{2\ell+1}{2} \int_{-1}^{1} \mathrm{d}\mu P_\ell(\mu) \int_{r_{12}-\Delta r/2}^{r_{12}+\Delta r/2} \mathrm{d}q q^2 \int_{r_{13}-\Delta r/2}^{r_{13}+\Delta r/2} \mathrm{d}p p^2 \zeta(p, q, \mu), \tag{6.16}$$

where P_ℓ are the Legendre polynomials, μ is the cosine of the angle between the two triangle sides r_{12} and r_{13}, and p and q are mute variables. This integral is significantly demanding in terms of computing resources, particularly if it is evaluated at each MCMC step. For this reason, we did not include it in our analysis. Further investigations are needed to understand whether the differences in the estimated parameters are fully due to this aspect.

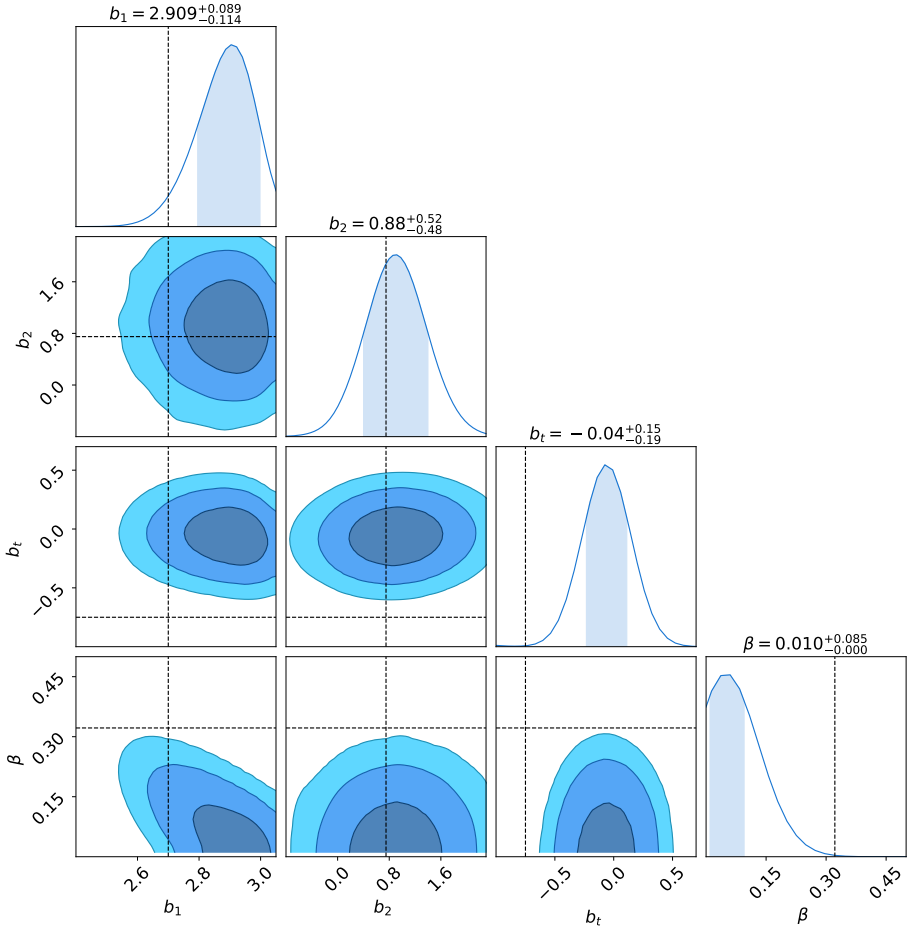

Figure 6.10: 1D and 2D posterior distributions for the free parameters b_1, b_2, b_t and β of our new 3PCF model, obtained from the analysis of the real-space MINERVA 3PCF at all scales. The darker/lighter shades in each box indicate the 68%, 95% and 99.7% confidence regions, respectively. The dashed lines represent the results of the analysis of Veropalumbo et al., 2022, which used a different cut in the scales.

6.7 Towards a joint analysis of 2PCF and 3PCF

After having modified the CBL functions for the 2PCF modelling letting b_t free to vary, having derived and implemented a new model for the redshift-space 3PCF, and having estimated the cross-covariance terms of both 2PCF and 3PCF, we are ready to prepare the computational framework for the joint analysis of 2PCF and 3PCF.

We consider the two probes as dependent, and we use a method to combine

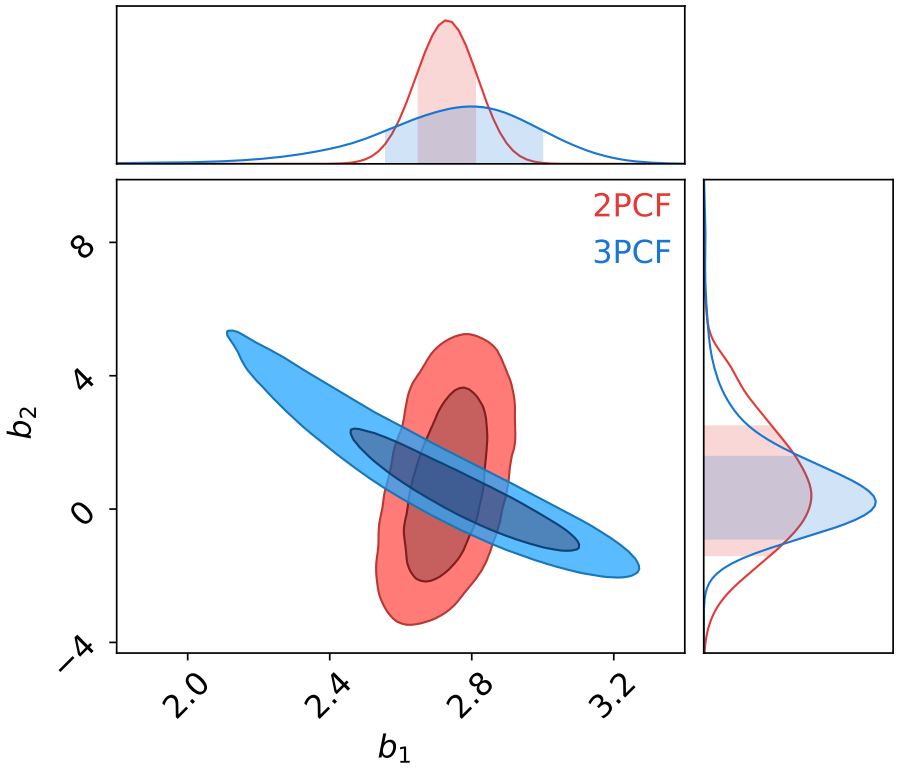

Figure 6.11: 1D and 2D posterior distributions for the parameters b_1 and b_2 obtained from the analyses of the 2PCF (red) and 3PCF (blue). The darker/lighter shades in each box indicate the 68% and 95% confidence regions, respectively.

the posterior distributions which consists of running the MCMC chain and, at each chain step θ_i, evaluating the logarithm of the posterior $\log\left[P_{\rm tot}(\theta_i)\right]$ as the sum of the logarithms of the posteriors of the considered probes $\log\left[P_j(\theta_i)\right]$. Considering N probes, we have:

$$\log\left[P_{\rm tot}(\theta_i)\right] = \sum_{j=1}^{N} \log\left[P_j(\theta_i)\right] . \tag{6.17}$$

In terms of likelihood $\mathcal{L}$ and prior π the latter equation can be written as:

$$\log\left[P_{\rm tot}(\theta_i)\right] = \sum_{j=1}^{N} \left\{\log\left[\mathcal{L}_j(\theta_i)\right] + \log\left[\pi_j(\theta_i)\right]\right\} . \tag{6.18}$$

This method is implemented in the CBL[1]. In this way, we are able to estimate

[1]It is implemented in the CBL class `cbl::statistics::CombinedPosterior`

the parameters of the two models exploiting information coming from both the 2PCF and 3PCF in the fit.

In Fig. 6.11 we show the comparison of the 1D and 2D posterior distributions of the parameters b_1 and b_2 obtained with the 2PCF analysis described in section 6.5 and the 3PCF analysis, in real space, discussed in section 6.6. This plot shows the interdependence and synergy between the 2PCF and the 3PCF. The combination of these two probes has the potential of increasing the accuracy of the constraints in cosmological parameters estimation and breaking degeneracies between them. Due to CPU limitations, it was not possible to include the results in this Thesis, but a full analysis is currently ongoing.

Chapter 7

Conclusions and future prospects

Scientific context

The extremely significant advancements in the study of the properties of the LSS of the Universe that we have witnessed in the last decades have led to the definition of the concordance ΛCDM model. As presented in Chapter 1, this model is based on Einstein's theory of GR, assuming the validity of the Cosmological Principle. The assumption of homogeneity and isotropy of space at large scales (typically hundreds of megaparsec) leads to the derivation of the FLRW metric, which describes the distances in the Universe, and to the formulation of the Friedmann equations, which describe the gravitational evolution of the cosmic fluid. The ΛCDM model predicts that the majority of the mass of the Universe is in the form of cold dark matter, and that the expansion of the Universe is accelerating due to the presence of dark energy. The existence of these two dark components, which completely dominate the energy content of the Universe, despite being generally accepted by the scientific community, opens several new challenges. In fact, the nature of CDM and dark energy represents one of the main open questions of modern cosmology.

The study of the clustering properties of the LSS represents an effective statistical tool to constrain cosmology and to widen our knowledge regarding the Universe. Specifically, the 2PCF and 3PCF (see sections 2.1 and 2.2 respectively), along with their Fourier counterparts, the power spectrum and the bispectrum, constitute extremely powerful cosmological probes. They encode information about the RSD, the distortions in the density field induced by the matter peculiar velocity field, and the BAO peak, a well defined distance imprinted in the matter distribution in the primordial Universe. The analysis of these clustering distortions represents one of the fundamental probes to obtain estimates of the cosmological parameters (see section 2.3).

Forthcoming redshift surveys, such as Euclid (Laureijs et al., 2006), the Vera

C. Rubin Observatory LSST (Ivezić et al., 2019), and the Nancy Grace Roman Space Telescope (Spergel et al., 2015), aim to push the accuracy of cosmological measurements to unprecedented levels. Within this context, it is necessary to go beyond two-point statistics, which is no longer sufficient to extract all the available cosmological information, and analyze higher-order clustering statistics. In particular, three-point statistics plays a crucial role, encoding information about deviations from Gaussianity in the primordial density field. Consequently, developing new and effective theoretical models, and exploring their application, becomes of utmost importance. Moreover, new and efficient statistical tools need to be implemented, in order to maximize the retrieved cosmological information.

This Thesis project aims to address these challenges, by deriving a new 3PCF redshift-space model, and developing the necessary computational framework for a joint analysis of the 2PCF and 3PCF.

Main results

We can summarize the main results of this Thesis work as follows:

- We proposed an extension of the redshift-space bispectrum model of Scoccimarro et al., 1999 (SCF99), by including the tidal bias b_t as an additional parameter, as shown in section 4.1. The SCF99 model for the bispectrum considered the following galaxy bias relation:

$$\delta_g(x) = b_1\delta(x) + \frac{b_2}{2}\delta^2(x), \tag{7.1}$$

  where δ_g is the galaxy density field, δ is the matter density field, x is a point in space, b_1 is the linear bias, and b_2 is the nonlinear bias. We included an additional bias parameter in our galaxy bias relation as follows:

$$\delta_g(x) = b_1\delta(x) + \frac{b_2}{2}\delta^2(x) + b_t s^2(x), \tag{7.2}$$

  where b_t is the tidal bias, and $s^2(x)$ is the square of the tidal field, which is proportional to the second derivative of the gravitational potential $\Phi(x)$. For this reason, we include an additional term, with respect to SCF99, in the definition of the second-order PT redshift-space kernel $Z_2(\mathbf{k}_1, \mathbf{k}_2)$. This term, $b_t S_2(\mathbf{k}_1, \mathbf{k}_2)$, accounts for the tidal field contribution through its kernel S_2. From this, we derived the resulting bispectrum model, which is presented in equation (4.12), along with equations (4.13), (4.14), (4.15), and (4.16).

- Using the previously derived expression for the bispectrum, we obtained a new formulation for the redshift-space model of the 3PCF, that improves with respect to the model of Slepian and Eisenstein, 2017 (SE17) by revising some incorrect terms in their calculations. We theoretically derived

the new 3PCF model from scratch, as described in section 4.2, using the Python library for symbolic mathematics SymPy, through which we solved all the integrals involved in the calculation. We began the derivation by expanding the 3PCF, ζ, onto Legendre polynomials, P_ℓ, as equation (4.19) shows. The coefficients of such expansion can be recovered through an integral of the so-called pre-cyclic 3PCF (see equation 4.20) that is the 3PCF prior to the summation over cyclic permutations of the triangle sides r_{12}, r_{13}, r_{23}, needed to ensure the indistinguishability of the triangle vertices. Then, we expanded the pre-cyclic 3PCF, again, onto Legendre polynomials (see equation 4.22). The coefficients of this expansion, $\zeta_{\mathrm{pc},\ell}$, can be recovered through the expression in equation (4.23), which depends on the multipole coefficients of the bispectrum, $B_{s,\ell}$. In order to derive the $\zeta_{\mathrm{pc},\ell}$, we proceeded to expand the different terms of the bispectrum monopole B_s derived in section 4.1 onto Legendre polynomials. The final result is the line-of-sight averaged 3PCF model, which we report here since it

represents one of the major achievements of this Thesis work:

$$
\begin{aligned}
\ell = 0: \quad & \xi_1^{[0]}\xi_2^{[0]}\Bigg\{ b_1^4\left[\frac{2}{3}\beta + \frac{38}{45}\beta^2 + \frac{2}{5}\beta^3 + \frac{2}{25}\beta^4\right] \\
& + b_1^3\Bigg[\frac{34}{21}\left(1 + \frac{47}{51}\beta + \frac{163}{425}\beta^2 + \frac{201}{2975}\beta^3\right) \\
& + \gamma\left(1 + \frac{2}{3}\beta + \frac{1}{9}\beta^2\right) - \frac{4}{3}\gamma_t\left(1 + \frac{2}{3}\beta + \frac{7}{75}\beta^2\right)\Bigg]\Bigg\} \\
& + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_0(r_1, r_2) \\
\ell = 1: \quad & - \left[\xi_1^{[1+]}\xi_2^{[1-]} + \xi_2^{[1+]}\xi_1^{[1-]}\right]\Bigg\{ b_1^4\left[\frac{1}{3}\beta + \frac{3}{5}\beta^2 + \frac{67}{175}\beta^3 + \frac{3}{35}\beta^4\right] \\
& + b_1^3\left[1 + \beta + \frac{37}{75}\beta^2 + \frac{17}{175}\beta^3\right]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_1(r_1, r_2) \\
\ell = 2: \quad & \xi_1^{[2]}\xi_2^{[2]}\Bigg\{ b_1^4\left[\frac{16}{45}\beta^2 + \frac{16}{35}\beta^3 + \frac{32}{245}\beta^4\right] \\
& + b_1^3\Bigg[\frac{8}{21}\left(1 + \frac{4}{3}\beta + \frac{54}{35}\beta^2 + \frac{111}{245}\beta^3\right) \\
& + \frac{4\beta^2\gamma}{45} + \frac{4}{3}\gamma_t\left(1 + \frac{2}{3}\beta + \frac{1}{21}\beta^2\right)\Bigg]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_2(r_1, r_2) \\
\ell = 3: \quad & - \left[\xi_1^{[3+]}\xi_2^{[3-]} + \xi_2^{[3+]}\xi_1^{[3-]}\right]\Bigg\{ b_1^4\left[\frac{8}{175}\beta^3 + \frac{8}{315}\beta^4\right] \\
& + b_1^3\left[\frac{8}{75}\beta^2 + \frac{8}{175}\beta^3\right]\Bigg\} + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_3(r_1, r_2) \\
\ell = 4: \quad & \xi_1^{[4]}\xi_2^{[4]}\Bigg\{ \frac{128}{11025}\beta^4 b_1^4 + b_1^3\left[-\frac{32}{3675}\beta^2 + \frac{32}{8575}\beta^3 + \frac{32}{525}\beta^2\gamma_t\right]\Bigg\} \\
& + b_1^3\beta^2\left(7 + 3\beta\right)\kappa_4(r_1, r_2) \\
\ell \geq 5: \quad & b_1^3\beta^2\left(7 + 3\beta\right)\kappa_\ell(r_1, r_2).
\end{aligned}
$$

These are the multipole moments of the 3PCF, where $\xi_i^{[n]}$ and $\xi_i^{[n\pm]}$ are integrals of the power spectrum defined in equation (4.37), the κ_ℓ terms are defined in equation (4.39), and we used $\gamma \equiv b_2/b_1$, $\gamma_t \equiv b_t/b_1$, and $\beta \equiv f/b_1$, f being the linear growth rate.

- We validated the new 3PCF model by performing various tests, comparing its multipole moments to the ones of the 3PCF model of Veropalumbo et al., 2022 (see section 5.4.1). For each test, we used the model parameters showed in Tab. 5.1. As Figs. 5.2-5.5 show, the two models are in very good

agreement, with an average difference on ζ_ℓ smaller than 8×10^{-7}, roughly a factor 50 smaller than the average values of the multipole moments. The small discrepancies are accountable for the different method of integration adopted in the two models, since in this analysis we used the quadrature method, whereas Veropalumbo et al., 2022 used an FFT algorithm.

- We implemented the new codes used for our analysis in the *CosmoBolognaLib* (CBL) libraries. We can categorize the modifications in the CBL as follows:
  - we added two new functions for the estimate of the covariance matrix of a collection of data sets, aiming at making the use of this CBL class more user-friendly, mainly for Python users (see section 5.2);
  - we modified the model functions related to the 2PCF modelling in order to make the tidal bias b_t a free parameter. Indeed, since b_t is a free parameter in the 3PCF analysis, we need it to be free also in the 2PCF analysis. In this way, we can obtain constraints from the combination of both statistics (see section 5.3);
  - the only model in the CBL for the analysis of the redshift-space 3PCF was based on SE17, therefore it was not correct. For this reason, we implemented our new 3PCF model from scratch, organizing it into multiple functions, described in detail in Appendix C. Furthermore, the CBL offered the study of the 3PCF only with the parametrization depending on two triangle sides, r_{12} and r_{13}, and the angle between them, θ. Therefore, we introduced new functions which allow the expression of the 3PCF with a new parametrization, depending on the three sides of the triangles r_{12}, r_{13}, r_{23}. This enables to study the 3PCF at all scales, analyzing all the configurations, rather than just fixing two triangle sides, and vary the angle between them (see section 5.4).
- We performed two cosmological parameter estimation analyses: one with the MINERVA 2PCF (see section 6.5) and one with the MINERVA 3PCF (see section 6.6). In the first analysis, we used the eTNS model for the modelling of the MINERVA 2PCF multipoles, where we found a perfect agreement between the estimated parameters and the expected theoretical values. In the second analysis, we exploited the new 3PCF model for the analysis of the MINERVA 3PCF at all scales. The modelling of the 3PCF in real space is statistically consistent with the works of Binetti, 2020 in configuration space and Oddo et al., 2020 in Fourier space. The results in redshift space are promising, but still require some further optimization. As Fig. 6.11 shows, the probe combination of the 2PCF and 3PCF has a very strong potential both in in increasing the precision on cosmological parameter constraints, and in lifting parameter degeneracies.

All these results have the common purpose of developing a framework for the probe combination of the 2PCF and 3PCF. It was not possible to include the

results of such combination in this Thesis work, due to CPU constraints, however a full analysis is currently ongoing.

Future perspectives

The natural development of this Thesis project consists of exploiting the developed framework to perform a joint analysis of the 2PCF and 3PCF, which is currently in progress. Moreover, further investigations are needed to understand the nature of the differences in the estimates of the parameters from the 3PCF analysis with respect to the ones estimated by Veropalumbo et al., 2022. After that, another development of this Thesis project will consist of applying the derived models and pipelines on a sample of real data. The results of this work will be exploited for the next work of the "Cluster Clustering Cosmology (C^3)" project, whose first paper, Marulli et al., 2021, dealed with the analysis of the anisotropies of the 2PCF of a large spectroscopic sample of galaxy clusters extracted from the Sloan Digital Sky Survey (SDSS); while the second paper, Moresco et al., 2021, concerned the investigation of the 3PCF of these galaxy clusters up to the BAO scale. The next step of the C^3 project will consist of exploiting a joint analysis of the 2PCF and 3PCF of the SDSS galaxy clusters, with the goal of providing tighter constraints on cosmological parameters and breaking degeneracies between them.

Appendix A

Implementated methods: covariance matrix

For what concerns the covariance matrix computation, we added two new functions to the CBL class `cbl::data::CovarianceMatrix`. In particular, we added two overloaded versions of the method `CovarianceMatrix::measure()`[1], used to compute the covariance matrix of a given dataset. The two existing overloaded versions of this method took a vector of smart pointers (in particular `std::shared_ptr`[2] objects) to the CBL class `cbl::data::Data` or a vector of vectors of smart pointers to the CBL class `cbl::data::Data` as arguments, respectively. As a result, the user should have turned the dataset into a pointer to the `cbl::data::Data` class before measuring the covariance matrix. This can be problematic, mainly for Python users, since the handling of smart pointers in Python is not straightforward. Consequently, with the purpose of making this class more user-friendly, we added two overloaded versions of the method `CovarianceMatrix::measure()` which take a vector of vectors of doubles (a 2-dimensional vector in which each element is a vector of doubles) and a vector of vectors of vectors of doubles (a 3-dimensional vector in which each element is a 2-dimensional vector) as arguments, respectively.

[1] We dropped the `cbl::data::` prefix for clarity.
[2] For more information visit https://en.cppreference.com/w/cpp/memory/shared_ptr.

Appendix B

Implementated methods: two-point correlation function

We modified some existing functions of the CBL class `cbl::modelling::twopt::Modelling_TwoPointCorrelation_multipoles`, to make the tidal bias parameter b_t a free parameter of the fit. In particular, we modified the following functions[1]:

- `Modelling_TwoPointCorrelation_multipoles::set_model_eTNS()`;
- `Modelling_TwoPointCorrelation_multipoles::set_fiducial_PkDM()`;
- `xiMultipoles()`;
- `Pkmu()`;
- `Pkmu_eTNS()`.

[1] We dropped the `cbl::modelling::twopt::` prefix for clarity.

Appendix C

Implementated methods: three-point correlation function

We implemented our new model for the 3PCF (see section 4.2) from scratch and added it to the CBL class `cbl::cosmology::ThreePointCorrelation`. Since it is organized into multiple functions, we present here a detailed description of the implementation. In particular, let us start from the following methods[1]:

- `zeta_ell_0_factor()`;
- `zeta_ell_0_factor_tidal()`;
- `zeta_ell_1_factor()`;
- `zeta_ell_2_factor()`;
- `zeta_ell_2_factor_tidal()`;
- `zeta_ell_3_factor()`;
- `zeta_ell_4_factor()`;
- `zeta_ell_4_factor_tidal()`;

that provide the content of the curly brackets of equation (4.42). Indeed, as the names suggest, these functions compute the terms corresponding to the $\ell = 0, 1, 2, 3, 4$ orders of the Legendre expansion, the contribution from the tidal bias being included in the ones labelled as `tidal`. Then, implemented the following functions:

[1] We drop the `cbl::cosmology::ThreePointCorrelation::` prefix for clarity.

- `zeta_precyclic_RSD()`: this function computes the precyclic 3PCF $\zeta_{\rm pc}$ by multiplying the multiple moments expressed in equation (4.42) times the Legendre polynomials P_ℓ, obtaining the result of equation (4.22). Then it sums together the different precyclic 3PCFs in order to get the content of the square brackets of equation (4.20);

- `zeta_expansion_RSD()`: this method integrates the product of the result of `zeta_precyclic_RSD()` times the Legendre polynomials to get the coefficients of the Legendre expansion of the 3PCF, ζ_ℓ, which is the result of equation (4.20);

- `zeta_RSD()`: this function calculates the 3PCF ζ at fixed scales by expanding the coefficients ζ_ℓ computed by `zeta_expansion_RSD()` onto Legendre polynomials, as shown by equation (4.19). There are two overloaded versions of this function, the first one takes a single angle as an argument and returns a single ζ, while the second one takes a vector of angles as an argument and returns a vector of ζ;

- `zeta_RSD_all_scales()`: this method computes the 3PCF ζ at all scales by cycling over all the possible scales and calling `zeta_RSD()` at each cycle step.

After having described the implementation of the model in the class `cbl::cosmology::ThreePointCorrelation`, we can turn to the modifications to the methods belonging to the `cbl::modelling::threept` namespace. Specifically, the following functions were added[2]:

- `Modelling_ThreePointCorrelation_comoving_connected::set_model_RSD()`: this function defines the parameters to model the connected 3PCF in comoving coordinates at fixed scales;

- `Modelling_ThreePointCorrelation_comoving_connected::set_model_RSD_all_scales()`: this function defines the parameters to model the connected 3PCF in comoving coordinates at all scales;

- `zeta_RSD()`: this method selects `cbl::cosmology::ThreePointCorrelation::zeta_RSD()` as the function that must be used for the modelling of the data;

- `zeta_RSD_all_scales()`: this method selects `cbl::cosmology::ThreePointCorrelation::zeta_RSD_all_scales()` as the function that must be used for the modelling of the data.

[2]We drop the `cbl::modelling::threept::` prefix for clarity.

Bibliography

Abdalla, E., Abellán, G. F., Aboubrahim, A., Agnello, A., Akarsu, Ö., Akrami, Y., Alestas, G., Aloni, D., Amendola, L., Anchordoqui, L. A., Anderson, R. I., Arendse, N., Asgari, M., Ballardini, M., Barger, V., Basilakos, S., Batista, R. C., Battistelli, E. S., Battye, R., ... Zumalacárregui, M. (2022). *J. High Energy Phys.*, *34*, 49–211. https://doi.org/10.1016/j.jheap.2022.04.002

Alam, S., Aubert, M., Avila, S., Balland, C., Bautista, J. E., Bershady, M. A., Bizyaev, D., Blanton, M. R., Bolton, A. S., Bovy, J., Brinkmann, J., Brownstein, J. R., Burtin, E., Chabanier, S., Chapman, M. J., Choi, P. D., Chuang, C.-H., Comparat, J., Cousinou, M.-C., ... Zheng, Z. (2021). *Phys. Rev. D*, *103*, 083533. https://doi.org/10.1103/PhysRevD.103.083533

Alcock, C., & Paczynski, B. (1979). *Nature*, *281*, 358. https://doi.org/10.1038/281358a0

Baldauf, T., Seljak, U. š., Desjacques, V., & McDonald, P. (2012). *Phys. Rev. D*, *86*, 083540. https://doi.org/10.1103/PhysRevD.86.083540

Barlow, R. J. (1991). *Statistics: A guide to the use of statistical methods in the physical sciences*. Wiley.

Barriga, J., & Gaztañaga, E. (2002). *Mon. Not. R. Astron. Soc.*, *333*, 443–453. https://doi.org/10.1046/j.1365-8711.2002.05431.x

Bassett, B. A., & Hlozek, R. (2009).

Bautista, J. E., Paviot, R., Vargas Magaña, M., de la Torre, S., Fromenteau, S., Gil-Marín, H., Ross, A. J., Burtin, E., Dawson, K. S., Hou, J., Kneib, J.-P., de Mattia, A., Percival, W. J., Rossi, G., Tojeiro, R., Zhao, C., Zhao, G.-B., Alam, S., Brownstein, J., ... Wang, Y. (2020). *Mon. Not. R. Astron. Soc.*, *500*, 736–762. https://doi.org/10.1093/mnras/staa2800

Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. (2002). *Physics Reports*, *367*, 1–248. https://doi.org/10.1016/S0370-1573(02)00135-7

Beutler, F., Saito, S., Seo, H.-J., Brinkmann, J., Dawson, K. S., Eisenstein, D. J., Font-Ribera, A., Ho, S., McBride, C. K., Montesano, F., Percival, W. J., Ross, A. J., Ross, N. P., Samushia, L., Schlegel, D. J., Sánchez, A. G., Tinker, J. L., & Weaver, B. A. (2014). *Mon. Not. R. Astron. Soc.*, *443*, 1065–1089. https://doi.org/10.1093/mnras/stu1051

Binetti, A. (2020). *A robust characterization of the statistical properties of higher-order correlation functions for next generations galaxy surveys* (Master's thesis). University of Milano.

Catelan, P., Lucchin, F., Matarrese, S., & Porciani, C. (1998). *Mon. Not. R. Astron. Soc.*, *297*, 692–712. https://doi.org/10.1046/j.1365-8711.1998.01455.x

Cen, R., & Ostriker, J. P. (2000). *Astrophys. J.*, *538*, 83–91. https://doi.org/10.1086/309090

Chan, K. C., Scoccimarro, R., & Sheth, R. K. (2012). *Phys. Rev. D*, *85*. https://doi.org/10.1103/physrevd.85.083509

Desjacques, V., & Seljak, U. (2010). *Class. Quantum Grav.*, *27*, 124011. https://doi.org/10.1088/0264-9381/27/12/124011

Di Benedetto, A. (2023). *Testing dynamic distortion models to infer the linear growth rate from clustering multipoles* (Master's thesis) [Available at https://amslaurea.unibo.it/28039/]. University of Bologna.

Dodelson, S. (2003). *Modern Cosmology*. Academic Press, Elsevier Science.

Einstein, A. (1915). *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften*, 778.

Eisenstein, D. J., Zehavi, I., Hogg, D. W., Scoccimarro, R., Blanton, M. R., Nichol, R. C., Scranton, R., Seo, H.-J., Tegmark, M., Zheng, Z., Anderson, S. F., Annis, J., Bahcall, N., Brinkmann, J., Burles, S., Castander, F. J., Connolly, A., Csabai, I., Doi, M., … York, D. G. (2005). *Astrophys. J.*, *633*, 560. https://doi.org/10.1086/466512

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). *Publ. Astron. Soc. Pac.*, *125*, 306. https://doi.org/10.1086/670067

Fry, J. N., & Gaztañaga, E. (1993). *Astrophys. J.*, *413*, 447. https://doi.org/10.1086/173015

Garcıéa-Farieta, J. E., Marulli, F., Moscardini, L., Veropalumbo, A., & Casas-Miranda, R. A. (2020). *Mon. Not. R. Astron. Soc.*, *494*, 1658–1674. https://doi.org/10.1093/mnras/staa791

Gardner, J. P., Connolly, A., & McBride, C. (2007). In R. A. Shaw, F. Hill, & D. J. Bell (Eds.), *Astronomical data analysis software and systems xvi* (p. 69). https://ui.adsabs.harvard.edu/abs/2007ASPC..376...69G

Gil-Marín, H., Bautista, J. E., Paviot, R., Vargas-Magaña, M., de la Torre, S., Fromenteau, S., Alam, S., Ávila, S., Burtin, E., Chuang, C.-H., Dawson, K. S., Hou, J., de Mattia, A., Mohammad, F. G., Müller, E.-M., Nadathur, S., Neveux, R., Percival, W. J., Raichoor, A., … Vivek, M. (2020). *Mon. Not. R. Astron. Soc.*, *498*, 2492–2531. https://doi.org/10.1093/mnras/staa2455

Gil-Marín, H., Noreña, J., Verde, L., Percival, W. J., Wagner, C., Manera, M., & Schneider, D. P. (2015). *Mon. Not. R. Astron. Soc.*, *451*, 539–580. https://doi.org/10.1093/mnras/stv961

Goodman, J., & Weare, J. (2010). *Comm. App. Math. Comp. Sci.*, *5*, 65–80. https://doi.org/10.2140/camcos.2010.5.65

Gray, A. G., Moore, A. W., Nichol, R. C., Connolly, A. J., Genovese, C., & Wasserman, L. (2004). In F. Ochsenbein, M. G. Allen, & D. Egret

(Eds.), *Astronomical data analysis software and systems (adass) xiii* (p. 249). https://doi.org/10.48550/arXiv.astro-ph/0401121

Grieb, J. N., Sánchez, A. G., Salazar-Albornoz, S., & Dalla Vecchia, C. (2016). *Mon. Not. R. Astron. Soc.*, *457*, 1577–1592. https://doi.org/10.1093/mnras/stw065

Hamilton, A. J. S. (2000). *Mon. Not. R. Astron. Soc.*, *312*, 257–284. https://doi.org/10.1046/j.1365-8711.2000.03071.x

Hartlap, Simon, P., & Schneider, P. (2007). *Astron. Astrophys.*, *464*, 399–404. https://doi.org/10.1051/0004-6361:20066170

Heavens, A. F., Matarrese, S., & Verde, L. (1998). *Mon. Not. R. Astron. Soc.*, *301*, 797–808. https://doi.org/10.1046/j.1365-8711.1998.02052.x

Huterer, D., & Shafer, D. L. (2018). *Rep. Prog. Phys.*, *81*, 016901. https://doi.org/10.1088/1361-6633/aa997e

Icaza-Lizaola, M., Vargas-Magaña, M., Fromenteau, S., Alam, S., Camacho, B., Gil-Marín, H., Paviot, R., Ross, A., Schneider, D. P., Tinker, J., Wang, Y., Zhao, C., Prakash, A., Rossi, G., Zao, G.-B., Cruz-Gonzalez, I., & de la Macorra, A. (2019). *Mon. Not. R. Astron. Soc.*, *492*, 4189–4215. https://doi.org/10.1093/mnras/stz3602

Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., Angel, J. R. P., Angeli, G. Z., Ansari, R., Antilogus, P., Araujo, C., Armstrong, R., Arndt, K. T., Astier, P., Aubourg, É., … Zhan, H. (2019). *Astrophys. J.*, *873*, 111. https://doi.org/10.3847/1538-4357/ab042c

Kaiser, N. (1987). *Mon. Not. R. Astron. Soc.*, *227*, 1–21. https://doi.org/10.1093/mnras/227.1.1

Katz, N., Hernquist, L., & Weinberg, D. H. (1999). *Astrophys. J.*, *523*, 463–479. https://doi.org/10.1086/307744

Labatie, A., Starck, J.-L., Lachièze-Rey, M., & Arnalte-Mur, P. (2012). *Statistical Methodology*, *9*, 85–100. https://doi.org/10.1016/j.stamet.2011.05.001

Landy, S. D., & Szalay, A. S. (1993). *Astrophys. J.*, *412*, 64. https://doi.org/10.1086/172900

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. .-., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., … Zucca, E. (2006).

Lewis, A., Challinor, A., & Lasenby, A. (2000). *Astrophys. J.*, *538*, 473. https://doi.org/10.1086/309179

Marulli, F., Veropalumbo, A., García-Farieta, J. E., Moresco, M., Moscardini, L., & Cimatti, A. (2021). *Astrophys. J.*, *920*, 13. https://doi.org/10.3847/1538-4357/ac0e8c

Marulli, F., Veropalumbo, A., & Moresco, M. (2016). *Astron. Comput.*, *14*, 35. https://doi.org/10.1016/j.ascom.2016.01.005

McDonald, P., & Roy, A. (2009). *J. Cosm. Astro-Particle Phys.*, *8*, 20. https://doi.org/10.1088/1475-7516/2009/08/020

Meurer, A., Smith, C. P., Paprocki, M., Čertıék, O., Kirpichev, S. B., Rocklin, M., Kumar, A., Ivanov, S., Moore, J. K., Singh, S., et al. (2017). *PeerJ Comput. Sci.*, *3*, e103. https://doi.org/10.7717/peerj-cs.103

Mohammad, Granett, B. R., Guzzo, L., Bel, J., Branchini, E., de la Torre, S., Moscardini, L., Peacock, J. A., Bolzonella, M., Garilli, B., Scodeggio, M., Abbas, U., Adami, C., Bottini, D., Cappi, A., Cucciati, O., Davidzon, I., Franzetti, P., Fritz, A., ... Moutard, T. (2018). *Astron. Astrophys.*, *610*, A59. https://doi.org/10.1051/0004-6361/201731685

Moresco, M., Veropalumbo, A., Marulli, F., Moscardini, L., & Cimatti, A. (2021). *Astrophys. J.*, *919*, 144. https://doi.org/10.3847/1538-4357/ac10c9

Oddo, A., Sefusatti, E., Porciani, C., Monaco, P., & Sánchez, A. G. (2020). *J. Cosmol. Astropart. Phys.*, *2020*, 056–056. https://doi.org/10.1088/1475-7516/2020/03/056

Percival, W. J., Ross, A. J., Sánchez, A. G., Samushia, L., Burden, A., Crittenden, R., Cuesta, A. J., Magana, M. V., Manera, M., Beutler, F., Chuang, C.-H., Eisenstein, D. J., Ho, S., McBride, C. K., Montesano, F., Padmanabhan, N., Reid, B., Saito, S., Schneider, D. P., ... Weaver, B. A. (2014). *Mon. Not. R. Astron. Soc.*, *439*, 2531–2541. https://doi.org/10.1093/mnras/stu112

Planck Collaboration, Aghanim, N., Akrami, Y., Ashdown, M., Aumont, J., Baccigalupi, C., Ballardini, M., Banday, A. J., Barreiro, R. B., Bartolo, N., Basak, S., Battye, R., Benabed, K., Bernard, J.-P., Bersanelli, M., Bielewicz, P., Bock, J. J., Bond, J. R., Borrill, J., ... Zonca, A. (2020). *Astron. Astrophys.*, *641*. https://doi.org/10.1051/0004-6361/201833910

Saito, S., Baldauf, T., Vlah, Z., Seljak, U., Okumura, T., & McDonald, P. (2014). *Phys. Rev. D*, *90*. https://doi.org/10.1103/physrevd.90.123522

Sánchez, A. G., Kazin, E. A., Beutler, F., Chuang, C.-H., Cuesta, A. J., Eisenstein, D. J., Manera, M., Montesano, F., Nichol, R. C., Padmanabhan, N., Percival, W., Prada, F., Ross, A. J., Schlegel, D. J., Tinker, J., Tojeiro, R., Weinberg, D. H., Xu, X., Brinkmann, J., ... Thomas, D. (2013). *Mon. Not. R. Astron. Soc.*, *433*, 1202. https://doi.org/10.1093/mnras/stt799

Scoccimarro, R. (2004). *Phys. Rev. D*, *70*. https://doi.org/10.1103/physrevd.70.083007

Scoccimarro, R., Couchman, H. M. P., & Frieman, J. A. (1999). *Astrophys. J.*, *517*, 531. https://doi.org/10.1086/307220

Sefusatti, E., Crocce, M., Pueblas, S., & Scoccimarro, R. (2006). *Phys. Rev. D*, *74*, 023522. https://doi.org/10.1103/PhysRevD.74.023522

Slepian, Z., & Eisenstein, D. J. (2015a). *Mon. Not. R. Astron. Soc.*, *448*, 9–26. https://doi.org/10.1093/mnras/stu2627

Slepian, Z., & Eisenstein, D. J. (2015b). *Mon. Not. R. Astron. Soc.*, *454*, 4142–4158. https://doi.org/10.1093/mnras/stv2119

Slepian, Z., & Eisenstein, D. J. (2017). *Mon. Not. R. Astron. Soc.*, *469*, 2059–2076. https://doi.org/10.1093/mnras/stx490

Smith, R. E., Sheth, R. K., & Scoccimarro, R. (2008). *Phys. Rev. D*, *78*. https://doi.org/10.1103/physrevd.78.023523

Spergel, D., Gehrels, N., Baltay, C., Bennett, D., Breckinridge, J., Donahue, M., Dressler, A., Gaudi, B. S., Greene, T., Guyon, O., Hirata, C., Kalirai, J., Kasdin, N. J., Macintosh, B., Moos, W., Perlmutter, S., Postman, M., Rauscher, B., Rhodes, J., ... Zhao, F. (2015).

Springel, V. (2005). *Mon. Not. R. Astron. Soc.*, *364*, 1105–1134. https://doi.org/10.1111/j.1365-2966.2005.09655.x

Szapudi, I. (2005).

Szapudi, I., & Szalay, A. S. (1998). *Astrophys. J.*, *494*, L41. https://doi.org/10.1086/311146

Taruya, A., Bernardeau, F., Nishimichi, T., & Codis, S. (2012). *Phys. Rev. D*, *86*, 103528. https://doi.org/10.1103/PhysRevD.86.103528

Taruya, A., Nishimichi, T., & Saito, S. (2010). *Phys. Rev. D*, *82*, 063522. https://doi.org/10.1103/PhysRevD.82.063522

Tegmark, M., Eisenstein, D. J., Strauss, M. A., Weinberg, D. H., Blanton, M. R., Frieman, J. A., Fukugita, M., Gunn, J. E., Hamilton, A. J. S., Knapp, G. R., Nichol, R. C., Ostriker, J. P., Padmanabhan, N., Percival, W. J., Schlegel, D. J., Schneider, D. P., Scoccimarro, R., Seljak, U. š., Seo, H.-J., ... York, D. G. (2006). *Phys. Rev. D*, *74*, 123507. https://doi.org/10.1103/PhysRevD.74.123507

Vargas-Magaña, M., Ho, S., Cuesta, A. J., O'Connell, R., Ross, A. J., Eisenstein, D. J., Percival, W. J., Grieb, J. N., Sánchez, A. G., Tinker, J. L., Tojeiro, R., Beutler, F., Chuang, C.-H., Kitaura, F.-S., Prada, F., Rodrıéguez-Torres, S. A., Rossi, G., Seo, H.-J., Brownstein, J. R., ... Thomas, D. (2018). *Mon. Not. R. Astron. Soc.*, *477*, 1153–1188. https://doi.org/10.1093/mnras/sty571

Verde, L., Treu, T., & Riess, A. G. (2019). *Nat. Astron.*, *3*, 891–895. https://doi.org/10.1038/s41550-019-0902-0

Veropalumbo, A., Binetti, A., Branchini, E., Moresco, M., Monaco, P., Oddo, A., Sánchez, A., & Sefusatti, E. (2022). *J. Cosmol. Astropart. Phys.*, *2022*, 33. https://doi.org/10.1088/1475-7516/2022/09/033

Veropalumbo, A., Sáez Casares, I., Branchini, E., Granett, B. R., Guzzo, L., Marulli, F., Moresco, M., Moscardini, L., Pezzotta, A., & de la Torre, S. (2021). *Mon. Not. R. Astron. Soc.*, *507*, 1184–1201. https://doi.org/10.1093/mnras/stab2205

Wang, Y., Xu, L., & Zhao, G.-B. (2017). *Astrophys. J.*, *849*, 84. https://doi.org/10.3847/1538-4357/aa8f48

Wang, Y., Zhao, G.-B., Zhao, C., Philcox, O. H. E., Alam, S., Tamone, A., de Mattia, A., Ross, A. J., Raichoor, A., Burtin, E., Paviot, R., de la Torre, S., Percival, W. J., Dawson, K. S., Gil-Marín, H., Bautista, J. E., Hou, J., Koyama, K., Peacock, J. A., ... Schneider, D. P. (2020). *Mon. Not. R. Astron. Soc.*, *498*, 3470–3483. https://doi.org/10.1093/mnras/staa2593

Yankelevich, V., & Porciani, C. (2019). *Mon. Not. R. Astron. Soc.*, *483*, 2078–2099. https://doi.org/10.1093/mnras/sty3143