Alma Mater Studiorum · Università di Bologna

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

Beyond 2-Point Statistics: Constraining Cosmology through a Joint 2- and 3- Point Correlation Functions Analysis of BOSS DR12

Graduation Thesis

Presented by:
Lorenzo Cavazzini

Supervisor:
Prof. Michele Ennio Maria Moresco

Co-supervisor:
Dr. Massimo Guidi

Academic Year 2024/2025
Graduation Date V

Abstract

The study of the Large-Scale Structure of the Universe has become one of the most prominent cosmological probes. Thanks to the advent of large galaxy surveys, we are now able to map our cosmos with unprecedented accuracy; while providing support to the establishment of the ΛCDM model, this progress is also opening new questions in modern cosmology. The increasing volume of data collected by current and future spectroscopic surveys now calls for methods beyond the standard ones to maximise the scientific return of such missions. Clustering analyses suggest that the combination of lower- and higher-order statistics is one of the most promising routes, especially in configuration space where limitations in measurement and theoretical modelling have so far prevented such analyses.

In this Thesis, we derive, for the first time on BOSS data, cosmological constraints from a joint analysis of two-point (2PCF) and three-point correlation functions (3PCF), quantifying the gain that can be achieved by including higher-order statistics. We develop a versatile Bayesian pipeline, capable of combining measurements, estimating covariance matrices, computing models, and deriving constraints via an MCMC approach across different configurations. We thoroughly validate this pipeline by creating dedicated synthetic data, testing the robustness of the code, inspecting the degeneracies between parameters, and determining the optimal regimes to minimise them. We apply the pipeline to the BOSS galaxies, fitting the first three multipoles of the 2PCF and the full 3PCF. We derive constraints on the bias, Alcock-Paczyński, and growth of structure parameters in two redshift bins, finding an improvement of a factor of 2 to 5 in the accuracy of the bias parameters from the joint fit.

This work demonstrates how crucial it is to include higher-order clustering statistics in future surveys, and contributes to the paper Guidi et al. (including L. Cavazzini) (2026, in prep.).

Contents

Introduction

Scientific background

Over the past two decades, several complementary observational probes have contributed to shaping a coherent cosmological model. One of the most important discoveries is the accelerated expansion of the Universe (Perlmutter et al. 1999, Riess et al. 1998). While the cosmological constant Λ provides a phenomenological description of this acceleration, its physical origin is still unclear, leaving room for a possible evolution of the dark energy component or for alternative explanations involving a modification of General Relativity (Huterer and Shafer 2017). Moreover, the increased precision of cosmological observations has revealed tensions between measurements obtained from the early- and late-Universe. The most notable one is the so-called "Hubble tension", a 5σ discrepancy between the value of the Hubble constant obtained from Cosmic Microwave Background (CMB) and from the local distance ladder based on type Ia Supernovae (SNe Ia).

The Large-Scale Structure (LSS) of the Universe provides a natural testing ground to probe its expansion history. Baryon Acoustic Oscillations (BAO) imprint a characteristic scale in the clustering of galaxies and can thus be used as a *standard ruler* (Eisenstein et al. 2005). A detailed measurement of the growth rate as a function of redshift (f) can also shed light on the underlying theory of gravity (Guzzo et al. 2008). The Baryon Oscillations Spectroscopic Survey (BOSS) (Eisenstein et al. 2011) was designed to measure the expansion history by means of BAO, with the goal of reaching an uncertainty of $\sim 1-2\%$ in the angular diameter distance $D_A(z)$ and Hubble parameter $H(z)$ out to redshift $z \sim 0.3-0.6$ (Alam et al. 2017).

The standard approach involves measuring and analysing the correlation functions, starting with the 2-point statistics, i.e. the 2-point correlation function (2PCF) or its Fourier-space counterpart (the power spectrum $P(k)$). During the last years, it has been demonstrated that additional information can be extracted by going beyond lower-order analysis; this is done by exploring 3-point statistics, and in particular the *3-point correlation function* (3PCF) and its Fourier counterpart, the bispectrum. The 3PCF signal is sensitive to non-Gaussianity of the density field, either related to gravitational instability, which naturally develops non-Gaussian features, or to primordial origin (for a review Celoria and Matarrese 2018). Moreover, the complementary sensitivity of the 3PCF with respect to the 2PCF to galaxy biasing helps break the degeneracies between parameters beyond the bias terms, thereby tightening the constraints on the cosmological parameters of interest (Veropalumbo et al. 2022, Gil-Marín et al. 2016, Euclid Collaboration: Guidi et al. 2025). As new data from spectroscopic surveys such as *DESI* (DESI Collaboration

et al. 2016) and *Euclid* (Euclid Collaboration et al. 2025) will increase the precision of these measurements, the implementation of complementary probes such as higher-order statistics will be crucial to maximise the scientific exploitation of these surveys.

Despite the large amount of information contained in 3-point statistics, several problems challenge their implementation. First of all, large volumes are needed to measure their signal with high accuracy, and building well-conditioned covariance matrices for the estimation of uncertainties requires a large amount of realisations, to be able to exploit all triangle configurations. While some cosmological analyses that explore the bispectrum on real data have been carried out (e.g., in BOSS Gil-Marín et al. 2016, Philcox and Ivanov 2022), the implementation in configuration space hasn't seen the same amount of attention. The main advantage of the 3PCF is that it can naturally account for the typically complex geometry that characterises the volume of galaxy surveys, differently from Fourier space analysis, in which the survey footprint imprints an intricated coupling of the wave modes that contribute to the statistic. Correcting for these convolution effects, in 3 dimensions, is far from trivial and constitutes a good reason to carry on the analysis in configuration space.

Measuring the 3PCF, however, has a high computational cost since it relies on estimators based on direct triplet counting. Until a few years ago, this was a real computational bottleneck, but the recent introduction of an estimator based on spherical harmonic decomposition (SHD) (Slepian and Eisenstein 2015) has significantly reduced the computational burden, allowing for the first time the extension of measurements to larger scales and datasets.

The challenges in implementing the 3PCF also show up in its modelling. Theoretical predictions of clustering statistics are based on cosmological perturbation theory, in which one solves the equations of motion that describe cosmic fields such as $\delta(\boldsymbol{k})$ with a perturbative expansion. This approach is best treated in Fourier space since wave modes k are decoupled (at least in the linear approximation). To get to the configuration space counterparts, one has to perform an inverse Fourier transformation. Galaxy surveys introduce distortions in the clustering that make the distribution of galaxies anisotropic, and the resulting bispectrum depends on several variables. Even when restricting to the isotropic part, the transform is computationally demanding. The 2D-FFTlog algorithm (Fang et al. 2020) is by far the most efficient way of performing the two-dimensional Hankel transform to go from the Legendre coefficients of the bispectrum to those of the 3PCF. Still, direct exploration of the cosmological parameter space is out of reach for *full-shape* studies. To overcome this issue, one can rely on *emulators*, machine learning algorithms (mostly Neural Networks) trained on libraries of theoretical models. These systems are capable of producing accurate predictions, reducing the time needed to compute a single model by an order of 10^5 in the case of the 3PCF. This approach has been recently introduced by Euclid Collaboration: Guidi et al. (2025) for a *full-shape* analysis of the joint 2- and 3- point correlation functions in numerical simulations, in preparation for Euclid. This work represents the first *full-shape* study of the real space 3PCF and demonstrates the advantages of combining the 3PCF with the 2PCF, which yields tighter constraints on shape parameters than the 2PCF alone.

A complementary, more computationally light strategy is the *template-fitting* ap-

proach. Theoretical models of correlation functions fundamentally depend on the linear power spectrum, whose shape is largely determined by parameters tightly constrained by CMB anisotropy measurements (Planck Collaboration: Aghanim et al. 2020). By fixing this "early-time physics" using a Planck prior, the theoretical predictions can be linearized and expressed in terms of a set of fitting parameters combined with precomputed templates. These fitting parameters comprise both cosmological parameters that depend on late-time evolution and a collection of nuisance parameters. The late-time parameters primarily capture distortions in the observed galaxy clustering that arise when inferring positions from spectroscopic survey data.

Goal of the thesis

A joint 2- and 3-point correlation function cosmological analysis that exploits the full scale dependence of the three-point correlation function without restrictions on particular triangle configurations has never been made on BOSS data. The exploration of the cosmological parameter space has been made on BOSS for the bispectrum, both in the *template fit* (Gil-Marín et al. 2016) and *full-shape* (Philcox and Ivanov 2022) approaches, and for the 3PCF it has been restricted to the isotropic AP parameter α with only a few triangle configurations used (Slepian et al. 2017). Addressing this challenge is fundamental in preparation for the full scientific exploitation of the wealth of data that are currently collected by Stage IV surveys, and that will be delivered by Stage V surveys.

The primary objective of this work is to investigate the additional information gained by combining the 3PCF with the 2PCF in constraining the linear growth rate f and in assessing the possibility of breaking its degeneracy with the present-day root-mean-square of density fluctuations smoothed over spheres of $8h^{-1}$Mpc, σ_8. In addition, we provide an independent measurement of cosmological distances with respect to other BOSS studies, although this constraint is derived solely from the 2PCF. This work also presents the first analysis of the BOSS reduced 3PCF, which has received even less attention in the literature than the connected 3PCF.

Taking advantage of the analysis performed in Guidi et al. (including L. Cavazzini) (2026, in prep.), we extract measurements of the BOSS DR12 2PCF monopole, quadrupole, hexadecapole, and isotropic 3PCF Legendre coefficients from $\ell = 0$ to 10, as well as covariance matrices estimated by means of 2000 MultiDark-`Patchy` mocks (Kitaura et al. 2016).

Crucial to the analysis is the computation of theoretical predictions. We utilize `MElCorr` (Guidi et al. (including L. Cavazzini) 2026, in prep.), which disposes of theoretical tools based on perturbation theory schemes most recently developed.

We constructed a pipeline for the exploration of the parameter space in which the posterior distribution is sampled with a Markov Chain Monte Carlo (MCMC) algorithm. This framework allows us to combine the BOSS 2- and 3-point correlation functions (and the reduced 3PCF) measures and theoretical models to derive and analyse cosmological constraints. We can fit the two statistics separately or jointly by considering the cross-covariance between the two. Since BOSS is divided into four separate data sets, either at different effective redshift (low-z and high-z) or spatially separate (North Galactic Cap or

South Galactic Cap), we can fit them separately or in a combined way by summation of the likelihood function at each step.

Outline

This thesis is structured as follows:

- In Chapter 1, we introduce the ΛCDM model and briefly review the fundamental equations of Cosmology in a homogeneous and isotropic Universe. Then, we discuss the clustering of matter and galaxies, starting from the 2PCF, and the power spectrum, followed by the 3PCF and bispectrum. We then address the differences between the two main approaches when it comes to cosmological analysis of clustering statistics and highlight where the information content of each is. We summarize the ingredients of our models under the *template fitting* approach. Finally, we motivate the work made in this thesis, by highlighting the current tensions and unexplained assumptions of the standard model and how higher-order statistics might aid us in testing them.

- In Chapter 2, we describe the BOSS catalogue and the estimators used for measuring the 2PCF and 3PCF multipoles in spectroscopic surveys. Then, we comment on the measured statistics and the different ways we can visualise, and utilize the 3PCF. We also describe how to obtain covariant errors for the different statistics.

- In Chapter 3, we present the new Bayesian framework that was developed, as well as the analysis performed on synthetic mock data created to validate our code and verify which parameters can be confidently constrained given the data and models at hand. We address the degeneracies between b_1, f, and σ_8, emerging in both the 2PCF, 3PCF, reduced 3PCF, and joint 2PCF and 3PCF, by employing various strategies and finding the best setup, which we consider as our baseline.

- In Chapter 4, we first fit the BOSS 2- and 3-point correlation functions separately, to explore which scale cuts minimize systematic errors in the modelling while retaining the highest signal. Applying these cuts, we present the results from fitting the 2- and 3-point correlation functions of combined BOSS data sets. Firstly, we show results from the 2PCF alone in terms of AP parameters; we extract their one-dimensional marginalized contours and perform a distance measurement comparing the result with other BOSS papers and the ΛCDM prediction. Secondly, we compare the marginalized posteriors from the 2PCF alone, 3PCF alone, and joint 2PCF and 3PCF when σ_8 is fixed to a CMB-based prior, to assess the ability to constrain parameters for each. Lastly, we show the results from a combined fit of the joint 2PCF and 3PCF with a uniform prior on σ_8 and compare the results with the previous fit and measurements of f and σ_8 from other studies.

- In Chapter 5, we draw our conclusions and discuss future prospects.

Chapter 1

Theoretical Framework

In this chapter, we review the principles of modern cosmology, highlight some of its currently unexplained phenomena, and give an overview of the theoretical framework needed to describe the Large Scale Structure (LSS) of the Universe, which we can use as a probe to test cosmological models. We first introduce the ΛCDM model in Sec. 1.1 which constitutes the current cosmological paradigm. In Sec. 1.2, the framework for probing the LSS is presented by defining the N-point statistics, which are central to this thesis. We explore the current approaches to the study of N-point correlation functions in Sec. 1.3, mainly how we extract cosmological information in *Full-shape* and *Template fitting* analyses. In Sec. 1.4, we review all the crucial elements of the theoretical models; the theoretical description of N-point statistics is based on perturbation theory, first evaluated in Fourier space. To get to the 2-point and 3-point correlation function models, an inverse Fourier transform is performed. Finally, in Sec. 1.5 we explore the questions that have arisen with the increased precision of cosmological probes and how higher-order correlation functions can aid us in answering them.

1.1 The ΛCDM Model

The model currently able to reproduce cosmological observations most robustly and consistently is the so-called ΛCDM model (Λ Cold Dark Matter) or Standard Model of Cosmology. According to the ΛCDM model, the Universe originated, through the Big Bang, from an initial singularity, that is, from a point-like state with divergent density and temperature. The distribution of energy at this point should be completely uniform and isotropic; if this were the case, however, we should not observe structures that appear today as inhomogeneities throughout space-time. Leading theories postulate that an inflation period (Guth 1981), in the very early Universe, provided the seed perturbations that would go on to form today's structures. While little observational evidence is provided for any particular inflation model, it remains a pillar in the standard view of cosmology as it provides the resolution to many problems associated with it. According to the ΛCDM model, the current energy budget of the Universe is dominated by the following components:

- **Dark Energy**: it is a component of unknown physical origin, with negative pressure and distributed homogeneously in space, which could justify the accelerated expansion of the Universe (Huterer and Shafer 2017). It is estimated that Dark Energy represents about 68% of the mass-energy of the Universe. This component is generally described in terms of a Cosmological Constant Λ;

- **Cold dark matter (CDM)**: It is a non-relativistic and non-collisional matter component, with an electromagnetic interaction cross section so low as to make it invisible to direct observations. It is estimated that CDM constitutes about 27% of the energy budget of the Universe. The introduction of a component with these characteristics is justified by both dynamical observations on galactic scales (rotation curves of spiral galaxies) and extragalactic scales (measurements of virial mass of clusters), as well as those related to gravitational lensing. Its introduction also allows us to explain, through the mechanism of gravitational instability, the formation of structures, such as those that characterize the current Universe;

- **Baryonic matter**: It is composed mainly of light elements, such as Hydrogen and Helium and it represents about 5% of the mass-energy of the Universe;[1]

- **Radiation**: It makes up less than 0.1% of all energy density in the Universe. The majority of the photon energy density comes from the Cosmic Microwave Background, which permeates the Universe and is well described by a black body spectrum;

- **Neutrinos**: In the standard model, neutrinos are massless and their energy density is comparable to that of radiation. It was discovered that neutrinos have a non-negligible mass; in this case, they may play a significant role in structure formation through their free streaming. Pointing down the precise mass of neutrino species is one of the current endeavours of cosmological observations.

The ΛCDM model has become the standard model of cosmology largely because it provides an excellent fit to a wide range of observational data. Among the most significant pieces of evidence supporting it is the Cosmic Microwave Background (CMB), the nearly perfect blackbody radiation that permeates the Universe with a temperature $T_0 = 2.7255 \pm 0.0006$ K. The CMB is remarkably isotropic, exhibiting temperature anisotropies of order $\Delta T \approx 10^{-5}$. These small fluctuations contain a wealth of information about the primordial Universe. In particular, the *Planck* mission (Planck Collaboration: Aghanim et al. 2020) measured the CMB anisotropies with unprecedented precision, allowing tight constraints to be placed on the parameters of the standard cosmological model. In the context of CMB anisotropy measurements, the ΛCDM model is specified by six fundamental parameters: the physical baryon density $\Omega_b h^2$, the physical cold dark matter density $\Omega_c h^2$, the amplitude of the primordial power spectrum A_s, its spectral

[1] In Astrophysics 'baryons' are defined as all of the ordinary matter in the Universe, so everything made by protons, neutrons and electrons.

index n_s, the ratio of the sound horizon to the angular diameter distance at last scattering θ_*, and the reionization optical depth τ (Scott 2018). From this set, many other commonly discussed cosmological parameters can be derived.

Additional key probes of cosmology include Baryon Acoustic Oscillations (BAO) measured in Large Scale Structure (LSS) surveys (Eisenstein et al. 2005) and observations of Type Ia supernovae (Riess et al. 1998, Perlmutter et al. 1999). These provide independent measurements of the distance-redshift relation: BAO act as a *standard ruler*, corresponding to a feature of known physical size, while Type Ia supernovae serve as *standard candles*, objects of known intrinsic luminosity. Together, these probes have provided strong evidence for the accelerated expansion of the Universe, supporting the interpretation that the present cosmic expansion is driven by a cosmological constant Λ. Another important probe is Big Bang Nucleosynthesis (BBN), which places tight constraints on the physical baryon density $\Omega_b h^2$ (Cooke et al. 2014).

1.1.1 Background evolution

The ΛCDM model sits on a few core assumptions: on the largest scales, the Universe is homogeneous and isotropic, an assumption known as the *Cosmological Principle*; the only interaction relevant at large scales is gravity, described by *General Relativity* (GR) (Coles and Lucchin 2003). Under these assumptions, a formal mathematical framework can be constructed that ultimately describes the Universe at large scales with only six degrees of freedom.

In GR, the force of gravity can be interpreted as the curvature of space-time induced by the energy and momentum distribution, which in turn responds to the local curvature. This interaction is well described by the Einstein field equations

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{1.1.1}$$

These are a system of coupled, non-linear equations for the metric tensor $g_{\mu\nu}$; they relate the 'space-time' curvature encoded in the left-hand side of the equation with the energy-momentum tensor $T\mu\nu$ on the right-hand side. The $g_{\mu\nu}$ tensor is a $(0, 2)$ rank symmetric tensor which specifies the geometric properties of the local space-time. We can define the infinitesimal interval between two events in space-time as

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{1.1.2}$$

No general solution exists for the field equations and we must specify some symmetry argument that reduces the number of degrees of freedom. The *Cosmological Principle* postulates that on sufficiently large scales the Universe is homogeneous (distributed uniformly) and isotropic (looks the same in every direction). Under this principle the 'mixed' terms $g_{0\nu}$ vanish and we are left with the general form

$$ds^2 = -c^2 dt^2 + dl^2.$$

Taking dl^2 to be any geometry with curvature K and rescaling the overall spatial component of the metric by a factor $a(t)$, in spherical coordinates, we get:

$$ds^2 = -c^2dt^2 + a(t)^2\left(\frac{dr^2}{1-Kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right) \tag{1.1.3}$$

The *Friedmann-Lemaitre-Robertson-Walker* (FLRW) metric. This form of the metric shows that the dynamics of the Universe on the largest scales is determined by one degree of freedom, the scale factor $a(t)$, once we have fixed the geometry K, which can be rescaled to take values $K = \{-1,\ 0,\ 1\}$ respectively denoting an *open*, *flat* and *closed* spatial geometry. r, θ, ϕ are the comoving coordinates and t is the proper time. While the comoving coordinates are time independent, the scale factor acts to rescale spacetime intervals at any time t.

It is not trivial to define distances in an expanding universe. We introduce the proper distance as the distance between two events at a fixed proper time $dt = 0$

$$d_p(t) = a(t)\int_0^r \frac{dr'}{\sqrt{1-Kr'^2}} = a(t)\, F(r) \tag{1.1.4}$$

Here we see that the proper distance is time-dependent and it changes based on the geometry encoded in the $F(r)$ factor. By taking the time derivative of the proper distance we derive the radial recession velocity of the origin

$$\dot{d}_p = \dot{a}(t)F(r) = \frac{\dot{a}}{a}a(t)\, F(r) = H(t)\, d_p \tag{1.1.5}$$

This is the Hubble law where we have defined $H(t) \equiv \frac{\dot{a}(t)}{a(t)}$[2] the Hubble parameter. The value of $H(t)$ at the present time is the Hubble constant H_0. Experimentally, the expansion of the Universe was first discovered by Hubble in 1929 (Hubble 1929), who measured the recession velocity of galaxies and found a linear relation between the velocity and their distance from us.

Another important consequence of an expanding universe is cosmological redshift

$$\frac{a(t_o)}{a(t_e)} = \frac{\lambda_o}{\lambda_e} = 1 + z \tag{1.1.6}$$

where z is the redshift, i.e., the change in wavelength due to the recession of the emitted source from the observer. Cosmological redshift is not a proper Doppler effect; it is instead due to the mere dilation of space that makes photons lose energy as they travel.

A system consisting of many particles, such as the Universe, can be conveniently described through an Eulerian approach, that is, treating it as if it were a fluid characterized by macroscopic variables defined over the entire spatial domain. In particular, the Universe is treated as a perfect fluid, that is, a fluid whose physical properties are completely specified by energy density ρ and isotropic pressure p. In this approximation, we neglect

[2] Cosmologists often use $h = \frac{H_0}{100\ \mathrm{km/s/Mpc}}$ the dimensionless hubble constant; approximately $H_0 \sim 70\ \mathrm{km/s/Mpc}$, $h \sim 0.7$

thermal conduction and viscosity. The energy-momentum tensor associated with a perfect fluid has the form:

$$T_{\mu\nu} = (\rho + \frac{p}{c^2})u_\mu u_\nu + p g_{\mu\nu} \tag{1.1.7}$$

where u_μ is the four-velocity. The solution to the Einstein equations associated with a FLRW metric for a perfect fluid are the so-called *Friedmann equations*:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} \tag{1.1.8}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) \tag{1.1.9}$$

1.1.8 and 1.1.9 are differential equations for the scale factor $a(t)$, they determine its evolution based on the energy budget of the Universe.

Under the assumption that the Universe is adiabatic we can write the continuity equation as

$$\frac{d\rho}{dt} + 3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0 \tag{1.1.10}$$

Eqs. 1.1.8, 1.1.9 and 1.1.10 are not, however, sufficient to describe the evolution of the Universe at large scales. To do so, it is necessary to specify the connection between pressure and energy density. In other words, it is necessary to provide an *equation of state* that allows us to characterise the intrinsic properties of the fluid, or fluids, that constitute the Universe. The equation of state of a perfect fluid can be expressed in the so-called Zel'dovich form:

$$p = w\rho c^2 \tag{1.1.11}$$

where, in the case of physical fluids, the parameter w is related to the speed of sound in the medium $c_s = \left(\frac{\partial p}{\partial \rho}\right)^{\frac{1}{2}}$. For this reason, since $c_s < c$ must be valid, w can only take values in the interval $[0, 1)$. For example, a fluid of non-relativistic and non-collisional matter has $w = 0$. A fluid of radiation, on the other hand, has $w = \frac{1}{3}$.

Consider the case of a Universe composed (or at least dominated) by a single cosmological fluid. Assuming that w is a constant, by making explicit the form of pressure in the continuity equation 1.1.10 and solving by separation of variables, we obtain:

$$\rho(t) = \rho_0 \left(\frac{a_0}{a}\right)^{3(1+w)} \tag{1.1.12}$$

Where ρ_0 indicates the current average density of the Universe. Some examples are:

- a dust fluid has $\rho \propto a^{-3}$, the evolution of the energy density of a matter-dominated Universe is completely determined by the increase in its volume;
- a radiative fluid has $\rho \propto a^{-4}$, In addition to the term a^{-3}, due to the increase in the volume of the Universe, photons lose energy due to the cosmological redshift, which is responsible for the additional factor a^{-1};

- a cosmological constant Λ has $\rho = const.$ therefore such a fluid is characterized by $w = -1$.

From Eq. 1.1.8 and by imposing $K = 0$ we find

$$\rho_{crit} = \frac{3H^2}{8\pi G}$$

is the critical density, i.e., the density required to have null curvature. It is useful to characterize the density relative to the critical one. We define the density parameter Ω_i as:

$$\Omega_i(t) = \frac{\rho_i(t)}{\rho_{crit}}$$

This is a particularly valuable relation, as it shows how the global curvature of space-time depends on the average density of the Universe. We can exploit this new definition in conjunction with Eq. 1.1.8 and 1.1.12 to express the evolution of the scale factor as:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2 \left[(\sum_i 1 - \Omega_{0,i}) + \sum_i \Omega_{0,i} \left(\frac{a_0}{a}\right)^{1+3w_i} \right] \tag{1.1.13}$$

In the case of a flat Universe, as current observations support, $\Omega_{0,tot} = 1$.

1.2 Clustering statistics

Present-day structures such as galaxies and clusters have formed thanks to a long process of gravitational instability that started in the primordial Universe. The leading theory is that Inflation (Guth 1981) seeded perturbations in the density field, small inhomogeneities in the distribution of dark matter, which we describe in terms of a scalar field defined as

$$\delta(\boldsymbol{x}, t) = \frac{\rho(\boldsymbol{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)} \tag{1.2.1}$$

where $\bar{\rho}$ is the mean value of the density field and is therefore only time dependent. There are other cosmic fields of interest for the evolution of structures such as the gravitational potential and velocity divergence. Here, we only describe the density field and invoke the others when needed.

We want to describe the cosmic fields in a statistical way rather than a deterministic one, for two reasons. First, we lack access to the initial conditions that would allow us to evolve the system in a deterministic way. Second, cosmological timescales are too long to follow the evolution of structures. We only witness the evolution of different structures at all cosmic times through our past light cone.

At the foundations of a statistical description of cosmic fields is the *cosmological principle* which can be defined in a statistical sense as follows. A random field is called *statistically homogeneous* if all the joint multi-point probability distribution functions

$p(\delta(\boldsymbol{x}_1), \ldots, \delta(\boldsymbol{x}_n))$ remain unchanged under spatial translations, thus the probabilities depend only on the relative positions. A stochastic field is called statistically isotropic if $p(\delta(\boldsymbol{x}_1), \ldots, \delta(\boldsymbol{x}_n))$ is invariant under spatial rotations. In Sec. 1.3.2 we abandon this assumption as spectroscopic surveys introduce significant deviations from statistical isotropy for the redshift-space density field.

Inflationary models suggest that the primordial contrast density field has a nearly Gaussian distribution. This means that given N points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$, the probability of obtaining the values of primordial contrasts $\delta_1, \ldots, \delta_N$ at $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ is well described by a multivariate Gaussian distribution. If we denote $\boldsymbol{V} \equiv (\delta_1, \ldots, \delta_N)$, this distribution writes

$$P_N(\boldsymbol{V}) = \frac{\det(M)^{\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}} \, e^{-\frac{1}{2}\boldsymbol{V}^T M \boldsymbol{V}}, \tag{1.2.2}$$

where M is the inverse of the covariance matrix C, whose element C_{ij} is

$$C_{ij} = \langle \delta(\boldsymbol{x}_i)\delta(\boldsymbol{x}_j) \rangle$$

the expectation value of the product of δ at the two positions $\boldsymbol{x}_i$, $\boldsymbol{x}_j$. Notice that the multivariate Gaussian distribution is fully determined by its covariance matrix; furthermore, Wick's theorem (Wick 1950) makes it possible to compute the higher-order moments of a multivariate Gaussian distribution in terms of its covariance matrix.

According to Wick's theorem, odd statistics of a Gaussian field, namely $\langle \delta(\mathbf{x}_1) \ldots \delta(\mathbf{x}_n) \rangle$ with n odd, are zero, since for a set with an odd number of elements it is not possible to find a partition into pairs. Even statistics, i.e. $\langle \delta(\mathbf{x}_1) \ldots \delta(\mathbf{x}_n) \rangle$ with an even n, are all functions of 2-point statistics $\langle \delta(\mathbf{x}_i)\delta(\mathbf{x}_j) \rangle$.

1.2.1 2-point statistics

We can define 2-point statistics in terms of the excess probability, with respect to a completely random distribution of matter particles, of finding two of these particles separated by a vector $\boldsymbol{r}$. The joint probability $dP_{1,2}$ of finding two particles in the comoving volume elements dV_1 and dV_2, centered at the positions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, and separated by a vector $\boldsymbol{r} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, which can be expressed, under the assumptions of statistical homogeneity and isotropy of the Universe, as:

$$dP_{1,2} \propto \bar{n}^2 \, dV_1 dV_2 \left[1 + \xi(r)\right], \tag{1.2.3}$$

where $r = |\boldsymbol{r}|$, $\bar{n}$ is the average number density of particles and $\xi(r)$ is the so-called *2-point correlation function* (2PCF), defined as the joint ensemble average of the density contrast at two different locations:

$$\xi(r) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x} + \boldsymbol{r}) \rangle, \tag{1.2.4}$$

where we have omitted, as we will do in the following if not otherwise specified, the time dependence in δ.

It is also convenient from the theoretical point of view to consider the Fourier space analogue of the 2PCF, called *power spectrum*. If we denote with $\tilde{\delta}(\boldsymbol{k})$ a Fourier mode of the density contrast field with wavevector $\boldsymbol{k}$, namely

$$\tilde{\delta}(\boldsymbol{k}) = \int d^3x\, \delta(\boldsymbol{x})\, e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}\,, \tag{1.2.5}$$

The power spectrum $P(k)$ is defined in terms of correlators in Fourier space through the relation

$$\left\langle \tilde{\delta}(\boldsymbol{k})\tilde{\delta}(\boldsymbol{k'}) \right\rangle \equiv (2\pi)^3 \delta^{(D)}(\boldsymbol{k}+\boldsymbol{k'})P(k), \tag{1.2.6}$$

where the symbol $\delta^{(D)}$ denotes the Dirac delta. It can be shown, by expanding the density contrast in Eq. 1.2.4 into Fourier modes and using Eq. 1.2.6, that the 2PCF and the power spectrum are a Fourier pair, i.e., they satisfy the relation

$$\xi(r) = \int \frac{d^3k}{(2\pi)^3} P(k)\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\,. \tag{1.2.7}$$

The power spectrum can be interpreted as the contribution of the k-mode to the variance σ^2 of the fluctuations of the density field δ, defined as $\sigma^2 \equiv < \delta^2 >$ (recall that the field δ has zero mean by definition). The variance of the density contrast is

$$\begin{aligned}
\sigma^2 \equiv \left\langle \delta(\boldsymbol{x})^2 \right\rangle &= \frac{1}{(2\pi)^6} \int d^3k\, d^3k' e^{i(\boldsymbol{k}+\boldsymbol{k'})\cdot\boldsymbol{x}} \left\langle \tilde{\delta}(\boldsymbol{k})\, \tilde{\delta}(\boldsymbol{k'}) \right\rangle \\
&= \frac{1}{(2\pi)^3} \int d^3k\, d^3k' e^{i(\boldsymbol{k}+\boldsymbol{k'})\cdot\boldsymbol{x}}\, \delta^{(D)}(\boldsymbol{k}+\boldsymbol{k'})P(k) \\
&= \frac{1}{2\pi^2} \int_0^\infty dk\, k^2 P(k),
\end{aligned} \tag{1.2.8}$$

where we used Eqs. 1.2.5 and 1.2.6, and in the last step, we switched to polar coordinates.

From a practical standpoint, the density contrast field cannot be measured with *infinite* resolution. Instead, the properties of the underlying continuous matter density field must be inferred from discrete tracers, such as dark matter halos or galaxies. Since these objects are associated with a characteristic spatial scale, it is convenient to consider the density contrast field δ smoothed below a given scale R. We denote the density contrast field filtered on the scale R by δ_R. Its value at a given position is obtained by convolving the original density field with a region of characteristic size R centered at that position, whose shape is specified by the choice of filter $W(\boldsymbol{x}, R)$. Recalling that the Fourier transform of a convolution is equal to the product of the Fourier transforms of the individual functions, the variance of the smoothed field can be written as

$$\sigma_R^2 = \frac{1}{2\pi^2} \int_0^\infty dk\, k^2\, P(k)\, W(k,R)^2 \tag{1.2.9}$$

Notice that σ_R provides information about the amplitude of perturbations at the scale R. For historical reasons, a common choice for specifying the density perturbation amplitude is the root-mean-square (RMS) of the present linear-theory density field, filtered on

a scale $R = 8\,h^{-1}$Mpc. This quantity is one of the fundamental cosmological parameters, known as σ_8.

If the field δ were *perfectly* Gaussian, then, as a direct consequence of Wick's theorem, any ensemble average of a product of Fourier modes, $\left\langle \tilde{\delta}(\boldsymbol{k}_1)\cdots\tilde{\delta}(\boldsymbol{k}_n)\right\rangle$, could be written entirely in terms of ensemble averages of pairs, $\left\langle \tilde{\delta}(\boldsymbol{k}_i)\tilde{\delta}(\boldsymbol{k}_j)\right\rangle$, with $i,j \in \{1,\ldots,n\}$. For a Gaussian field with vanishing mean, this implies that all statistical information is fully encoded in the two-point correlation function (2PCF) or, equivalently, in the power spectrum. However, there are several reasons why two-point statistics alone is not sufficient to fully characterize the density field of the Universe. First, primordial fluctuations may deviate from perfect Gaussianity, a possibility commonly referred to as *primordial non-Gaussianity* (PNG; see Celoria and Matarrese 2018 for a review). Second, the dynamics of gravitational instability is intrinsically nonlinear; even if the primordial density field were exactly Gaussian, gravitational evolution would inevitably generate non-Gaussian features as cosmic time progresses. These considerations motivate the study of higher-order statistics beyond the two-point level, starting with the three-point statistics, which we introduce in the following section.

1.2.2 3-point statistics

The lowest-order clustering statistics that can encode information about non-Gaussian properties of the LSS are 3-point statistics.

By following the same approach as in 2-point statistics, we can define the excess probability, with respect to a perfectly Gaussian distribution of matter particles, of finding three particles at the vertices of a triangle of given shape and size. This time, the joint probability $dP_{1,2,3}$ of finding three particles in the comoving volume elements dV_1, dV_2 and dV_3 at the positions $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$, separated by the vectors $\boldsymbol{r}_{12} = \boldsymbol{x}_2 - \boldsymbol{x}_1$, $\boldsymbol{r}_{13} = \boldsymbol{x}_3 - \boldsymbol{x}_1$, $\boldsymbol{r}_{23} = \boldsymbol{x}_3 - \boldsymbol{x}_2$, can be expressed, under the assumptions of statistical homogeneity and isotropy, as:

$$dP_{1,2,3} \propto \bar{n}^3\, dV_1 dV_2 dV_3\, [1 + \xi(r_{12}) + \xi(r_{13}) + \xi(r_{23}) + \zeta(r_{12}, r_{13}, r_{23})]\,, \tag{1.2.10}$$

$\zeta(r_{12}, r_{13}, r_{23})$ is the so-called *connected 3-point correlation function* (3PCF), defined as the joint ensemble average of the density contrast at three different locations:

$$\zeta(r_{12}, r_{13}, r_{23}) \equiv \langle \delta(\boldsymbol{x})\delta(\boldsymbol{x}+\boldsymbol{r}_{12})\delta(\boldsymbol{x}+\boldsymbol{r}_{13})\rangle\,, \tag{1.2.11}$$

with the constraint $|\boldsymbol{r}_{13} - \boldsymbol{r}_{12}| = r_{23}$.

The Fourier space analogue of the 3PCF, the *bispectrum* $B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)$ is

$$\left\langle \tilde{\delta}(\boldsymbol{k}_1)\tilde{\delta}(\boldsymbol{k}_2)\tilde{\delta}(\boldsymbol{k}_3)\right\rangle \equiv (2\pi)^3 \delta^{(D)}(\boldsymbol{k}_1 + \boldsymbol{k}_2 + \boldsymbol{k}_3) B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3). \tag{1.2.12}$$

With this definition, the 3PCF and the bispectrum are a Fourier pair since we have

$$\zeta(r_{12}, r_{13}, r_{23}) = \int \frac{d^3k_2}{(2\pi)^3}\frac{d^3k_3}{(2\pi)^3} B(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3)\, e^{i(\boldsymbol{k}_2\cdot\boldsymbol{r}_{12} + \boldsymbol{k}_3\cdot\boldsymbol{r}_{13})}\,. \tag{1.2.13}$$

Given ξ_N the N-point connected correlation function (so $\xi_2 \equiv \xi$ and $\xi_3 \equiv \zeta$) we can define the so-called *Hyerarchical Amplitudes* as

$$Q_N \equiv \frac{\xi_N}{\sum_{labellings} \prod_{edges\ i,j}^{N-1} \xi_2(r_{ij})} \tag{1.2.14}$$

In the perturbative description of LSS statistics, it is possible to prove that gravitational instability generates connected higher-order correlation functions that scale as $\xi_N \propto \xi_2^{N-1}$ at large scales, where Perturbation Theory (PT; Bernardeau et al. 2002) applies. This scaling can be represented by *tree diagrams* which we describe in Appendix A. Tree-level perturbation theory leads to functions,Q_N, independent of the amplitude of ξ_2. And for scale-free initial conditions, i.e., $P_{\rm in}(k) \propto k^n$, they are independent of overall scale. The Hierarchical amplitude for $N = 3$ is the first non-trivial function of this kind and it is often called the *reduced 3-point correlation function* (Peebles and Groth 1975)

$$Q(r_{12}, r_{13}, r_{23}) = \frac{\zeta(r_{12}, r_{13}, r_{23})}{\xi(r_{12})\xi(r_{13}) + \xi(r_{13})\xi(r_{23}) + \xi(r_{23})\xi(r_{12})} \tag{1.2.15}$$

Since at *tree-level* $\zeta \propto \xi^2$, Q should be independent of the amplitude of ξ (σ_8^2) at large scales. In this context, Q is better suited for visualizing the growth of non-Gaussian structure and the shape dependence of clustering than ζ.

From the practical point of view, at very large scales $\sim 120\,h^{-1}Mpc$ $\xi \to 0$ the reduced 3PCF becomes very unstable and its errors non-Gaussian, leading to an overestimation of the covariance (Bernardeau et al. 2002). Furthermore, modelling ξ in the mildly non-linear regime requires up to *one-loop* in perturbation theory, which breaks the scaling of Q from being independent of normalization to $Q \propto \sigma_8^4$.

1.3 Full-shape vs. Template-fitting

State-of-the-art cosmological analyses of clustering statistics can be broadly classified into two methodological approaches. Both seek to extract cosmological information through likelihood analysis of the power spectrum or 2PCF, and in our case, additionally, the bispectrum or 3PCF. Their theoretical descriptions are based on the linear matter power spectrum, whose shape is governed (in ΛCDM) by the parameter set

$$\{n_s,\ A_s,\ \omega_{cdm},\ \omega_b,\ h\},$$

which controls both the early-Universe physics and, through h, the late-time expansion history.

Spectroscopic galaxy surveys do not directly probe the matter density contrast $\delta(\boldsymbol{x})$, but rather the galaxy contrast $\delta_g(\boldsymbol{x})$. Moreover, galaxy positions are not observed in real space. Instead, surveys measure redshifts and angular coordinates on the sky (right ascension and declination), which are mapped to comoving coordinates according to

$$(z, \theta, \phi) \longrightarrow \boldsymbol{x}_{\rm obs}(z, \theta, \phi) = D_c(z), \hat{z}(\theta, \phi).$$

This observational procedure introduces two key effects.

1. Galaxies possess peculiar velocities that contribute to the observed redshift via Doppler shifts in addition to the Hubble flow. This induces anisotropies in the measured clustering pattern, enhancing correlations along the line-of-sight direction $\hat{z}$. This effect is known as *redshift-space distortions* (RSD).

2. The conversion from redshift to comoving distance requires the assumption of a fiducial cosmology,
$$D_c(z) = \int_0^z \frac{c}{H(z')}\,\mathrm{d}z', \tag{1.3.1}$$
which may differ from the true cosmological model. Any mismatch leads to apparent *geometric distortions* in the observed clustering signal.

Both effects constitute powerful sources of cosmological information. Accurate modelling of RSD enables constraints on the linear growth rate,
$$f \equiv \frac{\mathrm{d}\ln D_1}{\mathrm{d}\ln a}, \tag{1.3.2}$$
where D_1 denotes the fastest-growing linear mode of density perturbations (see Sec. 1.4.1). In parallel, the *Alcock-Paczynski* (AP) test (Alcock and Paczynski 1979) exploits geometric distortions to extract distance information, such as the comoving angular diameter distance $D_A(z)$ and the Hubble parameter $H(z)$.

Template-fitting analyses extract cosmological information primarily from distortions imprinted in the clustering signal around a fixed template. By fixing $P_{\rm lin}(k)$, all model contributions entering the power spectrum can be precomputed, effectively linearising the dependence on the fitting parameters. The resulting template takes the form
$$P_{\rm template}(k,\mu) = \sum_{p=1}^{N} X^{(p)}, P^{(p)}(k,\mu), \tag{1.3.3}$$
where the terms $P^{(p)}(k,\mu)$ are precomputed and independent of the fitting parameters, which instead appear in the coefficients $X^{(p)}$. These coefficients are linear combinations of the physical parameters of interest,
$$X^{(p)} = \{b_1^2\sigma_8^2,\ b_1 f\sigma_8^2,\ f^2\sigma_8^2,\ b_1 b_2\sigma_8^4,\ \alpha,\ \epsilon,\ \dots\} \tag{1.3.4}$$
The fitting parameters are a combination of bias parameters ($b_1,\ b_2, \dots$), encoding the relation between the matter and galaxy field, and cosmological parameters f,σ_8, α and ϵ which compress the information from dynamic and geometric distortions. During posterior sampling, only the coefficients $X^{(p)}$ are varied, while the precomputed tables $P^{(p)}(k,\mu)$ remain fixed. Consequently, model evaluation is computationally inexpensive. This strategy extends straightforwardly to the 2PCF and 3PCF by transforming the corresponding Fourier-space templates into configuration space and precomputing the relevant quantities.

Full-shape analyses, by contrast, aim to extract all the available cosmological information, encompassing both early-Universe physics and the late-time expansion history. Shape

parameters enter in the evaluation of the linear matter power spectrum and can not be linearised under a fixed template. The principal limitation of this approach is its computational cost. While theoretical predictions for the power spectrum and bispectrum can be evaluated relatively efficiently, obtaining fast and accurate predictions for the 2PCF and especially the 3PCF is more challenging, as it requires transforming Fourier-space quantities into configuration space via Eqs. 1.2.7 and 1.2.13.

Recent advances have mitigated this difficulty through the development of *emulators*, which enable rapid evaluation of theoretical models. In particular, they can reduce the computation time of a single 3PCF prediction by up to five orders of magnitude (Euclid Collaboration: Guidi et al. 2025).

In this work, we adopt the *template-fitting* approach to analyze the BOSS 2PCF and 3PCF. In the following section, we summarize the cosmological information accessible through *full-shape* analyses, and subsequently provide a detailed discussion of redshift-space distortions and the Alcock-Paczynski test.

1.3.1 Shape and distance information

While in our analysis we effectively marginalize over 'shape' parameters by putting strong CMB priors from Planck Collaboration: Aghanim et al. (2020), it is worthwhile to discuss their effect on 2-point statistics, as this forms the basis of our theoretical models.

The formation of the primordial distribution of fluctuations in cosmic fields, whose subsequent evolution gave rise to present-day cosmic structures, is due to a rapid phase of accelerated expansion that occurred in the early Universe, known as inflation. The absence of preferred scales predicted by standard models of inflation translates into a power-law primordial power spectrum $P_{\rm in}$,

$$P_{\rm in}(k) = A_s \left(\frac{k}{k_p}\right)^{n_s}, \tag{1.3.5}$$

where A_s and n_s are referred to as the *scalar amplitude* and *spectral index*, respectively. Current constraints from CMB are consistent with little scale dependence with $n_s = 0.9652 \pm 0.0042$ (Planck Collaboration: Aghanim et al. 2020).

As density perturbations evolve from inflation to the present epoch, their amplitudes are modified in a scale-dependent manner. Perturbations on a characteristic scale R lie in the *linear regime* if $\sigma_R^2 \ll 1$. At a generic epoch, the linear matter power spectrum can be written as

$$P_{\rm lin}(k, a) = D_1^2(a)\, T^2(k)\, P_{\rm in}(k), \tag{1.3.6}$$

where the time and scale dependence factorize into the *linear growth factor* $D_1(a)$ and the *transfer function* $T(k)$, respectively. While the growth factor can be derived analytically for a given cosmological model, the transfer function has a more complex dependence on the underlying physics and must be computed numerically by solving the Einstein-Boltzmann equations (Dodelson and Schmidt 2020).

In the primordial Universe, photons and baryons were tightly coupled and behaved as a single fluid. When the temperature dropped below approximately 3000 K, photons decoupled from baryons and began to stream freely. This event defines the last-scattering

surface, which we observe today through the Cosmic Microwave Background. At decoupling, sound waves propagating in the photon-baryon fluid ceased, imprinting a characteristic physical scale in the baryon distribution,

$$r_d = \int_{z_d}^{\infty} \frac{c_s(z)}{H(z)} \, \mathrm{d}z, \tag{1.3.7}$$

where $c_s(z)$ is the sound speed of the plasma and depends on its composition up to the drag redshift z_d. Since acoustic oscillations freeze out at decoupling, the scale $r_d \simeq 150\,\mathrm{Mpc}$ corresponds to a preferred separation with an excess of matter. This feature manifests itself as a peak in the two-point correlation function (2PCF) and as a series of oscillations, or 'wiggles', in the power spectrum. Known as *Baryon Acoustic Oscillations* (BAO), this feature provides a key observational pillar of the Big Bang model and acts as a standard ruler for cosmological distance measurements. Indeed, BAO measurements are a primary scientific driver of modern galaxy surveys such as BOSS (Eisenstein et al. 2011).

In Fig. 1.1 we illustrate how the matter two-point statistic varies as a function of the physical baryon density ω_b, at fixed $\omega_m = \omega_b + \omega_{cdm}$ (on the left), and of the cold dark matter density $\omega_{\rm cdm}$, in the limit $\omega_b \to 0$ (on the right). These predictions are computed using the `CosmoBolognaLib`[3] library (Marulli et al. 2016), which relies on the Boltzmann solver `CAMB`[4] for the calculation of transfer functions (Lewis et al. 2000).

For fixed values of A_s and n_s, the overall shape of the power spectrum is primarily determined by the physical densities ω_b and $\omega_{\rm cdm}$. The ratio $\omega_b/\omega_{\rm cdm}$ controls both the relative amplitude of the BAO wiggles and the suppression of power on small scales, with larger ratios enhancing these effects. The characteristic frequency of the BAO oscillations is set by the sound horizon r_d, which itself depends on these physical densities. In addition, the location of the turnover in the power spectrum and the degree of small-scale suppression depend on the total matter density ω_m and on r_d. In configuration space, BAO appear as a single peak at a comoving separation of $r \sim 105\,h^{-1}\mathrm{Mpc}$, although the precise position of this peak varies with the ratio $\omega_b/\omega_{\rm cdm}$.

The features described above correspond to well-defined comoving length scales for a given cosmological model and can therefore be used to constrain its parameters. In the context of ΛCDM, the physical densities of baryons and dark matter are fixed by the transfer function, leaving a single parameter, the Hubble constant H_0, or equivalently h, to determine the location of these features in the power spectrum.

[3] https://federicomarulli.github.io/CosmoBolognaLib/Doc/html/
[4] https://camb.info/

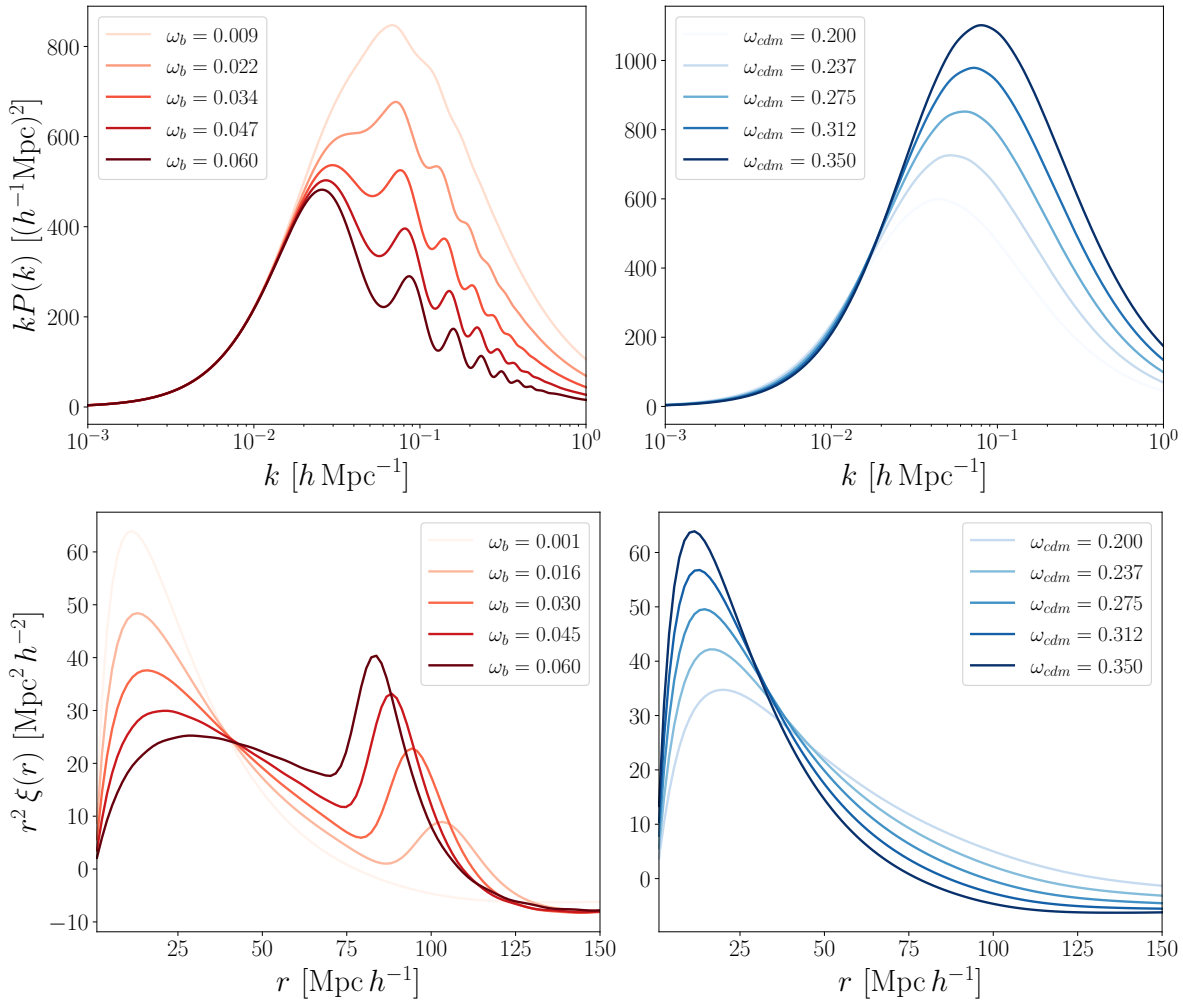

Figure 1.1: At the top, the linear matter power spectrum and at the bottom, the linear matter two-point correlation function. On the left, the ω_b parameter is varied and ω_{cdm} is adjusted to keep ω_m fixed; on the right $\omega_b \sim 0$ to illustrate shape modification exclusively due to ω_{cdm}. The curves are computed with the `CosmoBolognaLib` suite.

1.3.2 Redshift Space Distortions

The use of the observed redshift of a luminous source as a proxy for its comoving distance from the observer is responsible for the generation of anisotropic distortions in the distribution pattern of galaxies. This originates from the effect of galaxy peculiar velocities summing up to the velocity induced by the cosmic flow.

Peculiar velocities cannot be directly observed, so in the context of galaxy surveys, we only obtain information about the *redshift space* clustering of galaxies (or any other tracer of matter). This means that the observed comoving position $\boldsymbol{s}$ of an object is displaced with respect to its true position $\boldsymbol{x}$ as

$$\boldsymbol{s} = \boldsymbol{x} - f v_z(\boldsymbol{x})\hat{z} \tag{1.3.8}$$

where $v(\boldsymbol{x}) \equiv -u(\boldsymbol{x})/aHf$ and $u(\boldsymbol{x})$ is the peculiar velocity. The subscript z denotes the vector along the line of sight $\hat{z}$. In the definition of Eq. 1.3.8 we have assumed the *plane-parallel approximation*, for which all the considered separations are much smaller than the distance to the observer (Scoccimarro et al. 1999). In the case of two-point statistics, the angular dependence is characterised by the angle between the galaxy pair separation and the line of sight $\mu = \boldsymbol{k}\cdot\hat{z}/k$. While for the 2PCF a single variable μ is used to describe the anisotropy in the 3PCF, two variables are needed to describe the angle of the first pair with respect to the l.o.s and one for the second pair. Therefore, the redshift space 3PCF depends on 5 variables.

The density contrast in redshift space, $\delta_s(\boldsymbol{s})$, is obtained from the real-space density $\delta(\boldsymbol{x})$ by requiring that the redshift-space mapping conserves the number of galaxies

$$(1+\delta_s)d^3s = (1+\delta)d^3x \tag{1.3.9}$$

we can then take $d^3s = J(\boldsymbol{x})d^3x$ where the jacobian matrix $J(\boldsymbol{x}) = 1 - f\nabla_z u_z(\boldsymbol{x})$ encodes the exact mapping between the two spaces so that

$$\delta_s(\boldsymbol{s}) = \frac{\delta(\boldsymbol{x}) + 1 - J(\boldsymbol{x})}{J(\boldsymbol{x})} = \frac{\delta(\boldsymbol{x}) + f\nabla_z u_z(\boldsymbol{x})}{1 - f\nabla_z u_z(\boldsymbol{x})} \tag{1.3.10}$$

We are interested in the Fourier transform of the density contrast in redshift space:

$$\delta_s(\boldsymbol{k}) = \int \frac{d^3x}{(2\pi)^3} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} e^{ifk_z(\boldsymbol{x})}[\delta(\boldsymbol{x}) + f\nabla_z u_z(\boldsymbol{x})] \tag{1.3.11}$$

This equation describes the fully nonlinear density field in redshift space in the plane-parallel approximation. The standard approach involves taking a perturbative expansion of Eq. 1.3.11. If we limit the expansion to first order, we get the Kaiser result (Kaiser 1987). For the 2PCF, it reads

$$\xi_{\rm Kaiser}(s,\mu) = (b_1 + f\mu^2)^2 \xi_{\rm lin}(s) \tag{1.3.12}$$

where b_1 is the linear bias which encodes the contrast between galaxies and the underlying density field in the large-scale limit. The main effect introduced by RSD on the observed galaxy distribution is a squashing that can be explained by the fact that galaxies tend to fall towards regions of higher gravitational potential; objects further away are blueshifted and therefore look closer, while objects closer to the observer will look further away. This effect appears in the linear approximation of the 2-point correlation function with the factor $f\mu^2$. where f, the linear growth rate

$$f \propto \Omega_m^{\gamma}, \tag{1.3.13}$$

quantifies the growth of perturbations with respect to the growth in a *Einstein De Sitter* universe (f=1). Its proportionality to Ω_m can be quantified with $\gamma = 0.55 + 0.05(1 + w(z = 1))$ (Linder 2005), which depends on the equation of state parameter w in GR, but can have more complicated forms in Modified theories of gravity. Probing the f parameter can therefore quantify deviations from General Relativity.

Shown in Fig. 1.2 are the Legendre coefficients ξ_0 and ξ_2 obtained from the expansion

$$\xi_\ell(s) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu\, \mathcal{L}_\ell(\mu)\xi(s,\mu). \tag{1.3.14}$$

In the 2PCF monopole, f is quite degenerate with b_1, as it modifies the normalisation of ξ_0 in a scale-independent way. The degeneracy can be disentangled by introducing the quadrupole, the first term to encode the anisotropy of clustering, due to the angle dependence.

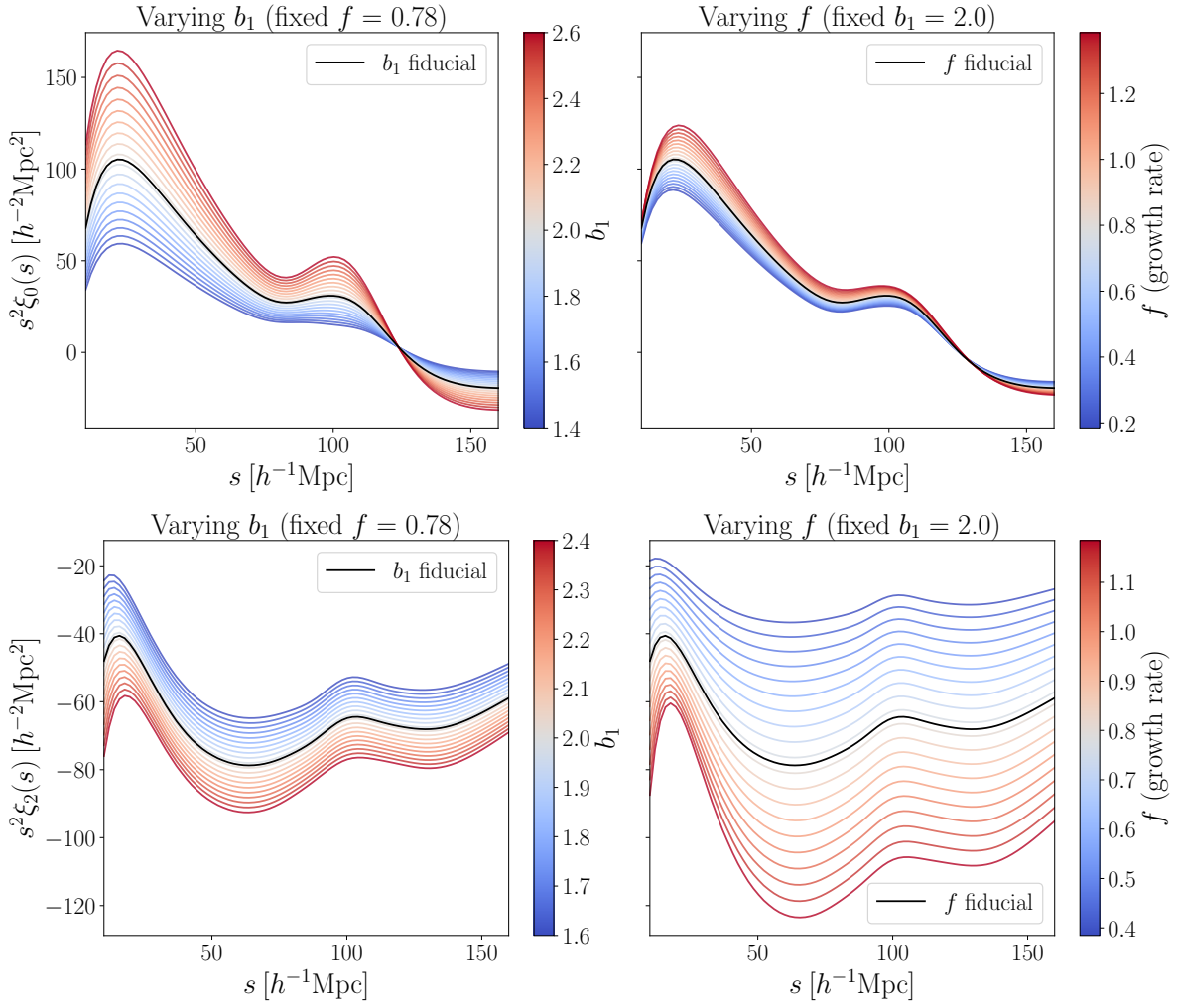

Figure 1.2: The monopole (top) and quadrupole (bottom) of the 2-point correlation function as a function of linear bias and growth rate $\{b_1,\ f\}$.

1.3.3 The Alcock-Paczyński test

Geometric distortions are introduced by the assumption of a fiducial cosmology, required both for the computation of comoving distance 1.3.1 and theoretical models (in the template fit approach). The so-called Alcock-Paczynski (AP) test (Alcock and Paczynski

1979) implements re-scaling parameters that account for the difference between the 'fiducial' and the 'true' underlying cosmology. The AP parameters are defined as

$$\alpha_{\parallel} = \frac{H_{\rm fid}(\bar{z})}{H(\bar{z})}\frac{r_{d,\rm fid}}{r_d} \quad , \quad \alpha_{\perp} = \frac{D_A(\bar{z})}{D_{A,\rm fid}(\bar{z})}\frac{r_{d,\rm fid}}{r_d} \tag{1.3.15}$$

where r_d is the scale of the sound horizon at decoupling time, $D_A(\bar{z}) = D_c(\bar{z})/(1+\bar{z})$ denotes the angular diameter distance and $H(\bar{z})$ is the Hubble expansion rate evaluated at the survey redshift $\bar{z}$. Quantities with and without subscript "fid" hereafter denote those in the fiducial cosmology and the true (unknown) cosmology, respectively.

In this thesis, we will follow the formalism for the multipole method based on the work of Padmanabhan and White (2008) where the AP parameters are defined by means of an isotropic shift α and an anisotropic warping ϵ

$$\alpha = \alpha_{\parallel}^{1/3}\alpha_{\perp}^{2/3}, \quad 1+\epsilon = \left(\frac{\alpha_{\parallel}}{\alpha_{\perp}}\right)^{1/3} \tag{1.3.16}$$

This corresponds to a shift of the coordinates parallel $r_{\parallel} = \alpha(1+\epsilon)^2 r_{\parallel,\rm fid}$ and perpendicular $r_{\perp} = \alpha(1+\epsilon)^{-1} r_{\perp,\rm fid}$ to the line of sight. Combining the two components results in

$$\begin{aligned} r =& \alpha\sqrt{(1+\epsilon)^4 r_{\parallel} + (1+\epsilon)^{-2} r_{\perp}} \\ \mu^2 =& \frac{\alpha^2(1+\epsilon)^4 r_{\parallel}}{\alpha^2(1+\epsilon)^4 r_{\parallel} + \alpha^2(1+\epsilon)^{-2} r_{\perp}} \end{aligned} \tag{1.3.17}$$

Note that if there is no isotropic shift, then $\alpha = 1$. Similarly, the lack of anisotropy implies $\epsilon = 0$. Under this parameterisation, the rescaling yields

$$\begin{aligned} \alpha &= \frac{D_V(\bar{z})}{D_{V,\rm fid}(\bar{z})}\frac{r_{d,\rm fid}}{r_d} = \left[\frac{D_A^2(\bar{z})}{D_{A,\rm fid}^2(\bar{z})}\frac{H_{\rm fid}(\bar{z})}{H(\bar{z})}\right]^{1/3}\frac{r_{d,\rm fid}}{r_d} \\ 1+\epsilon &= \left[\frac{D_{A,\rm fid}(\bar{z})}{D_A(\bar{z})}\frac{H_{\rm fid}(\bar{z})}{H(\bar{z})}\right]^{1/3} \end{aligned} \tag{1.3.18}$$

Where D_V is the so-called volume-averaged distance, constrained by the isotropic dilation parameter α.

As shown in Fig. 1.3, the monopole ξ_0 is sensitive to the "dilation" parameter, α, whose variation causes an isotropic shift of the BAO peak locations and a change in the power spectrum amplitude. On the other hand, the quadrupole ξ_2 is sensitive to the "warping" parameter ϵ, whose variation causes an anisotropic distortion in the redshift-space clustering (Kobayashi et al. 2020). The baryon acoustic oscillation (BAO) feature in the correlation functions of galaxies makes the AP effect very powerful to constrain the angular diameter distance and the Hubble expansion rate at the redshift of the survey.

Notice that using AP rescaling parameters in full-shape studies is not entirely correct, since when varying the linear power spectrum shape r_d changes too. In other words, it only makes sense to use AP parameters when r_d is fixed by a fiducial cosmology (Ivanov et al. 2020).

In the context of template fit studies, one does not need to generate these distortions on purpose; even if the fiducial cosmology exactly matches the true one, the distortions will be contained in the theoretical templates that are used to fit the data during MCMC sampling and hence the AP effect will still be effective (Philcox et al. 2020).

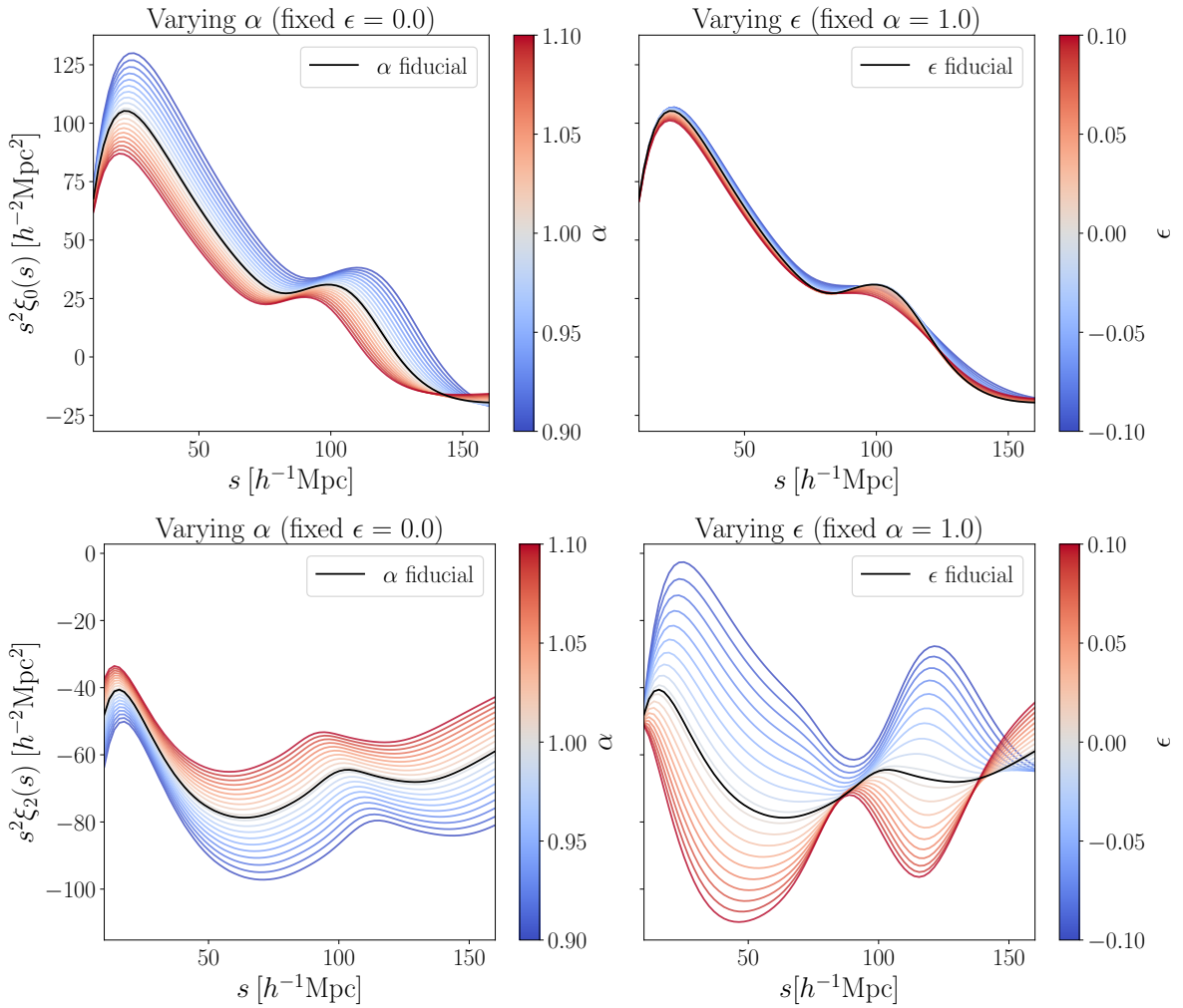

Figure 1.3: The monopole (top) and quadrupole (bottom) of the 2-point correlation function as a function of isotropic and anisotropic AP parameters $\{\alpha, \epsilon\}$.

1.4 Theoretical models of 2- and 3-point correlation functions

The theoretical models we adopt are based on cosmological perturbation theory (PT, Bernardeau et al. 2002). The usefulness of perturbation theory in interpreting results from galaxy surveys is based on the fact that in the gravitational instability scenario density fluctuations become small enough at large scales.

Solving equations of motion in configuration space is particularly hard as nearby regions of space are coupled. In Fourier space, however, wave vectors evolve independently from one another, in the linear approximation, and analytic results can be obtained. For this reason, N-point correlation function models are first treated in Fourier space and then transformed to configuration space. This is by far the most computationally demanding step for the modelling of the 3PCF, as the transformation from bispectrum to the 3PCF requires a 2-dimensional Hankel transform.

To take into account the non-linearity that arises from gravitational instability we consider an expression of the power spectrum up to *one-loop* in Eulerian perturbation theory, while the model for the bispectrum is at *tree-level*. The models we adopt implement a perturbative description of redshift space distortions as derived by Scoccimarro et al. (1999). Recent efforts to improve the modelling of clustering statistics in the mildly non-linear regime have achieved important results. Among these, we consider the full perturbative description of *galaxy bias* (explained in Sec. 1.4.2), and an accurate description of baryon acoustic oscillations by means of *infrared-resummation* (explained in Sec. 1.4.3).

Since we implement the *template fitting* approach, the linear power spectrum prediction is fixed by adopting the shape parameters of Planck Collaboration: Aghanim et al. (2020):

$$\{n_s = 0.966,\ A_s = 2.10 \times 10^{-9},\ \omega_{cdm} = 0.120,\ \omega_b = 0.022,\ h = 0.673\}$$

1.4.1 Standard Perturbation Theory

As previously mentioned, the formation of structures in the Universe is attributed to gravitational instability driven by initial perturbations generated by inflation.

To describe the dynamics of cosmic fields, such as the density contrast and peculiar velocity divergence (defined as $\theta(\boldsymbol{x}) \equiv \nabla \cdot \boldsymbol{u}(\boldsymbol{x})$), we use *Perturbation Theory* which allows us to describe the dynamics of these fields analytically. While extremely powerful, perturbation theory has its limitations, and its range of validity must fall within its assumptions. We can summarize the assumptions of PT as a set of considerations:

1. The perturbation theory approach is valid as long as fluctuations in the cosmic fields are small, i.e., the variance $\sigma_R^2 \ll 1$. This puts a scale limit on the validity of this approach at a given cosmic time. We expect perturbations to be small at very large scales, but as gravitational instability sets in, the same scale might be non-linear at a later time;

2. The only agent responsible for the change in cosmic fields is gravity, which acts on a matter component of collisionless nature we call Cold Dark Matter (CDM). This assumption is valid because CDM is the dominant component of matter and thus primarily responsible for gravitational instability;

3. CDM particles are non-relativistic so the equations of motion reduce to those of *Newtonian gravity*. In the context of the expanding Universe, the gravitational

potential is described by the Poisson equation

$$\nabla^2\Phi(\boldsymbol{x},\tau) = \frac{3}{2}\Omega_m\mathcal{H}^2(\tau)\delta(\boldsymbol{x},\tau); \tag{1.4.1}$$

where we have used the *conformal* time $d\tau = dt/a$ and the conformal Hubble parameter $\mathcal{H} = aH$.

4. The CDM particles number density in phase space is a function $f = f(\boldsymbol{x},\boldsymbol{p},\tau)$ and the phase space conservation implies the *Vlasov equation*

$$\frac{df}{d\tau} = \frac{\partial f}{\partial\tau} + \frac{\boldsymbol{p}}{ma}\cdot\nabla f - am\nabla\Phi\cdot\frac{\partial f}{\partial\boldsymbol{p}} = 0 \tag{1.4.2}$$

The full phase space distribution function f is truncated at first order (the 0th and 1st order moments are taken). By stating an equation of state ($p = w\rho c^2 \approx 0$), the hierarchy is closed, since the second order moment of f, the stress tensor $\sigma_{ij} \approx 0$.

Under these assumptions, the equations of motion correspond to the *continuity equation* and *Euler equation* in the expanding Universe. We state them here in Fourier space

$$\frac{\partial\tilde{\delta}(\boldsymbol{k},\tau)}{\partial\tau} + \tilde{\theta}(\boldsymbol{k},\tau) = -\int d^3\boldsymbol{k}_1 d^3\boldsymbol{k}_2\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12})\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2)\tilde{\theta}(\boldsymbol{k}_1,\tau)\tilde{\delta}(\boldsymbol{k}_2,\tau) \tag{1.4.3}$$

$$\begin{aligned}\frac{\partial\tilde{\theta}(\boldsymbol{k},\tau)}{\partial\tau} + \mathcal{H}(\tau)\tilde{\theta}(\boldsymbol{k},\tau) + \frac{3}{2}\Omega_m\mathcal{H}^2(\tau)\tilde{\delta}(\boldsymbol{k},\tau) = &-\int d^3\boldsymbol{k}_1 d^3\boldsymbol{k}_2\delta_D(\boldsymbol{k}-\boldsymbol{k}_{12}) \\ &\times\beta(\boldsymbol{k}_1,\boldsymbol{k}_2)\tilde{\theta}(\boldsymbol{k}_1,\tau)\tilde{\theta}(\boldsymbol{k}_2,\tau)\end{aligned} \tag{1.4.4}$$

The functions

$$\alpha(\boldsymbol{k}_1,\boldsymbol{k}_2) \equiv \frac{\boldsymbol{k}_{12}\cdot\boldsymbol{k}_1}{k_1^2}, \qquad \beta(\boldsymbol{k}_1,\boldsymbol{k}_2) \equiv \frac{k_{12}^2(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)}{2k_1^2k_2^2} \tag{1.4.5}$$

encode the non-linearity of the evolution (mode coupling) and come from the non-linear terms in the continuity and Euler equations.

To go beyond the linear regime, we retain the non-linear terms in the fluid equations (right-hand side of Eqs.1.4.3, 1.4.4) and solve perturbatively in Fourier space. The density contrast and velocity divergence are expanded as:

$$\begin{aligned}\tilde{\delta}(\mathbf{k},\tau) &= \sum_{n=1}^{\infty} D_1^n(\tau)\,\delta_n(\mathbf{k}), \\ \tilde{\theta}(\boldsymbol{k},\tau) &= -\mathcal{H}f\sum_{n=1}^{\infty} D_1^n(\tau)\,\theta_n(\mathbf{k})\end{aligned} \tag{1.4.6}$$

where D_1 is the fastest growing mode from the linear approximation and f is the linear growth rate defined in Eq. 1.3.2. Eq. 1.4.6 constitutes the main *ansatz* of perturbation theory as it supposes that it is possible to expand the density and velocity fields about their linear solutions, only valid when assumption number 1 is satisfied. Inserting Eqs. 1.4.6 in Eqs. 1.4.3 and 1.4.4 leads to coupled equations for δ_n and θ_n which are non-separable

in time due to the presence of f. However, this dependence vanishes for $f = \Omega_m^{1/2}$ which turns out ot be quite a good approximation. δ_n and θ_n are built from convolutions of n linear fields with mode-coupling kernels F_n and G_n:

$$\delta_n(\mathbf{k}) = \int_{\boldsymbol{q}_1} \cdots \int_{\boldsymbol{q}_n} \delta_D(\mathbf{k} - \mathbf{q}_{1\cdots n}) F_n(\mathbf{q}_1, \ldots, \mathbf{q}_n)\, \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n) \tag{1.4.7}$$

$$\theta_n(\mathbf{k}) = \int_{\boldsymbol{q}_1} \cdots \int_{\boldsymbol{q}_n} \delta_D(\mathbf{k} - \mathbf{q}_{1\cdots n}) G_n(\mathbf{q}_1, \ldots, \mathbf{q}_n)\, \delta_1(\mathbf{q}_1) \cdots \delta_1(\mathbf{q}_n) \tag{1.4.8}$$

where we have used a compact notation for the integrals $\int_{\boldsymbol{q}} \equiv \int d^3q/(2\pi)^3$. The recursion relations for the kernels F_n and G_n define the structure of Standard Perturbation Theory (SPT). They are constructed from the fundamental mode coupling functions of Eq. 1.4.5 according to recursion relations

$$\begin{aligned} F_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) &= \sum_{m=1}^{n-1} \frac{G_m(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m)}{(2n+3)(n-1)} [(2n+1)\alpha(\boldsymbol{k}_1, \boldsymbol{k}_2) F_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n) \\ &+ 2\beta(\boldsymbol{k}_1, \boldsymbol{k}_2) G_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n)] \\ G_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) &= \sum_{m=1}^{n-1} \frac{G_m(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_m)}{(2n+3)(n-1)} [3\alpha(\boldsymbol{k}_1, \boldsymbol{k}_2) F_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n) \\ &+ 2n\beta(\boldsymbol{k}_1, \boldsymbol{k}_2) G_{n-m}(\boldsymbol{q}_{m+1}, \ldots, \boldsymbol{q}_n)] \end{aligned} \tag{1.4.9}$$

where $\boldsymbol{k}_1 = \boldsymbol{q}_1 + \cdots + \boldsymbol{q}_m$ and $\boldsymbol{k}_2 = \boldsymbol{q}_{m+1} + \cdots + \boldsymbol{q}_n$ (Goroff et al. 1986 for a full derivation). Notice that for linear theory F_1, $G_1 = 1$ and the density contrast and divergence fields reduce to the linear approximations.

We have laid down the foundation of Standard Perturbation Theory (SPT), the simplest of the theories that aim to derive perturbative solutions for density contrast and peculiar velocity divergence fields in the nonlinear regime. These results can be used to describe the evolution of the statistical properties of the fields, such as the correlation functions. To illustrate this we compute, assuming a primordial Gaussian statistic, the power spectrum at first order (the tree-level power spectrum). To do this, we expand the overdensity field in standard perturbation theory to first order

$$\tilde{\delta}(\mathbf{k}, \tau) = D_1(\tau)\, \delta_1(\mathbf{k}) \tag{1.4.10}$$

given the definition of power spectrum (Eq. 1.2.6) we get

$$\begin{aligned} \left\langle \tilde{\delta}(\boldsymbol{k}, \tau) \tilde{\delta}(\boldsymbol{k'}, \boldsymbol{\tau}) \right\rangle &= \langle D_1(\tau) \delta_1(\boldsymbol{k}) D_1(\tau) \delta_1(\boldsymbol{k'}) \rangle \\ &= D_1^2(\tau)\, \langle \delta_1(\boldsymbol{k}) \delta_1(\boldsymbol{k'}) \rangle \\ &= (2\pi)^3 \delta_D(\boldsymbol{k} + \boldsymbol{k'}) D_1^2(\tau) P_{\rm lin}(k) \end{aligned} \tag{1.4.11}$$

Since in the linear regime wave modes evolve independently of each other, as time progresses, the power spectrum develops exclusively in amplitude, according to the rescaling $D_1^2(\tau)$. As for the nonlinear spectrum, keeping higher orders in δ_n leads to mode coupling: In addition to an increase in amplitude, there is a progressive increase in power

associated with large wave modes, and therefore small scales, which are more affected by nonlinearities. This is accompanied by the loss of Gaussian statistics. In the nonlinear regime, cosmological information is not completely contained in the variance (and therefore in the power spectrum or 2-point correlation function), but part of it shifts towards higher-order moments of the distribution.

In Appendix A, we introduce a diagrammatic representation of perturbation theory useful for the derivation of N-order correlation functions in Fourier space. We illustrate the *one-loop* power spectrum and *tree-level* bispectrum, which constitute the base of our theoretical models of 2-point and 3-point correlation functions.

Our theory models must take into account the redshift space distortions that are needed when taking observations of galaxy positions from spectroscopic surveys. Standard perturbation theory provides a way to write the density contrast in redshift space by making a perturbative expansion of Eq. 1.3.11 (Scoccimarro et al. 1999) where

- The mapping in redshift space is treated approximately by developing the exponential term in a power series.
- $\delta_s(\boldsymbol{k})$ and $\theta_s(\boldsymbol{k})$ are expanded in terms of linear fluctuations $\delta_1(k)$, as in Eq. 1.4.6.

Under these approximation we can write the n_{th} order redshift space density contrast as

$$\delta_{s,n}(\boldsymbol{k}) = \int_{\boldsymbol{q}_1} \cdots \int_{\boldsymbol{q}_n} \delta_D(\boldsymbol{k} - \boldsymbol{q}_{1,\ldots,n}) Z_n(\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n) \delta_1(\boldsymbol{q}_1) \ldots \delta_1(\boldsymbol{q}_n). \tag{1.4.12}$$

The redshift space mapping is encoded in the Z_n kernels. We can write their form at each order in perturbation theory by specifying a *bias model* described in the following section.

PT in redshift space breaks down at larger scales than in real space, because the redshift-space mapping is only treated approximately, through the expansion of the exponential in Eq. 1.3.11.

1.4.2 Galaxy Biasing

The objects we choose to trace the underlying density field are not able to perfectly reproduce its statistical properties, in other words, when we use galaxies to gain information about the density field we are introducing a bias. Bias describes, in a statistical sense, the relation of the distribution of these objects to that of matter. Galaxy formation is a complex process that unfolds over long timescales and may depend on a variety of physical mechanisms beyond gravity. As a result, the detailed connection between galaxies and the underlying matter distribution remains only partially understood.

There are two main approaches to construct a bias model: In the first approach (bottom-up) one starts with a model for how individual galaxies sit in the local small-scale mass density field, and then computes large-scale clustering by including the large-scale correlation of the relevant small-scale density feature. In the second approach (top-down) one starts from the fact that large-scale fluctuations are small and expands a completely unknown relation between galaxies and mass, with generally infinite freedom (except typically for the assumption of locality, relative to the scale of observations) into

a Taylor series in the density perturbations, where the coefficients of the first few terms in the series become the free parameters of the model. The models we adopt take this second, perturbative approach. A well-defined set of bias parameters determines the statistics of galaxies at a given order in perturbation theory. In this way, we effectively marginalize over the unknown details of the galaxy formation process, while robustly extracting cosmological information from galaxy surveys.

Early perturbative schemes have derived a bias model phenomenologically. The bias relation most often adopted in the literature is the nonlinear and local bias relation, in which the overdensity of the tracers is expressed in terms of a power series of the underlying matter (Coles 1993, Fry and Gaztanaga 1993)

$$\delta_g = \sum_{m=1}^{\infty} \frac{b_m}{m!}\delta^m = b_1\delta + \frac{b_2}{2}\delta^2 + \dots \tag{1.4.13}$$

By cutting the expansion to first order one finds (Kaiser 1984)

$$\delta_g(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}) \tag{1.4.14}$$

The linear bias parameter b_1 captures all our ignorance about how galaxies form from an overdensity. This relation is actually quite valid at large scales, but if one wants to recover a better description on more non-linear scales, other terms in the expansion should be kept.

The expansion of Eq. 1.4.13 does not take into account other terms that might develop in the formation of structures beyond the spherical collapse approximation. In fact, it has been proven that gravitational instability leads to the development of tidal fields (Baldauf et al. 2012).

More recently, a precise theory of galaxy bias has been obtained (for a review Desjacques et al. 2018) in which the bias model is derived from first principles of General Relativity, i.e, representing all possible local gravitational observables along the fluid trajectory. In other words, a full perturbative bias expansion should include all terms that present two spatial derivatives of the gravitational potential Φ, such as the density contrast and the tidal field. Higher derivative terms such as $\nabla^2\delta$ should also be included (McDonald and Roy 2009) even if they are sub-leading in the limit of long-wavelength density perturbations[5]. A complete bias expansion up to third order can be written as:

$$\begin{aligned}\delta_g(\boldsymbol{x}) =& b_1\delta(\boldsymbol{x}) + b_{\nabla^2\delta}\nabla^2\delta(\boldsymbol{x}) + \\ & \frac{b_2}{2}\delta^2(\boldsymbol{x}) + b_{\mathcal{G}_2}\mathcal{G}_2(\Phi_\nu|\boldsymbol{x}) + \\ & \frac{b_3}{6}\delta^3 + b_{\mathcal{G}_3}\mathcal{G}_3(\Phi_\nu|\boldsymbol{x}) + b_{\mathcal{G}_3}\mathcal{G}_2\delta(\boldsymbol{x}) + b_{\Gamma_3}\Gamma_3(\boldsymbol{x})\end{aligned} \tag{1.4.15}$$

where, in addition to the local bias operators δ, δ^2 and δ^3 we consider the *Galileon operators* G_2, G_3 and Γ_3 defined in terms of the density and velocity divergence potentials

[5]In Fourier space, the higher derivative term is $-k^2\tilde{\delta}(\boldsymbol{k})$ and is thus sub-leading in the large-scale limit.

Φ and Φ_ν, the latter defined as $\nabla^2\Phi_\nu \equiv \theta$:

$$\begin{aligned}\mathcal{G}_2(\Phi, \boldsymbol{x}) \equiv&[\partial_i\partial_j^2\Phi(\boldsymbol{x})]^2 - [\nabla^2\Phi(\boldsymbol{x})]^2;\\ \mathcal{G}_3(\Phi, \boldsymbol{x}) \equiv&[\nabla^2\Phi]^3 + 2\partial_i\partial_j\Phi\partial_j\partial_k\Phi\partial_k\partial_i\Phi - 3(\partial_i\partial_j\Phi)^2\nabla^2\Phi;\\ \Gamma_3(\boldsymbol{x}) \equiv&\mathcal{G}_2(\Phi, \boldsymbol{x}) - \mathcal{G}_2(\Phi_\nu, \boldsymbol{x}).\end{aligned}$$

This basis is complete up to cubic order, but in the context of LSS statistics, not all are needed to specify the power spectrum up to one-loop. The higher derivative term $\nabla^2\delta$, the third order local operator δ^3 and the terms $\mathcal{G}_2\delta, \mathcal{G}_3$ get renormalized into other parameters or their contributions produce a loop integral which vanishes and are therefore not part of the final model. The complete expansion reduces to

$$\delta_g(\boldsymbol{x}) = b_1\delta(\boldsymbol{x}) + \frac{b_2}{2}\delta^2(\boldsymbol{x}) + b_{\mathcal{G}_2}\mathcal{G}_2(\Phi_\nu|\boldsymbol{x}) + b_{\Gamma_3}\Gamma_3(\boldsymbol{x}). \tag{1.4.16}$$

Notice that we have neglected any stochastic contribution which needs to be taken into account given the randomness associated with our ignorance about the real location of galaxies (which depend on the unknown initial conditions). This contribution cancels out when transforming the Fourier space model to configuration space.

The free parameters of Eq. 1.4.16 are $\{b_1,\ b_2,\ b_{\mathcal{G}_2},\ b_{\Gamma_3}\}$ and can be summarized as follows:

- The linear bias term, b_1, quantifies the relationship between the galaxy density and the dark matter density at large scales. At leading order, it is the dominant contribution and simply rescales the matter density contrast by a constant factor.
- Moving to mildly non-linear scales, higher-order correlations of the density field appear, starting with a term proportional to δ^2, characterised by a quadratic local bias b_2. This factor is expected from a spherically symmetric gravitational collapse, in which higher powers of δ become more relevant at progressively smaller scales.
- Even starting from the local bias relation of Eq. 1.4.13, gravitational evolution is expected to generate large-scale tidal fields (Baldauf et al. 2012). At leading order, the corrections given by the tidal stress tensor are represented by the second-order Galileon operator $\mathcal{G}_2$.
- The next-to-leading order correction to the tidal field can be obtained by considering terms up to second order in the displacement field (given by the difference between density contrast and velocity divergence fields). This contribution generates an additional operator which appears at third order in PT and is denoted with Γ_3 (for cubic bias in simulations: Abidi and Baldauf 2018).

1.4.3 Infrared Resummation

The amplitude of BAO wiggles is sensitive to large-scale bulk flows (*infrared* modes), whose main effect is a smearing of the BAO ring. Standard perturbation theory is unable to properly model this effect, and it has been addressed in the literature with a number of

phenomenological methods. More recently, this problem has been resolved in a precise way under various beyond perturbation theory approaches, such as the Effective Field Theory (EFT) (Senatore and Zaldarriaga 2015) and Time Sliced Perturbation Theory (TSPT) (Blas et al. 2016). These schemes are able to predict the smearing of the BAO peak to very high precision by effectively resumming all infrared modes $q < k_S$. Remarkably, infrared resummation introduces no extra fitting parameters.

In practical terms, the linear power spectrum is expressed as the sum of a smooth P_{nw} and wiggly P_w component, such that

$$P_{\rm lin}(k) = P_{nw}(k) + P_w(k) \tag{1.4.17}$$

The smooth component consists of a linear power spectrum with the BAO feature removed ("no-wiggle" power spectrum), $P_{\rm nw}(k)$. A damping factor is introduced to absorb the washing out of the BAO feature (Ivanov and Sibiryakov 2018), in redshift space, it amounts to

$$\Sigma^2_{\rm tot}(\mu) = \Sigma^2 + \mu^2[2f\Sigma^2 + f^2(\Sigma^2 - \delta\Sigma^2)] + \mu^4 f^2 \delta\Sigma^2 \tag{1.4.18}$$

where

$$\begin{aligned} \Sigma^2 =& \frac{1}{6\pi^2}\int_0^{k_s} dq\, P_{nw}(q)\left[1 - j_0\left(\frac{q}{k_{osc}}\right) + 2j_2\left(\frac{q}{k_{osc}}\right)\right] \\ \delta\Sigma^2 =& \frac{1}{2\pi^2}\int_0^{k_s} dq\, P_{nw}(q) j_2\left(\frac{q}{k_{osc}}\right) \end{aligned} \tag{1.4.19}$$

are the BAO damping functions, $j_\ell(x)$ are spherical Bessel functions evaluated at $k_{osc} = 1/\ell_{osc}$, i.e., the wave number associated with the BAO scale $\ell_{osc} = 110h^{-1}$Mpc. k_s is the separation scale defining IR modes that need to be resummed, we choose $k_s = 0.2\,h$Mpc^{-1} following Ivanov et al. (2022).

At Leading Order (LO) the linear power spectrum is replaced by

$$P_{\rm lin}(k) \to P^{\text{IR-res}}_{\rm LO}(k,\mu) = P_{nw}(k) + P_w(k) e^{-k^2\Sigma^2_{\rm tot}(\mu)} \tag{1.4.20}$$

This means modelling of clustering statistics at tree-level will need to be evaluated with the following form. Since we have a one-loop model for the power spectrum, the next-to-leading order contribution of the resummed spectrum is required. This leads to an extra term appearing in the power spectrum at tree level

$$P_{\rm lin}(k) \to P^{\text{IR-res}}_{\rm NLO}(k,\mu) = P_{nw}(k) + e^{-k^2\Sigma^2_{\rm tot}} P_w(k)(1 + k^2\Sigma^2_{\rm tot}(\mu)) \tag{1.4.21}$$

and the one-loop integrals need to be evaluated with $P^{\text{IR-res}}_{\rm LO}(k,\mu)$.

1.4.4 Configuration space modelling

The elements described in the previous sections are assembled in the following manner. The *up to one-loop redshift space galaxy power spectrum* can be written as

$$P^s_g(k,\mu) = P^{s,tree}_g(k,\mu) + P^{s,1-loop}_g(k,\mu). \tag{1.4.22}$$

Expanding this form gives

$$\begin{aligned} P_g^{s,\mathrm{tree}}(k,\mu) =& Z_1^2(\mu) P_{\mathrm{lin}}(k) \\ P_g^{s,1-loop}(k,\mu) =& P_{g,22}^{s,1-loop}(k,\mu) + P_{g,13}^{s,1-loop}(k,\mu) \end{aligned} \tag{1.4.23}$$

where the loop corrections are

$$\begin{aligned} P_{g,22}^{s,1-loop}(k,\mu) =& 2\int_{\boldsymbol{q}} Z_2^2(\boldsymbol{q}, \boldsymbol{k}-\boldsymbol{q}) P_{\mathrm{lin}}(q) P_{\mathrm{lin}}(|\boldsymbol{k}-\boldsymbol{q}|) \\ P_{g,13}^{s,1-loop}(k,\mu) =& 6 Z_1(\mu) P_{\mathrm{lin}}(k) \int_{\boldsymbol{q}} Z_3(\boldsymbol{q}, -q, \boldsymbol{k}) P_{\mathrm{lin}}(q) \end{aligned} \tag{1.4.24}$$

We used a compact notation for the integral $\int_{\boldsymbol{q}} \equiv \int d^3q/(2\pi)^3$. The $\mathbf{Z}$ kernels encode the redshift space mapping and the bias model as in Eq. 1.4.12. Their final form is

$$\begin{aligned} &Z_1(\boldsymbol{k}_1) = (b_1 + f\mu_1^2) \\ &Z_2(\boldsymbol{k}_1, \boldsymbol{k}_2) = b_1 F_2(k_1, k_2) + f\mu^2 G_2(k_1, k_2) + \frac{f\mu k}{2}\left[\frac{\mu_1}{k_1}(b_1 + f\mu_2^2) + \frac{\mu_2}{k_2}(b_1 + f\mu_1^2)\right] \\ &+ \frac{b_2}{2} + b_{\mathcal{G}_2}\left(\frac{(\boldsymbol{k}_1\cdot\boldsymbol{k}_2)^2}{k_1^2 k_2^2} - 1\right) \\ &Z_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = b_1 F_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + f\mu^2 G_3(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) + \ldots \end{aligned} \tag{1.4.25}$$

where k and μ are the amplitude and angle with respect to the sum of $\boldsymbol{k}_i$.

The IR-resummation substitutes the linear power spectrum prediction in both the tree-level and 1-loop contributions. The IR-resummed redshift space galaxy power spectrum up to one loop is

$$P_{g,\mathrm{NLO}}^{s}(k,\mu) = P_g^{s,tree}[P_{\mathrm{NLO}}^{\mathrm{IR-res}}(k,\mu)] + P_g^{s,1-loop}[P_{\mathrm{LO}}^{\mathrm{IR-res}}(k,\mu)] \tag{1.4.26}$$

where the square brackets indicate that the tree-level and loop terms need to be evaluated by substituting each linear power spectrum with its resummed version.

To account for the apparent geometrical distortion in the observed pattern of galaxy clustering, the so-called Alcock-Paczynski (AP) parameters are integrated in the power spectrum model

$$P_{g,\mathrm{NLO}}^{s}(k_{\mathrm{fid}}, \mu_{\mathrm{fid}}) = \frac{1}{\alpha^3} P_{g,\mathrm{NLO}}^{s}(k,\mu) \tag{1.4.27}$$

where the relation between true and observed (fiducial) coordinates and angles were given in Eq. 1.3.17 in configuration space but can be equivalently derived for the Fourier space wave vectors.

The *tree-level redshift space galaxy bispectrum* is

$$B_g^{s,tree}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = 2 Z_2(\boldsymbol{k}_1, \boldsymbol{k}_2) Z_1(\boldsymbol{k}_1) Z_1(\boldsymbol{k}_2) P_{\mathrm{lin}}(k_1) P_{\mathrm{lin}}(k_2) + \mathrm{cycl.} \tag{1.4.28}$$

where the cyclic permutation is made over all pairs of wavevectors $\boldsymbol{k}_i$. The IR-resummed, redshift space galaxy bispectrum at tree-level is obtained by substituting the linear power spectrum with its resummed version at leading order

$$B_{g,\mathrm{LO}}^{s}(\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3) = 2 Z_2(\boldsymbol{k}_1, \boldsymbol{k}_2) Z_1(\boldsymbol{k}_1) Z_1(\boldsymbol{k}_2) P_{\mathrm{LO}}^{\mathrm{IR-res}}(k_1, \mu_1) P_{\mathrm{LO}}^{\mathrm{IR-res}}(k_2, \mu_2) + \mathrm{cycl.} \tag{1.4.29}$$

To simplify the notation, we drop the g, s, and NLO subscripts and superscripts. To compute the configuration space 2PCF model, the power spectrum model is expanded in Legendre coefficients,

$$P_\ell(k) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu\, \mathcal{L}_\ell(\mu)\, P(k,\mu), \tag{1.4.30}$$

where $\mathcal{L}_\ell(\mu)$ are the Legendre polynomials that encode the angular dependence of the anisotropic power spectrum. The power spectrum multipoles are related to those of the 2PCF through

$$\xi_\ell(s) = \frac{i^\ell}{2\pi^2}\int_0^\infty dk\, k^2 P_\ell(k) j_\ell(ks). \tag{1.4.31}$$

In practice, the one-dimensional Hankel transform for the Legendre coefficients with $\ell = 0, 2, 4$ is computed using a 1D-FFTLog method (Hamilton 2000).

The 1D-FFTLog algorithm has been widely used to efficiently evaluate Fourier transforms involving logarithmic variables and single Bessel integrals of the form $\int_0^\infty dk\, f(k) j_\ell(kr)$, where $f(k)$ is a generic smooth function. The central idea is to expand the function as

$$f(k) = \sum_m c_m k^{z_m},$$

where z_m are, in general, complex numbers. By computing the sum over the Fourier coefficients c_m, the algorithm avoids the need to directly evaluate the one-dimensional integral.

The bispectrum in redshift space depends on two angular variables defined with respect to the line of sight. The first is the angle between the line of sight $\hat{z}$ and $\boldsymbol{k}_1$, denoted by θ, and the second is the azimuthal angle of $\boldsymbol{k}_2$ around $\boldsymbol{k}_1$, denoted by φ. Consequently, the redshift-space bispectrum depends on five variables in total. We expand it in spherical harmonics $Y_{\ell m}(\theta,\varphi)$,

$$\begin{aligned} B(k_1,k_2,k_3,\theta,\varphi) &= \sum_{\ell,m} B_{\ell,m}(k_1,k_2,k_3) Y_{\ell m}(\theta,\varphi), \\ B_{\ell,m}(k_1,k_2,k_3) &= \frac{2\ell+1}{2}\int_0^{2\pi} d\varphi \int_{-1}^{1} d(\cos\theta)\, Y^*_{\ell m}(\theta,\varphi)\, B(k_1,k_2,k_3,\theta,\varphi). \end{aligned} \tag{1.4.32}$$

In this work, we retain only the $\ell = 0$, $m = 0$ coefficient, effectively focusing on the isotropic component of the three-point correlation function. Equation (1.4.32) therefore corresponds to a spherical average of the full anisotropic bispectrum. This step represents the most computationally demanding part of the bispectrum evaluation, as it requires a two-dimensional integration over the full sphere.

The bispectrum monopole depends only on three variables: k_1, k_2, and μ_{12} (or, equivalently, k_3), which are related by the law of cosines, $k_3^2 = k_1^2 + k_2^2 + 2k_1k_2\mu_{12}$. This allows the bispectrum monopole to be expanded in Legendre polynomials,

$$B_\ell(k_1,k_2) = \frac{2\ell+1}{2}\int_{-1}^{1} d\mu_{12}\, \mathcal{L}_\ell(\mu_{12})\, B_{0,0}(k_1,k_2,\mu_{12}). \tag{1.4.33}$$

The inner bispectrum multipoles $B_\ell(k_1, k_2)$ are therefore two-dimensional functions depending only on k_1 and k_2. These multipoles play a key role in computing the galaxy 3PCF in configuration space, since the isotropic (connected) 3PCF corresponds to the inverse Fourier transform of the redshift-space bispectrum monopole. The Legendre coefficients of the bispectrum are related to those of the 3PCF through (Szapudi 2004)

$$\zeta_\ell(r_{12}, r_{13}) = (-1)^\ell \int \frac{dk_1}{2\pi^2} \frac{dk_2}{2\pi^2} \, k_1^2 k_2^2 B_\ell(k_1, k_2) j_\ell(k_1 r_{12}) j_\ell(k_2 r_{13}), \tag{1.4.34}$$

where r_{12} and r_{13} denote the side lengths of a triangle whose vertices are labelled $1, 2, 3$. Evaluating this integral constitutes the main computational burden in the 3PCF modelling procedure. When computed on a grid of size $N_r \times N_r$ using standard quadrature methods, the cost scales as $\mathcal{O}(N_r^2 N_k^2)$, where N_r and N_k denote the number of sampled r bins and k modes, respectively. In addition, the rapid and oscillatory behaviour of the Bessel functions requires a large number of integration steps, making an accurate evaluation of the integrals computationally demanding.

To compute Eq. 1.4.34 in an efficient way, a two-dimensional *FFTLog* method (Fang et al. 2020) provides the best solution. It relies on the same idea as the 1D-FFTLog, but uses a double power-law expansion to accurately evaluate integrals involving the product of two Bessel functions. Crucially, the 2D-FFTLog method computational cost scales as $\mathcal{O}(N^2 \log N)$.

In the choice of number of grid points for the numerical integration we follow Guidi et al. (2023) and use a fixed grid of dimension $256 \times 256 \times 51$, where 256 are the wavenumbers along the k_1 and k_2 directions and consider 51 angle values sampled using the Gauss-Legendre quadrature method. We use $k_{min} = 3 \times 10^{-4}\, hMpc^{-1}$ and $k_{max} = 5\, hMpc^{-1}$, to avoid aliasing and to damp nonlinear contributions to the power spectra from scales in which the perturbative approach fails. For better numerical convergence and shortcomings in the perturbative modelling of short scales, we dampen the bispectrum multipoles

$$B_{damp}(k_1, k_2, k_3) = B(k_1, k_2, k_3)\, e^{-(k_1^2+k_2^2+k_3^2)/k_0^2} \tag{1.4.35}$$

Where for the damping scale we choose a value of $k_0 = 0.5\, hMpc^{-1}$; smaller than that advised in Guidi et al. (2023) because we expect perturbative modelling in redshift space to break down at larger scales than in real space.

Templates in configuration space

Ultimately, we extract the cosmological parameters of interest using a template-fitting approach. In this framework, the fitting parameters enter the model linearly and can therefore be factored out from the remaining theoretical contributions. The multipoles of the IR-resummed, redshift-space galaxy two-point correlation function are written as

$$\xi_\ell(s) = \sum_{p=1}^{25} X^{(p)} \xi_\ell^{(p)}(s) \tag{1.4.36}$$

where the coefficients $X^{(p)}$ are linear combinations of the fitting parameters, and the functions $\xi_\ell^{(p)}(s)$ are precomputed templates. The coefficients have 25 independent terms

$$\begin{aligned} X^{(p)} =&\{b_1^2\sigma_8^2, b1f\sigma_8^2, f^2\sigma_8^2, b_1^2\sigma_8^4, b1f\sigma_8^4, f^2\sigma_8^4, \\ &b_2^2\sigma_8^4, b_{\mathcal{G}_2}^2\sigma_8^4, b_2b_{\mathcal{G}_2}\sigma_8^4, f^4\sigma_8^4, b_1b_{\mathcal{G}_2}\sigma_8^4, f^3\sigma_8^4, b_1b_2\sigma_8^4, \\ &b_1f^2\sigma_8^4, b_1f^3\sigma_8^4, b_{\mathcal{G}_2}f^2\sigma_8^4, b_{\mathcal{G}_2}f\sigma_8^4, b_1^2f^2\sigma_8^4, b_2f^2\sigma_8^4, \\ &b_2f\sigma_8^4, b_1^2f\sigma_8^4, b_1b_2f\sigma_8^4, b_1b_{\mathcal{G}_2}f\sigma_8^4, b_1b_{\Gamma_3}\sigma_8^4, b_{\Gamma_3}f\sigma_8^4\} \end{aligned} \tag{1.4.37}$$

A factor of σ_8 comes for every instance of the linear density field, thus the leading order contribution of the 2-point correlation function contains two instances of the linear density field and is thus multiplied by a normalisation factor of σ_8^2 (first three terms in Eq. 1.4.37). Every contribution that appears at the next-to-leading order in perturbation theory should instead contain four instances of the density field and is thus multiplied by a factor σ_8^4. The set of independent fitting parameters entering the two-point correlation function model is

$$\{\, b_1,\ b_2,\ b_{\mathcal{G}_2},\ b_{\Gamma_3},\ f,\ \sigma_8,\ \alpha,\ \epsilon\,\}. \tag{1.4.38}$$

The first four parameters describe galaxy bias up to cubic order in perturbation theory, as required for consistency with the one-loop power spectrum modelling. The remaining four parameters are cosmology dependent: the growth rate f enters through the modelling of redshift-space distortions, while α and ϵ parametrize the geometric distortions.

For the 3PCF we have 14 (p) terms of $\zeta_\ell^{(p)}$ for ℓ up to 10. In the main analysis we don't directly fit the Legendre coefficients but instead re-sum them with the following equation

$$\zeta^{(p)}(r_{12}, r_{13}, \cos\theta) = \sum_{\ell=0}^{10} \zeta_\ell^{(p)}(r_{12}, r_{13})\mathcal{L}_\ell(\cos\theta) \tag{1.4.39}$$

The fitting parameters are then fed to

$$\zeta(r_{12}, r_{13}, \cos\theta) = \sum_{p=1}^{14} X^{(p)}\zeta^{(p)}(r_{12}, r_{13}, \cos\theta) \tag{1.4.40}$$

where $X^{(p)}$ is

$$\begin{aligned} X^{(p)} = \{&b_1^3, b_1b_2, b_1b_{\mathcal{G}_2}, b_1fb_2, b_1fb_{\mathcal{G}_2}, b_2f^2, b_{\mathcal{G}_2}f^2, \\ &b_1^2f, b_1^3f, b_1^2f^2, b_1f^2, b_1f^3, f^3, f^4\} \end{aligned}$$

All 3PCF terms are multiplied by an overall factor of σ_8^4, since three-point statistics involve four powers of the density field at leading order in perturbation theory. The independent parameters entering the 3PCF model are therefore

$$\{b_1,\ b_2,\ b_{\mathcal{G}_2},\ f,\ \sigma_8\}. \tag{1.4.41}$$

The bias expansion is truncated at second order, as required for the tree-level bispectrum. We do not exploit geometric distortions as a probe of cosmology in the 3PCF but only

in the 2PCF. Finally, predictions for the reduced 3PCF are obtained by combining the 2PCF monopole ξ_0 with the connected 3PCF using Eq. 1.2.15.

A scheme of the modelling procedure is shown in Fig. 1.4.

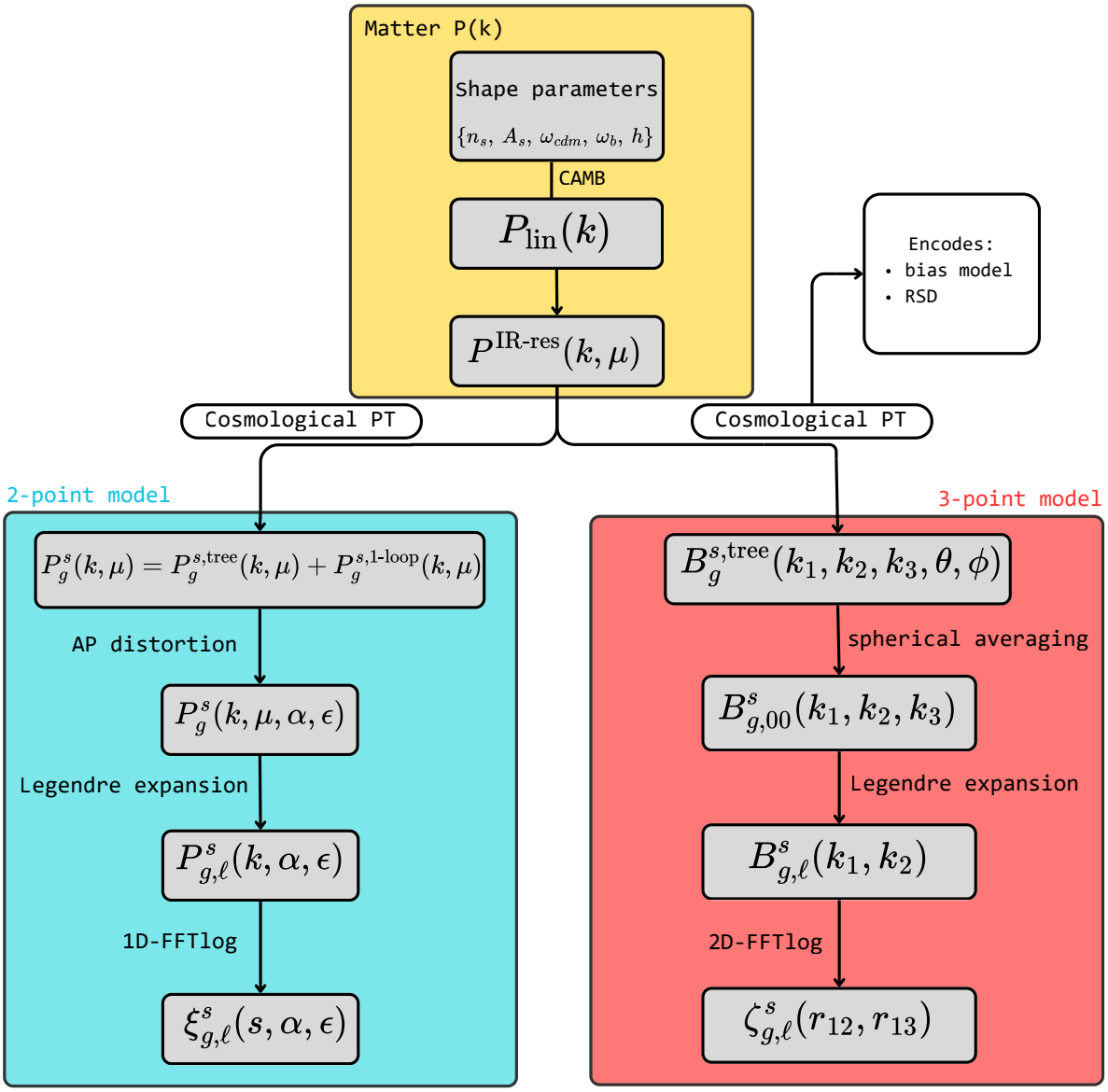

Figure 1.4: Schematic outline of the theoretical prediction of observed galaxy 2- and 3-point statistics. Starting from the computation of the matter power spectrum, with shape parameters fixed by a flat ΛCDM model with precise values of Planck Collaboration: Aghanim et al. (2020).

1.5 Open questions in cosmology

Despite the effectiveness of the ΛCDM model in explaining several results from cosmological observations, there are unexplained phenomena and tensions that need to be addressed. We give a summary of the most relevant ones regarding this work:

Nature of Dark Energy. Dark energy is responsible for the acceleration of the Universe as a cosmological constant, but a physical explanation of this phenomenon is still lacking. It can be interpreted as the vacuum energy, which should be non-zero in the context of Quantum Field Theory (QFT); however, this has led to one of the worst predictions in history, since values predicted by cosmology deviate from those in QFT by between 50 and 120 orders of magnitude (Zel'Dovich 1967). Perhaps, the cosmological constant is not a physical component at all and instead is simply a property of space-time under GR (Bianchi and Rovelli 2010) or it might be that GR on large scales loses its descriptive power and we need to resort to modified theories of gravity (for a review Shankaranarayanan and Johnson 2022). Recent observational evidence from DESI (DESI Collaboration: Abdul Karim et al. 2025) suggests that dark energy might not be constant at all, but a changing dynamical process. With the upcoming observational data from stage IV surveys, we will rule out many of the theories surrounding dark energy and whether it is constant or not.

H_0 Tension. This tension arises from a discrepancy between the values of the current expansion rate of the Universe, the Hubble constant H_0, estimated either indirectly from CMB observables and BAO data (standard ruler), or from local observables such as Type Ia SNe (standard candles). The best estimate from CMB data comes from the *Planck* collaboration (Planck Collaboration: Aghanim et al. 2020) with a measurement of $H_0 = (67.36 \pm 0.54)\,\mathrm{km\,s^{-1}Mpc^{-1}}$ at the 68% confidence level. The second method relies on the calibration of Type Ia Supernovae (SNIa) through the observation of Cepheids. The most precise result comes from the *SH0ES* collaboration, with a measurement of $H_0 = (73.29 \pm 0.90)\,\mathrm{km\,s^{-1}Mpc^{-1}}$ (Murakami et al. 2023), in 5.6σ discrepancy with Planck result.

S_8 Tension. The S_8 parameter is defined as $S_8 \equiv \sqrt{\Omega_m/0.3}\sigma_8$. The tension arises between direct measurements of S_8 and the analysis of CMB data assuming ΛCDM. For example, the value obtained by *Planck* (Planck Collaboration: Aghanim et al. 2020), $S_8 = 0.834 \pm 0.016$, is in 2.9σ discrepancy with the directly measured values from weak lensing surveys, in particular the *Kilo Degree Survey* (KiDS) (Heymans, Catherine et al. 2021) and the *Dark Energy Survey* (DES) (Abbott et al. 2022), which give $S_8 = 0.769 \pm 0.016$ after combining the two.

While these are the more relevant topics for our study, there are many more questions to answer regarding a complete description of cosmology, namely, what is the nature of dark matter? Is the *cosmological principle* a solid assumption? Can we detect primordial gravitational waves? Can we detect primordial non-Gaussianity? How well can we detect the mass of neutrinos? While some may be harder than others to answer, exploiting the full information content of the LSS of the observable Universe will be of great help.

1.5.1 The importance of higher-order statistics

In Sec. 1.3 we outlined the two main approaches for the exploration of the cosmological parameter space and discussed the primary sources of information in each case. One may ask what the purpose of going beyond 2-point statistics is, since it seems the main source of cosmological information is contained within the power spectrum or the 2PCF.

On sufficiently large scales, which have evolved linearly since the beginning of the Universe, the matter density field is well described by the 2PCF or power spectrum. However, gravitational evolution is inherently non-linear and proceeds in a bottom-up fashion. Mode coupling is induced as structures form, transferring information from two-point statistics to higher-order correlations. Even if the primordial density field were Gaussian, the late-time matter distribution becomes increasingly non-Gaussian toward smaller scales and lower redshifts, naturally generating higher-order statistics.

Although two-point statistics typically have the highest signal-to-noise, the amount of information available in higher-order statistics grows rapidly as one probes a wider range of scales. In particular, the number of distinct triangular configurations entering three-point statistics increases much faster than the number of pairs contributing to the 2PCF, allowing additional information to be extracted from the distribution of galaxies.

Furthermore, at all scales, the galaxy distribution exhibits higher-order correlations because galaxies are biased tracers of the underlying matter field. This biasing is encoded in non-linear relations between galaxy and matter densities. Unlike the 2PCF, which is sensitive only to scale information (i.e., pair separations), the 3PCF also probes shape information through the geometry of galaxy triplets. This additional dependence leads to a different sensitivity to bias parameters, making the 3PCF highly complementary to the 2PCF and enabling degeneracies among bias parameters to be broken. From the perspective of perturbative modelling, this difference is particularly clear. At leading order, the galaxy 2PCF depends only on the linear bias parameter b_1. Higher-order correlators such as the galaxy 3PCF depend already at leading order on both local and non-local quadratic bias parameters, b_2 and $b_{\mathcal{G}_2}$. As a result, these parameters can be constrained with significantly greater accuracy when higher-order statistics are taken into account.

By lifting the degeneracies inherent in the biasing relation at the two-point level, measurements of the 3PCF or bispectrum provide access to complementary information and improve cosmological constraints. In particular, they help to break the degeneracy between the growth rate f and the amplitude of matter fluctuations σ_8 (Veropalumbo et al. 2021).

Chapter 2

Dataset and Measurements

In this chapter we review the estimators used to measure correlation functions and discuss recent improvements that have significantly reduced the computational cost of their evaluation. In particular, the Spherical Harmonics Decomposition (SHD) algorithm (Slepian and Eisenstein 2015) reduces the computational complexity of measuring the isotropic 3PCF from $\mathcal{O}(N^3)$ to $\mathcal{O}(N^2)$ (Sec. 2.1). In Sec. 2.2, we introduce the BOSS data and the large-scale structure (LSS) catalogue constructed from it for the measurement of clustering statistics. We also provide a brief overview of the main scientific goals of BOSS and how they have been achieved. Finally, we describe measurements of the 2PCF and 3PCF from the BOSS dataset in Sec. 2.3, and from a set of mock catalogues, we estimate the covariance matrices associated with each estimator in Sec. 2.4.

2.1 Clustering estimators

One of the main challenges in the analysis of Large Scale Structure (LSS) data is obtaining a reliable method for the estimation of clustering statistics. Fourier space correlation functions have been widely used for clustering analysis; the efficient estimators that have been developed (based on FKP estimators Feldman et al. 1993) make them reliable and fast to obtain. One downside to their use on real data is the need to account for the survey geometry. Configuration space correlation functions have the advantage that complex survey geometries can be easily considered in the estimator by introducing edge corrections. However, estimators in configuration space, particularly for the 3-point correlation function (3PCF), have historically been slower and computationally more intensive to use in a cosmological analysis. Recently, there have been substantial efforts to make configuration space statistics faster to measure. The Spherical harmonics decomposition method (SHD) developed by Slepian and Eisenstein (2015) has reduced the computational cost of three-point estimation down to $\mathcal{O}(N^2)$. We briefly review the progress made in the last century for configuration space correlation functions estimators.

Given the nature of correlation functions in configuration space, one might expect that to measure them is to do a counting exercise. After all, the 2-point correlation function is a measure of the excess (or defect) of pairs of objects with respect to a random distribution. So, by definition, to measure the galaxy 2PCF, one has to measure all data

pair counts and compare them with random pair counts at a given distance r (Peebles and Groth 1975):

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1 \tag{2.1.1}$$

Where $DD(r)$ is the number of pairs of galaxies at distance r, and $RR(r)$ is the expected number of pairs at distance r for a random catalogue with the same mean density and sampling geometry of the data catalogue. To obtain a random galaxy distribution, we must resort to numerical simulations that reproduce the survey's footprint and take into account all the systematics that might be involved when taking real spectroscopic data. Landy and Szalay (1993) demonstrated that this estimator is suboptimal, its variance is at order $\mathcal{O}(1/n)$ which is not close to the optimal Poisson variance of $\mathcal{O}(1/n^2)$[1]. A better estimate is given by:

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)} \tag{2.1.2}$$

where the DR pairs are cross-correlating galaxies from the data and a random catalogue. This estimator improves the variance down to the Poisson limit and requires no additional computational work. Supposedly, this estimator can be extended to higher order correlation functions (Szapudi and Szalay 1998); the general form of an Nth order correlation function is

$$\xi_N = \frac{(D_1 - R_1)(D_2 - R_2)\dots(D_N - R_N)}{R_1 \dots R_N} \tag{2.1.3}$$

Therefore, the unbiased estimator for the connected three-point function writes

$$\zeta = \frac{DDD - 3DDR + 3DRR - RRR}{RRR}. \tag{2.1.4}$$

Data provided by spectroscopic surveys, such as BOSS, allow us to measure the three-dimensional position of galaxies (or other tracers of the LSS). The 3-point correlation function depends on 9 degrees of freedom (3 triangle sides and 6 angles, 2 angles for each side) needed to specify the position of a triplet of objects:

- Three are constrained by the property of spatial homogeneity. In other words, triangles with the same geometry but placed in different positions contribute identically to the 3-point statistics.
- Three are constrained by the property of statistical isotropy. This implies that triangles geometrically identical but with different spatial orientations provide the same contribution to the 3PCF.

Redshift Space Distortions (see Sec. 1.3.2) break the assumption of statistical isotropy since they introduce a preferred direction for clustering. In particular, clustering statistics become dependent on the orientation of pairs (in the case of 2-point statistics) or triplets (for 3-point statistics) with respect to the line of sight vector. This dependence makes

[1] In case of a cubic box with periodic boundary conditions, Eq. 2.1.1 is still optimal and therefore used for measurements of correlation functions in numerical simulations.

the 2PCF/power spectrum depend on two variables: the norm of the vector s separating a given pair and the cosine of the angle between s and the l.o.s. direction $\hat{z}$, $\mu \equiv \boldsymbol{s} \cdot \hat{\boldsymbol{z}}/s$. The estimator for the 2PCF becomes

$$\xi(s,\mu) = \frac{DD(s,\mu) - 2DR(s,\mu) + RR(s,\mu)}{RR(s,\mu)} \tag{2.1.5}$$

The anisotropic 2PCF can be expanded in Legendre polynomials

$$\xi_\ell(s) = \frac{2\ell+1}{2} \int_{-1}^{1} d\mu \, \mathcal{L}_\ell(\mu)\xi(s,\mu). \tag{2.1.6}$$

Taking the first three non-valishing multipoles, the monopole ($\xi_0(s)$), quadrupole ($\xi_2(s)$) and hexadecapole ($\xi_4(s)$) is advantageous, since the leading order approximation (see Eq. 1.3.12) generates anisotropy fully described by these three coefficients.

Two additional variables (e.g., two angles) are instead needed to describe the relative orientation of a triangle with respect to the line of sight. These two variables add to the three of the isotropic case, which constrain the size and shape of a triangle, so the redshift space bispectrum/3PCF depends on five variables. Our analysis is limited to the isotropic or spherically averaged 3PCF ($\ell = 0, m = 0$ part of Eq. 1.4.32). We refer the interested reader to Slepian and Eisenstein (2018) for the measurement of the anisotropic 3PCF in $\mathcal{O}(N^2)$ and to Sugiyama et al. (2019) for the measurement of the anisotropic 3PCF based on the multipole expansion on a Tripolar Spherical Harmonics (TripoSH) basis.

When measuring the isotropic 3PCF a convenient choice of the three variables that specify it is to consider two of the three triangle sides r_{12}, r_{13}, and the angle $\boldsymbol{r}_{12} \cdot \boldsymbol{r}_{13} = \cos(\theta)$ between them (where we label the triangle vertices with 1, 2 and 3). In this way, we can expand the 3PCF $\zeta(r_{12}, r_{13}, \cos\theta)$ into Legendre polynomials.

$$\zeta(r_{12}, r_{13}, \cos\theta) = \sum_{l=0}^{\infty} \zeta_\ell(r_{12}, r_{13}) \mathcal{L}_\ell(\cos\theta) \tag{2.1.7}$$

Where the coefficients $\zeta_\ell(r_{12}, r_{13})$ are given by the relation:

$$\zeta_\ell(r_{12}, r_{13}) = \frac{2l+1}{2} \int_{-1}^{1} d\cos\theta \, \zeta(r_{12}, r_{13}, \cos\theta) \mathcal{L}_\ell(\cos\theta) \tag{2.1.8}$$

The advantage of this formalism is threefold:

- **Efficiency**: The 3-point correlation function has a continuous and slowly varying dependence on the aperture angle. Therefore, it is natural to expect that the number of non-zero multipoles is small.
- **Visualization**: The Legendre polynomial decomposition allows us to move from the study of a function of three variables to that of some functions of two variables. For the latter, it is simpler to provide a general graphical representation, without the need to reduce to some details.

- **Computational cost**: The multipole moments of the 3PCF can be obtained with greater speed than other decompositions of the 3PCF without sacrificing accuracy.

Measurement of the 3PCF with Eq. 2.1.4 fundamentally scales as the number of possible triangles in a survey. If one wishes to measure the 3PCF out to some scale R_{max}, there are $N(nV_{R_{max}})^2$ relevant triangles, where N is the number of objects in the survey, n is the survey number density and $V_{R_{max}} = (4/3)\pi R_{max}^3$. A direct way to measure ζ_ℓ would be to sit on every possible origin and compute the angle between pairs of vectors pointing to all possible sets of two galaxies out to the radius R_{max} to which one wishes to measure the 3PCF, but this also scales as $N(nV_{R_{max}})^2$. The SHD algorithm exploits the spherical harmonic addition theorem to decompose the Legendre polynomials into factors that depend only on one angular variable each. First, the $\zeta_\ell(r_{12}, r_{13})$ multipoles are defined for each galaxy of the survey with vector $\boldsymbol{s}$ and radially binned around it, yielding the following estimator

$$\bar{\hat{\zeta}}_\ell(r_{12}, r_{13}, \boldsymbol{s}) = \frac{2\ell+1}{(4\pi)^2} \int d\hat{r}_{12} d\hat{r}_{13} \delta(\boldsymbol{s}) \bar{\delta}(r_{12}, \hat{r}_{12}, \boldsymbol{s}) \bar{\delta}(r_{13}, \hat{r}_{13}, \boldsymbol{s}) \mathcal{L}_\ell(\hat{r}_{12}, \hat{r}_{13}). \tag{2.1.9}$$

Then, using the spherical harmonics addition theorem, we get

$$\bar{\hat{\zeta}}_\ell(r_{12}, r_{13}, \boldsymbol{s}) = \frac{1}{4\pi} \delta(\boldsymbol{s}) \sum_{m=-\ell}^{\ell} \int d\Omega_{12} \bar{\delta}(r_{12}, \hat{r}_{12}, \boldsymbol{s}) Y_{\ell m}(\hat{r}_{12}) \times \int d\Omega_{13} \bar{\delta}(r_{13}, \hat{r}_{13}, \boldsymbol{s}) Y^*_{\ell m}(\hat{r}_{13}). \tag{2.1.10}$$

Next, we can find the expansion coefficients of the density field (as a function of angle alone) in spherical harmonics, in a fixed radial bin. In other words, if we compute for each radial bin r:

$$a_{\ell m}(r, \boldsymbol{s}) = \int d\Omega \; \delta(r, \hat{r}, \boldsymbol{s}) Y^*_{\ell m}(\hat{r}) \tag{2.1.11}$$

we can construct all combinations dictated by r_{12} and r_{13}, without ever needing to do an $\mathcal{O}(N^2)$ operation. Explicitly, inserting Eq. 2.1.11 into Eq. 2.1.10, we find:

$$\bar{\hat{\zeta}}_\ell(r_{12}, r_{13}, \boldsymbol{s}) = \frac{1}{4\pi} \delta(\boldsymbol{s}) \sum_{m=-\ell}^{\ell} a_{\ell m}(r_{12}, \boldsymbol{s}) a^*_{\ell m}(r_{13}, \boldsymbol{s}) \tag{2.1.12}$$

By precomputing the $a_{\ell m}(r, \boldsymbol{s})$ coefficients in each radial bin, we then only need to construct $(N_{bins}+1)N_{bins}/2$ combinations of these coefficients out to R_{max}. Computing these coefficients scales as the number of objects contained in the sphere, i.e, $(nV_{R_{max}})$. We still must integrate over all possible choices of origin, because the galaxies are discrete; this will reduce to a sum with N terms

$$\bar{\zeta}_\ell(r_{12}, r_{13}) = \frac{1}{V} \sum^{N} \bar{\hat{\zeta}}_\ell(r_{12}, r_{13}, \boldsymbol{s}) \tag{2.1.13}$$

Thus, this algorithm will scale as $N(nV_{R_{max}})$; linear in both the total number of galaxies and the number within a sphere of radius R_{max}, and a factor of order $(nV_{R_{max}})$ faster than

the naive counting approach. A visual representation of the steps performed is shown in Fig. 2.1.

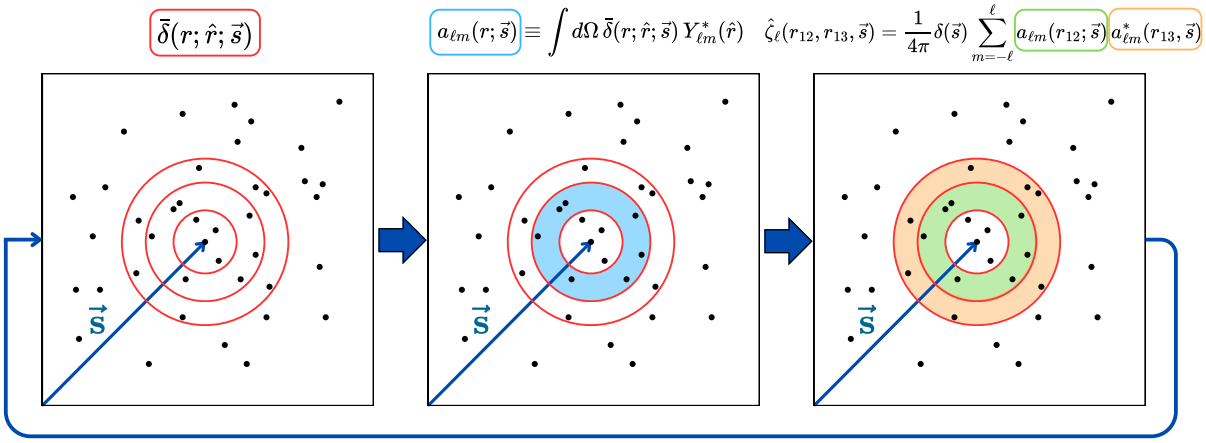

Figure 2.1: The algorithm sits on each galaxy in the survey at position denoted by vector $\vec{s}$ (left panel); It computes the spherical harmonic expansion of the density field in concentric spherical shells (radial bins) around that point (central panel). The $a_{\ell m}$s can be combined to yield the multipole moments around this galaxy (sum over m) and then translation-averaged to yield ζ_ℓ for the survey (right panel).

2.2 The Baryon Oscillation Spectroscopic Survey

The Baryon Oscillation Spectroscopic Survey (BOSS) was one of the primary observational programs of SDSS-III (York et al. 2000). Over the course of the survey, spectroscopic redshifts were obtained for approximately 1.37 million massive galaxies, along with Lyα forest spectra for roughly 150,000 quasars (Eisenstein et al. 2011). Its name hints at the ability of baryon acoustic oscillations to be used as a standard ruler for the measurement of distances. In particular, the survey was designed to enable percent-level constraints on the angular diameter distance $D_A(z)$ and the Hubble expansion rate $H(z)$, at redshifts $z \sim 0.3$ and $z \sim 0.6$.

To achieve these goals, BOSS was optimized to sample a large volume with sufficient tracer density. For the main galaxy sample at redshifts $0.1 < z < 0.6$, this required coverage of approximately 10,000 deg^2 and a comoving galaxy number density of $\sim 3 \times 10^{-4}\, h^3\, \mathrm{Mpc}^{-3}$. Given the observational constraints of the Sloan 2.5-m telescope and its spectrographs, galaxies were selected from a much larger photometric catalogue using carefully defined criteria.

The target selection strategy was designed to satisfy several requirements: the selected galaxies needed to populate the desired redshift range with sufficient density, exhibit well-characterized stellar populations that isolate a certain galaxy type, and possess spectral features that allow reliable redshift measurements within relatively short exposure times. In addition, the selection was required to minimize contamination from non-galaxy sources, remain spatially uniform across the survey footprint, and be robust against systematic uncertainties in the imaging data.

Photometric targets were drawn from SDSS imaging in the *ugriz* bands, covering approximately 7,600 deg^2 in the Northern Galactic Cap and an initial 600 deg^2 in the Southern Galactic Cap. Additional imaging obtained during SDSS-III expanded the southern coverage to a contiguous area of about 3,200 deg^2. Spectroscopic observations were carried out using multi-object spectroscopy, in which fibers were plugged into manually drilled aluminum plates, each spanning a field of roughly 3° in diameter. Observations consisted of repeated 900-second exposures, accumulated until a predefined signal-to-noise threshold was achieved for the faintest galaxy targets.

Galaxy targets were identified using combinations of colour-magnitude and colour-colour cuts chosen to select luminous galaxies with prominent spectral features. At lower redshifts ($z \lesssim 0.4$), the sample extends the luminous red galaxy population observed in earlier phases of SDSS to fainter magnitudes. At higher redshifts, the selection broadens to include galaxies of diverse intrinsic colours, approximating a stellar-mass-limited sample rather than one restricted to red galaxies.

2.2.1 A Large-Scale Structure galaxy catalogue

The BOSS mask is divided into two redshift selections: the LOWZ catalogue has galaxies in the range $0.1 < z < 0.5$ and the CMASS catalogue has galaxies in the range $0.4 < z < 0.7$. The mask is constructed of spherical polygons that resemble the target line of sight of the spectroscopic tiles used during the observations. Fig. 2.2 shows the fraction of the total BOSS area that has target completeness greater than a specified value. While in previous BOSS data releases, completeness cuts were applied to the mask, in the 12th Data Release (DR12), they are not, as accounting for them only removes a negligible amount of area (Reid et al. 2016).

While the basic geometry of the survey is encapsulated in the survey mask described, there are many small regions within it where galaxies could not have been observed. These go under the name of *Veto masks* and they include:

- *Centerpost mask*: Each Sloan plate is secured to the focal plane by a central bolt, no targets coinciding with the centerpost of a spectroscopic tile can be observed;
- *Bright objects mask*: Area surrounding stars and bright local galaxies which saturate the image are masked;
- *Seeing cut*: regions where the point spread function full width half maximum is greater than 2.3, 2.1, 2.0 in the g, r, and i bands, respectively, are discarded;
- *Extinction cut*: for similar reasons, areas where the E(B−V) extinction exceeds 0.15.

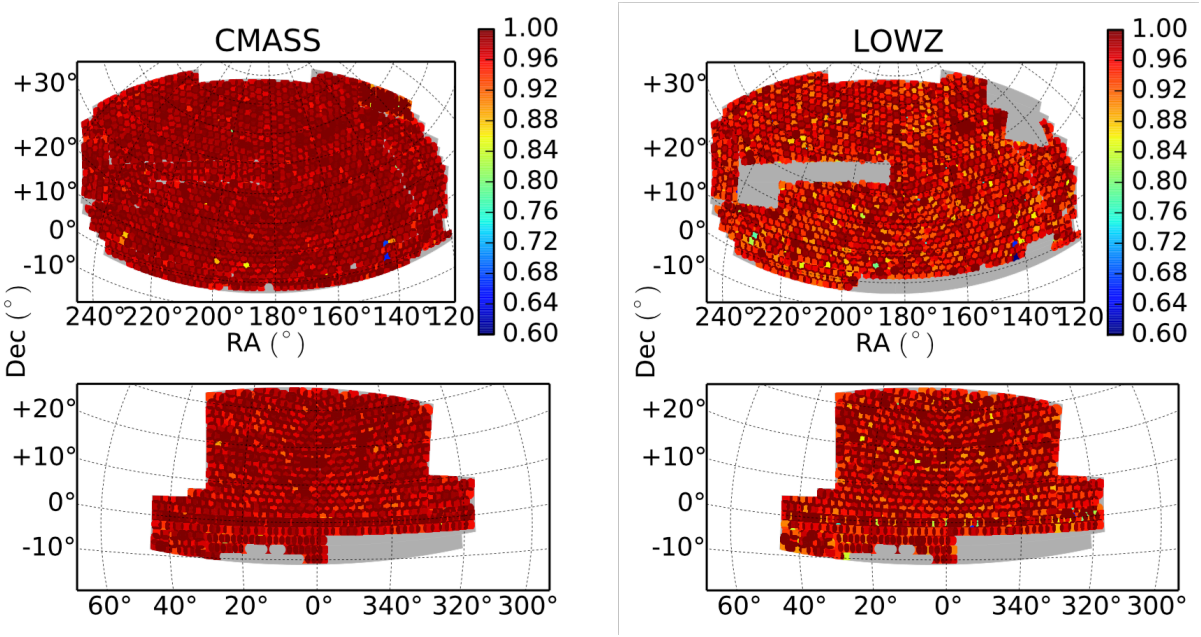

Figure 2.2: The BOSS survey footprint divided from left to right in redshift bins. The top figures show the Northern Galactic Cap (NGC) and the bottom ones show the Southern Galactic Cap (SGC) selections. Each spherical polygon represents a spectroscopic tile. The color on each tile is weighted by completeness. The images are taken from Reid et al. (2016)

In order to minimize the impact of observational artefacts on the estimate of the true galaxy overdensity field, galaxies are weighted differently in the final LSS catalogue. The total weight for each galaxy, denoted with i, is

$$w_{tot,i} = w_{systot,i}(w_{cp,i} + w_{noz,i} - 1) \tag{2.2.1}$$

Where $w_{systot,i} = w_{star,i}w_{see,i}$ takes into account two terms. These effects correspond to

- $w_{star,i}$ accounts for the fact that the projected density of galaxies depends on the local stellar density;
- $w_{see,i}$ accounts for the fact that the projected density of galaxies depends on the seeing at the time of observation of a particular field;
- $w_{cp,i}$ accounts for fibre collisions. Some selected objects were not assigned a spectroscopic fibre because too close to another fibre. In such cases, the nearest galaxy from the same target class that was assigned a fibre is upweighed;
- $w_{noz,i}$ accounts for redshift failures, i.e., galaxies for which it is impossible to obtain an accurate value of redshift. Their "weight" is transferred to their nearest neighbor.

The final data release (DR12) LSS catalogue[2] contains, in the LOWZ sample, 361 762 galaxies in the range $0.15 < z < 0.43$, with 248 237 in the North Galactic Cap (NGC) and

[2]https://www.sdss4.org/dr12/

113525 in the South Galactic Cap (SGC). The CMASS sample contains 777 202 galaxies in the range $0.43 < z < 0.70$, with 568 776 in the NGC and 208 426 in the SGC to a total of 1 138 964 galaxies. In our analysis, we make use of the so-called *optimally binned combined sample*. This catalogue was created in the DR12 release and used in all of the main BOSS papers; The method and motivation for it are detailed in Alam et al. (2017). In this new catalogue the LOWZ and CMASS selection functions are combined into two non-overlapping redshift bins with $0.2 < z < 0.5$ at $z_{\rm eff} = 0.38$ and $0.5 < z < 0.75$ at $z_{\rm eff} = 0.61$. These new selections have very similar effective volumes, unlike the LOWZ-CMASS catalogue. A third overlapping bin with galaxies in the range $0.4 < z < 0.6$ is also part of the optimally binned combined sample but we do not utilize it. Throughout the analysis we call these two bins "*low-z*" and "*high-z*" respectively to not draw confusions with the other samples' names. Each redshift bin sample is additionally divided into two sub-samples depending on the Galactic hemisphere where the galaxies are observed, called the South and North Galactic Cap (SGC and NGC). The number of galaxies and effective volumes used in our analysis are shown in 2.1. The effective volumes quantify the final constraining power of clustering statistics, as those can be better estimated by larger volumes, i.e., more galaxies and smaller uncertainty in the estimation of statistical quantities such as correlation functions. The effective volume is computed as

$$V_{\rm eff} = \int dV \left(\frac{n(z)P_0}{1 + n(z)P_0} \right)^2 \tag{2.2.2}$$

where P_0 is an estimate of the amplitude of the power spectrum at the BAO scale, here assumed to be $P_0 = 10000\, h^{-3}Mpc^3$ (Reid et al. 2016), and $n(\bar{z})$ is the galaxy number density, which can be estimated by assuming a fiducial cosmology as

$$n(\bar{z}) = \frac{N(\bar{z})}{\Delta\Omega \int_{z_1}^{z_2} \frac{cD_C^2(z)}{H(z)} dz} \tag{2.2.3}$$

where $\Delta\Omega$ is the solid angle subtended by the survey and $\bar{z}$ refers to the redshift bin that goes from z_1 to z_2.

		N_{gals}	$V_{\rm eff}$ (Gpc^3)	V (Gpc^3)
$z_{\rm eff} = 0.38$	NGC	429 182	2.7	4.7
	SGC	174 819	1.0	1.7
	total	604 001	3.7	6.4
$z_{\rm eff} = 0.61$	NGC	435 741	3.0	9.0
	SGC	158 262	1.1	3.3
	total	594 003	4.1	12.3

Table 2.1: Number of galaxies, effective and total survey volumes for the *optimally binned* samples at different redshift and sky regions. $P_0 = 10000\, h^{-3}Mpc^3$ is used for the computation of effective survey volumes. (Alam et al. 2017).

The number density as a function of redshift for the BOSS survey is shown in Fig. 2.3. The result shows that the CMASS+LOWZ number density goal of $n(z) \sim 3 \times$

$10^{-4}h^{-3}Mpc^3$ is reached for the entire redshift range, expect around a local minimum in the overlap region of $\bar{n}(z = 0.41) = 2.2 \times 10^{-4}h^{-3}Mpc^3$.

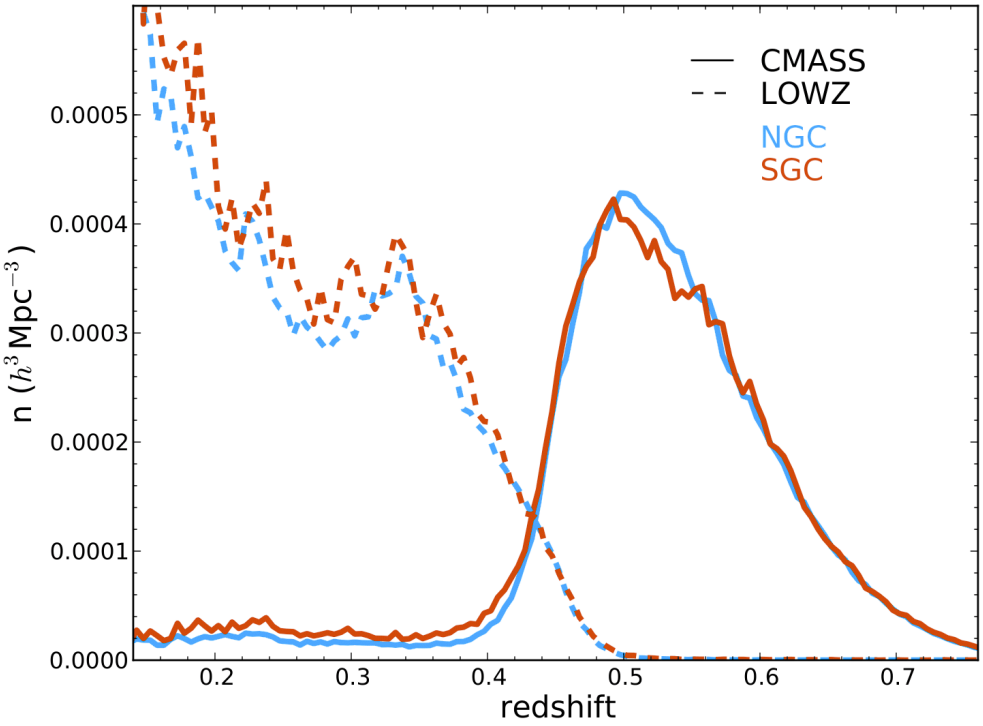

Figure 2.3: Number density of all four target classes assuming a fiducial cosmology with $\Omega_m = 0.31$. Image taken from Ross et al. 2017

2.2.2 Combining the Southern and Northern Galactic Caps

The DR12 combined sample is observed across the two Galactic hemispheres, referred to as the Northern and Southern galactic caps (NGC and SGC, respectively). As these two regions do not overlap, they are prone to slight offsets in their photometric calibration as first reported by Schlafly and Finkbeiner (2011). For this reason, NGC and SGC subsamples probe slightly different galaxy populations; this variation is more evident in the low redshift sample, LOWZ. However, as explained in Ross et al. (2017), there are no strong discrepancies between the NGC and SGC configuration-space clustering of BOSS galaxies at scales relevant to BAO or RSD studies. They also demonstrate that the results applied to each hemisphere individually are consistent with the combined constraints and that the BAO results are thus robust to any concerns about combining the NGC and SGC results when only scales above 20 h^{-1}Mpc are considered. Alam et al. (2017) show discrepancies between the two hemispheres are more apparent at small scales when studying the power spectrum. They decide to simply allow sufficient freedom when fitting models to the clustering statistics in each galactic cap, as to allow for this slight change

in galaxy population. In practice, the different Fourier-space statistics are modelled with different nuisance parameters in the two hemispheres. This approach is also considered by Philcox and Ivanov (2022) in a joint power spectrum and bispectrum analysis.

2.2.3 Main results from BOSS clustering analyses

We review the BOSS DR12 results from 2- and 3-point clustering analyses. As mentioned above, the main goal of BOSS was to constrain $D_A(z)$ and $H(z)$ down to the % level. This has been archived by several research teams working independently with a variety of analysis tools and modelling assumptions.
Extracting distance information from clustering statistics is possible in 2 main ways:

Template-fitting analyses fix the linear power spectrum shape, compressing the cosmological information on few parameters that govern the late-time expansion. Constraints are extracted from redshift-space distortions, which probe the growth rate of structure, and Alcock-Paczynski parameters, which constrain distance measures such as $D_A(z)$ and $H(z)$. By expressing the model as a linear combination of precomputed templates, likelihood evaluations are computationally inexpensive.

Full-shape analyses aim to extract all available cosmological information and are usually used to constrain shape parameters such as $\{n_s, A_s, \omega_{cdm}, \omega_b, h\}$. This approach retains sensitivity to primordial physics and the late-time expansion history. Its main drawback is the high computational cost, particularly for configuration-space statistics (2PCF and 3PCF), which require Fourier-to-configuration-space transformations. Recent emulator-based methods substantially alleviate this limitation but have yet to be explored for the 3PCF in actual spectroscopic surveys.

To achieve the level of precision required by BOSS goals, template fitting analyses often exploit *reconstruction*, which attempts to reverse non-linear effects present at the level of the galaxy overdensity field. This has the effect of restoring information from higher-order functions back to the 2-point-function and sharpens the BAO peak, therefore increasing measurement precision. Naturally, this type of analysis only exploits 2-point statistics as reconstruction purposefully removes signal from 3-point statistics.

In Tab. 2.2 we list BOSS DR12 clustering analyses which go under the category of *Template fit* and *Full-shape*[3]. We list as *BAO (det.)* those studies that aim to detect the BAO peak in 3-point statistics. The AP parameters are measured using the anisotropic two-point correlation function multipoles in Ross et al. (2017) and Cuesta et al. (2016) and using the anisotropic power spectrum in Beutler et al. (2017). The full shape of the anisotropic two-point correlation function is computed and analysed using multipoles in Satpathy et al. (2017) and using μ-wedges in Sánchez et al. (2017). The full shape of the power spectrum is analyzed by Beutler et al. (2016) and more recently by Ivanov et al. (2020).

[3]The studies that are listed under 2-point clustering (mainly) come from the BOSS collaboration, but many more studies have been made by other research groups.

Studies at the level of 3-point statistics have been made primarily in Fourier space. Gil-Marín et al. (2016) uses a combined measurement of power spectrum multipoles with isotropic bispectrum in the template fitting approach, showing that by combining the two statistics it is possible to disentangle the degeneracy between f and σ_8. Philcox and Ivanov (2022) combine the full-shape power spectrum and bispectrum and constrains shape parameters with and without the use of Planck priors on A_s and n_s. In configuration space, studies have explored the detectability of the BAO peak in the 3PCF by focusing on a few triangle configurations. Slepian et al. (2017) makes a BAO significance analysis and finds the BAO peak in the BOSS 3PCF with a 4.5 σ significance. While Moresco et al. (2021) detects the BAO peak in the 3PCF of SDSS clusters. Slepian et al. (2017) also makes a distance measurement using AP parameterization of the 3PCF template model but marginalizes over f by fixing it in the fit.

An exploration of the parameter space that exploits the full information content of the 3PCF triangle configurations is still lacking for BOSS. This work attempts to fill in the gap by performing a joint analysis of the 2PCF and 3PCF of BOSS in the template fitting approach. While Guidi et al. (including L. Cavazzini) (2026, in prep.) will perform the first full-shape analysis of the joint 2PCF and 3PCF.

2-point clustering		
	$P(k)$ multipoles/μ-wedges	$\xi(r)$ multipoles/μ-wedges
Template fit	Beutler et al. (2017)	Ross et al. (2017) Cuesta et al. (2016)
Full-shape	Grieb et al. (2017) Beutler et al. (2016) Ivanov et al. (2020)	Satpathy et al. (2017) Sánchez et al. (2017)
3-point clustering		
	$P(k) + B(k_1, k_2, k_3)$	$\xi(r) + \zeta(r_{12}, r_{13}, r_{23})$
BAO (det.)		Slepian et al. (2017) Moresco et al. (2021)
Full-shape	Philcox and Ivanov (2022)	**Guidi in prep.**
Template fit	Gil-Marín et al. (2016)	**this work**

Table 2.2: BOSS literature on clustering studies. The table is divided by studies made on 2-point statistics (top) and 3-point statistics (bottom). BAO (det.) studies focus on the detectability of the BAO peak in the 3PCF.

2.3 Measurements of the BOSS 2PCF and 3PCF

We use measures, taken by Guidi et al. (including L. Cavazzini) (2026, in prep.), of the monopole, quadrupole, and hexadecapole of the galaxy 2PCF (ξ_0, ξ_2, ξ_4), and the isotropic multipoles of the galaxy 3PCF ζ for ℓ in the range $0 \leq \ell \leq 10$ of the 'optimally binned' BOSS DR12 LSS catalogue.

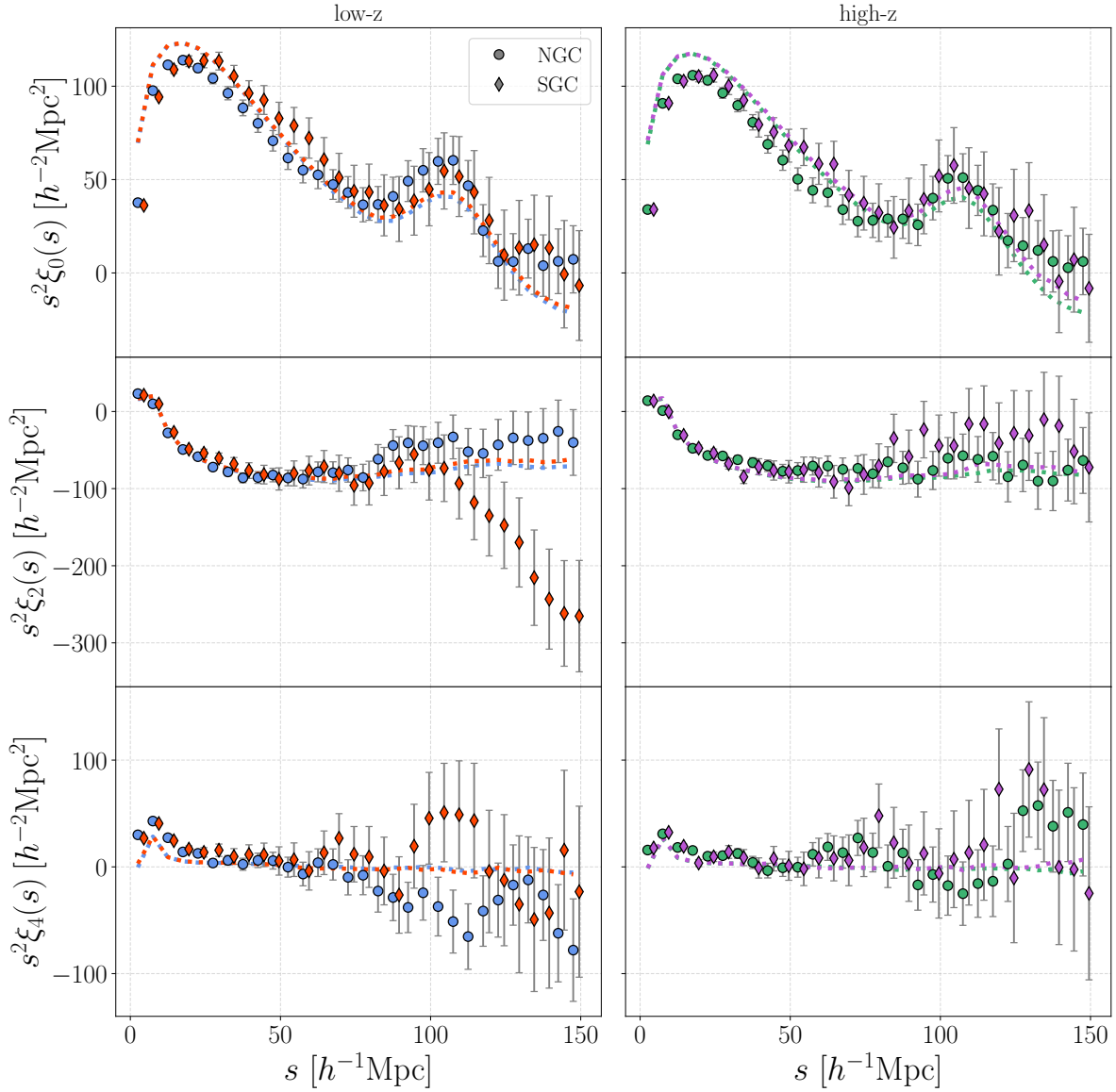

Figure 2.4: From top to bottom the measured 2PCF multipoles ξ_0,ξ_2 and ξ_4. Figures on the left are the low redshift sample (low-z) and on the right the high redshift sample (high-z). The error bars represent the uncertainty of each bin. The dotted lines represent the mean value of each bin computed from 2000 mock catalogues.

The measures have been extracted using `MeasCorr`[4](Farina et al. 2026) to estimate 2PCF multipoles in the range $0 \leq r \leq 150\,h^{-1}$Mpc and 3PCF multipoles in the range $0 \leq r_{12}, r_{13} \leq 150\,h^{-1}$Mpc , with a binning of $\Delta r = 5\,h^{-1}$Mpc; with bin centers at $2.5, 7.5, 12.5, \ldots\, h^{-1}$Mpc.

The measured 2PCF multipoles are shown in Fig. 2.4, together with the mean of ξ obtained from the 2000 mock catalogues. The role of mock catalogues in estimating statistical uncertainties is discussed in the following section. The BOSS measurements exhibit an excess of power on large scales ($r \gtrsim 100\,h^{-1}$Mpc) relative to the mock mean. In particular, the monopole is expected to cross zero at $r \simeq 120\,h^{-1}$Mpc, whereas the BOSS data do not show a clear transition to negative values.

The NGC and SGC samples display small mutual variations, with the most noticeable difference appearing in the quadrupole ξ_2 of the low-redshift sample. As discussed in Sec. 2.2.2, such differences between the two caps are expected due to observational and systematic effects. These variations are statistically significant only on small scales ($r < 20\,h^{-1}$Mpc). On larger scales, they are sufficiently small that the NGC and SGC samples can be treated as probing statistically equivalent galaxy populations.

While the 2PCF is relatively straightforward to visualize and interpret, the three-point correlation function (3PCF) poses additional challenges owing to its dependence on three distinct length scales. In the following, we describe the different visualization strategies adopted for the 3PCF, highlighting their respective advantages and limitations. Inevitably, there is a trade-off between visual clarity and information content: the more restrictive the cuts applied to emphasize specific features, the greater the loss of information about the full dataset. To avoid unnecessary redundancy, we present results only for the NGC high-z sample.

3PCF multipoles. The measured 3PCF multipoles are shown in Fig. 2.5. Within each multipole ℓ block, the colour-coded values represent the amplitude of ζ_ℓ for each (r_{12}, r_{13}) pair. Each block corresponds to a symmetric matrix, with larger positive (negative) values indicated by redder (bluer) colours.

The multipole decomposition corresponds to an expansion of the 3PCF in Legendre polynomials. Each multipole therefore captures the contribution of polynomials of increasing order to the full signal: ζ_0 represents the isotropic (constant) component, ζ_1 the contribution from first-order polynomials, ζ_2 from second-order polynomials, and so forth.

The BOSS measurements exhibit a strong signal on small, nearly isosceles scales. As the multipole order increases, the off-diagonal regions of the matrices rapidly approach zero for $\ell \gtrsim 5$ (see, for instance, the lower-right corner of the colormaps and, by symmetry, the upper-left corner). In contrast, along the diagonal corresponding to isosceles configurations with $r_{12} = r_{13}$ the signal remains large. This behaviour reflects a known limitation of the SHD estimator, which struggles to accurately reconstruct isosceles triangle configurations. Since the reconstruction of ζ via Eq. 2.1.7 requires the multipole sum to converge after a finite number of terms, this feature is particularly relevant.

[4]https://gitlab.com/veropalumbo.alfonso/meascorr

In addition, a localized enhancement is visible near $r \simeq 100\,h^{-1}$Mpc, especially in the low-order multipoles, and can be associated with the baryon acoustic oscillation (BAO) feature.

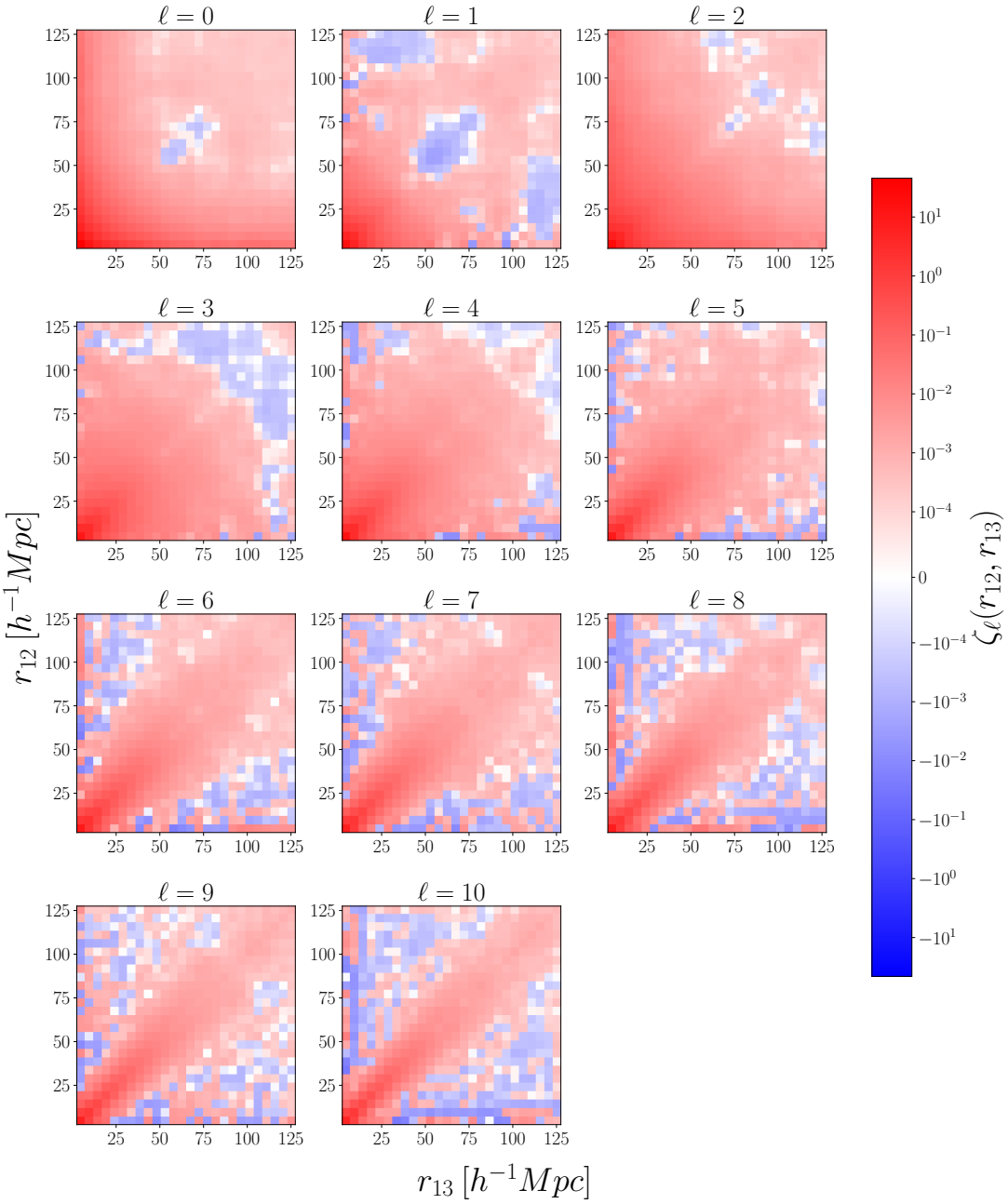

Figure 2.5: Isotropic 3PCF multipoles from $\ell = 0$ to $\ell = 10$ for the NGC high-z sample. The colorbar displays the amplitude of ζ_ℓ in each (r_{12}, r_{13}) bin.

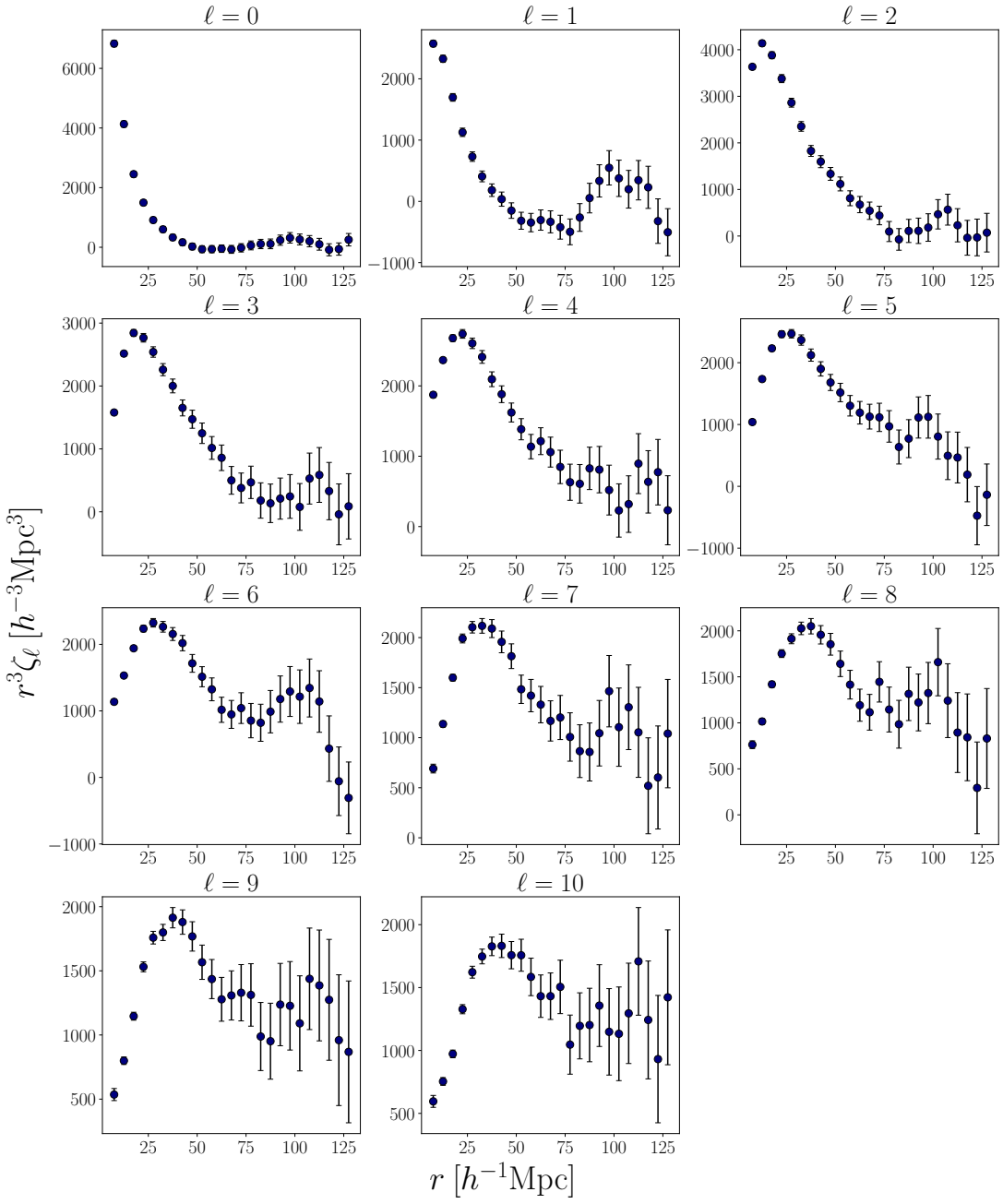

Figure 2.6: Isotropic 3PCF multipoles from $\ell = 0$ to $\ell = 10$ for the NGC high-z sample. Only isosceles configurations are displayed ($r = r_{12} = r_{13}$).

An alternative visualization of the multipoles can be obtained by taking one-dimensional slices through the two-dimensional matrices. For example, a cut along the diagonal isolates isosceles configurations, while a horizontal cut selects a fixed value of r_{12} and displays ζ_ℓ as a function of r_{13}. In Fig. 2.6, the semi-isosceles ($r_{12} = r_{13} + \Delta r$) multipoles are shown, essentially an off-diagonal cut of the full 2-D matrices. This cut provides a more intuitive representation, with a distinctive peak around $r \simeq 100\,h^{-1}\mathrm{Mpc}$.

Single-scale 3PCF The single-scale three-point correlation function (3PCF) provides a direct way to inspect the angular dependence of the signal by fixing two sides of a triangle, r_{12} and r_{13}, and varying the angle θ between them. Because only a finite number of multipoles can be measured, the angular dependence $\zeta(r_{12}, r_{13}, \theta)$ is reconstructed by truncating the Legendre expansion at $\ell = \ell_{\max}$,

$$\zeta^{r_{12},r_{13}}(\theta) = \sum_{\ell=0}^{\ell_{\max}} \zeta_\ell(r_{12}, r_{13})\, \mathcal{L}_\ell(\cos\theta). \tag{2.3.1}$$

This reconstruction is accurate only if the contributions from multipoles with $\ell > \ell_{\max}$ are negligible, a condition that generally depends on the specific values of r_{12} and r_{13}.

A useful comparison can be made with measurements obtained from direct triplet counts rather than the SHD estimator. In particular, Veropalumbo et al. (2021) show that for nearly isosceles configurations ($r_{12} \simeq r_{13}$) the 3PCF measured from direct counts steepens rapidly as $\theta \to 0$, corresponding to the limit $r_{23} \to 0$. This sharp feature cannot be accurately reconstructed using the SHD estimator, even for $\ell_{\max} \sim 30$. The reconstruction improves as the difference between r_{12} and r_{13} increases.

Figure 2.7 illustrates this behaviour for two representative configurations. For an isosceles triangle with $r_{12} = r_{13} = 32.5\,h^{-1}$Mpc, the panel showing the individual multipole contributions (yellow curves) demonstrates that truncation at $\ell_{\max} = 10$ is insufficient to accurately reproduce the angular dependence. In contrast, for the configuration $r_{12} = 42.5\,h^{-1}$Mpc and $r_{13} = 87.5\,h^{-1}$Mpc, the multipoles rapidly converge to zero, yielding a well-behaved reconstruction.

This representation constitutes the most intuitive visualization of the 3PCF, as it provides direct insight into its physical interpretation. In Fig. 2.7, the angular dependence exhibits a characteristic U-shape, reflecting the filamentary structure of the cosmic web: the correlation is strongest for elongated triangles, corresponding to $\theta/\pi \simeq 0$ and $\theta/\pi \simeq 1$. For the configuration with $r_{12} = 42.5\,h^{-1}$Mpc and $r_{13} = 87.5\,h^{-1}$Mpc (top panel), a secondary peak is visible near $\theta/\pi \simeq 0.5$, corresponding to a third side length of $r_{23} \sim 100\,h^{-1}$Mpc. This feature is commonly interpreted as the baryon acoustic oscillation (BAO) signature in the 3PCF.

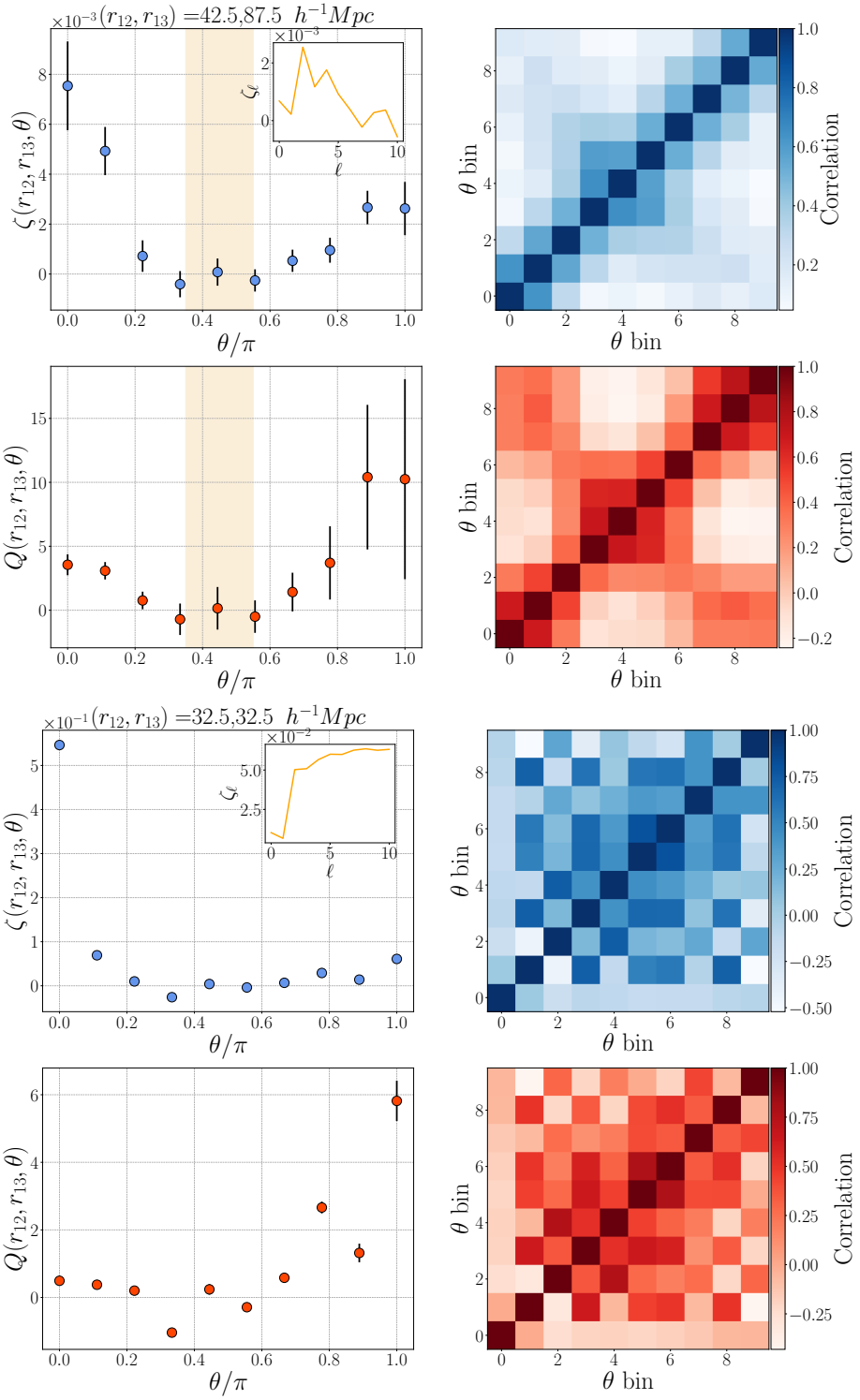

Figure 2.7: The single scale connected 3PCF in blue and reduced 3PCF in red, for the fixed scales $(r_{12}, r_{13}) = 42.5,\ 87.5\,h^{-1}Mpc$ (top) and $(r_{12}, r_{13}) = 32.5,\ 32.5\,h^{-1}Mpc$ (bottom). Alongside their associated correlation matrices. The small figure inside the ζ plots displays the amplitude of each ζ_ℓ multipole used in the resummation. Gray bars indicate the supposed location of the BAO peak.

To assess the statistical significance of the BAO feature in the 3PCF, Moresco et al. (2021) developed two complementary approaches:

- *BAO difference:* two templates are fitted to the data, one physical template including BAO features and one "no-wiggle" template in which the BAO has been removed. The square root of the difference in χ^2 between the two fits provides a measure of the BAO detection significance.
- *BAO contrast:* this method identifies the BAO feature as a local maximum embedded between two broader peaks at $0 < \theta < \pi$. The peak position is determined using derivatives, while the adjacent local minima define a baseline via linear interpolation. The BAO amplitude is then quantified as the height of the peak above this baseline, yielding a model-independent measure of BAO significance.

Applying these metrics to both the connected and reduced single-scale 3PCF, Moresco et al. (2021) found a higher BAO significance in the reduced 3PCF of SDSS galaxy clusters, illustrating one of the potential advantages of this estimator.

Despite its high degree of physical interpretability, the single-scale 3PCF contains limited information, as it probes only a restricted subset of triangle configurations. Consequently, it is suboptimal for constraining cosmological parameters in a Bayesian analysis.

All-scales 3PCF. This function is useful to show the information contained in the 3PCF for all the considered side-binned triangles. The *all-scales* 3PCF is estimated by constructing the following function:

$$\zeta_{\rm all}(r_{12}, r_{13}, r_{23}) = \sum_{\ell=0}^{\ell_{\rm max}} \zeta_\ell(r_{12}, r_{13})\, \mathcal{L}_\ell(r_{12}, r_{13}, r_{23}), \tag{2.3.2}$$

where the side lenghts r_{ij} are the centers of equally spaced bins and the triplets (r_{12}, r_{13}, r_{23}) satisfy the following conditions:

$$\begin{cases} r_{12} \leq r_{13} \leq r_{23} \\ r_{23} \leq r_{12} + r_{13} \end{cases} \tag{2.3.3}$$

and

$$\mathcal{L}_\ell(r_{12}, r_{13}, r_{23}) \equiv \mathcal{L}_\ell(\cos\theta), \quad \cos\theta = \frac{r_{12}^2 + r_{13}^2 - r_{23}^2}{2\, r_{12} r_{13}}. \tag{2.3.4}$$

The first condition in Eq. 2.3.3 avoids redundancy, while the second one ensures that the triplet (r_{12}, r_{13}, r_{13}) represents the sides of a triangle. Notice that one can associate an index to order each triplet (r_{12}, r_{13}, r_{23}) satisfying the above conditions

by deciding how to loop over the sides. In this work, we always loop through r_{23} in the innermost loop and r_{12} in the outermost loop.

When including all scales, it is favorable to exclude pathological configurations, as seen from single scales ζ, isosceles triangles are not well reconstructed by the SHD algorithm when $\ell_{max} \sim 10$. To exclude such configurations Veropalumbo et al. (2022) introduces

$$\eta \equiv \frac{|r_{13} - r_{12}|}{\Delta r}. \tag{2.3.5}$$

Imposing a lower bound $\eta \geq \eta_{\rm min}$ effectively removes pathological configurations, with the specific choice of $\eta_{\rm min}$ determining how restrictive the cut is. In particular, $\eta_{\rm min} = 1$ excludes only triangles with $r_{12} = r_{13}$, while $\eta_{\rm min} = 2$ also removes configurations with $r_{12} = r_{13} + \Delta r$, and so on.

Although the all-scales 3PCF offers limited direct visualization power, this can be partially alleviated by explicitly ordering the triangle sides as (r_{21}, r_{13}, r_{23}), as shown in the bottom panels of Fig. 2.8. This ordering aids in identifying the general triangle shape associated with each measurement. Additional insight can be gained by classifying triangle configurations according to simple geometric criteria. In Fig. 2.8, we distinguish three triangle types: equilateral triangles are defined as having all three sides equal within one bin, $+\Delta r = 5\,h^{-1}$Mpc; isosceles triangles satisfy the condition $r_{12} = r_{13} + \Delta r$; all remaining configurations are classified as scalene. Under this scheme, it is straightforward to identify which triangle types contribute the strongest signal.

An alternative classification is shown in Fig. 2.9, where different symbols and colours are used to denote specific physical regimes. Green dots correspond to small-scale triplets, defined as configurations with $r_{12} \leq 50\,h^{-1}$Mpc. Red triangles indicate BAO triplets, namely configurations with at least one side in the range $[97.5, 117.5]\,h^{-1}$Mpc, where baryon acoustic oscillations imprint their characteristic oscillatory features on the correlation functions. Yellow stars identify squeezed BAO configurations, consisting of one side much smaller than the other two, at least one of which lies in the BAO range. All remaining configurations are represented by light-blue dots.

The primary strength of the all-scales 3PCF does not lie in its visualization, but rather in its information content. Including all triangle configurations maximizes the constraining power of the data and is therefore best suited for Bayesian analyses. For this reason, the main analysis of this work focuses on the all-scales 3PCF.

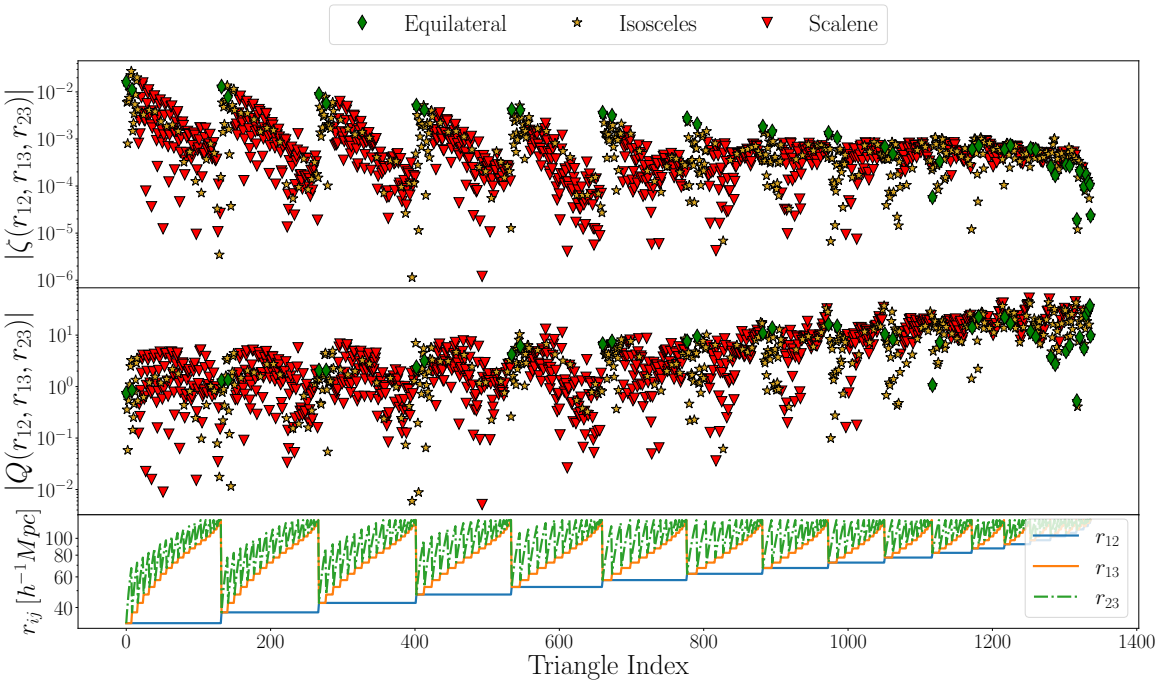

Figure 2.8: The measured NGC high-z, all-scales ζ and Q for all triangles between $r = 30 - 130\,h^{-1}$Mpc. Triangles are classified by their side lengths as equilateral, isosceles and scalene. At the bottom, the scale ordering (r_{12}, r_{13}, r_{23}) is shown.

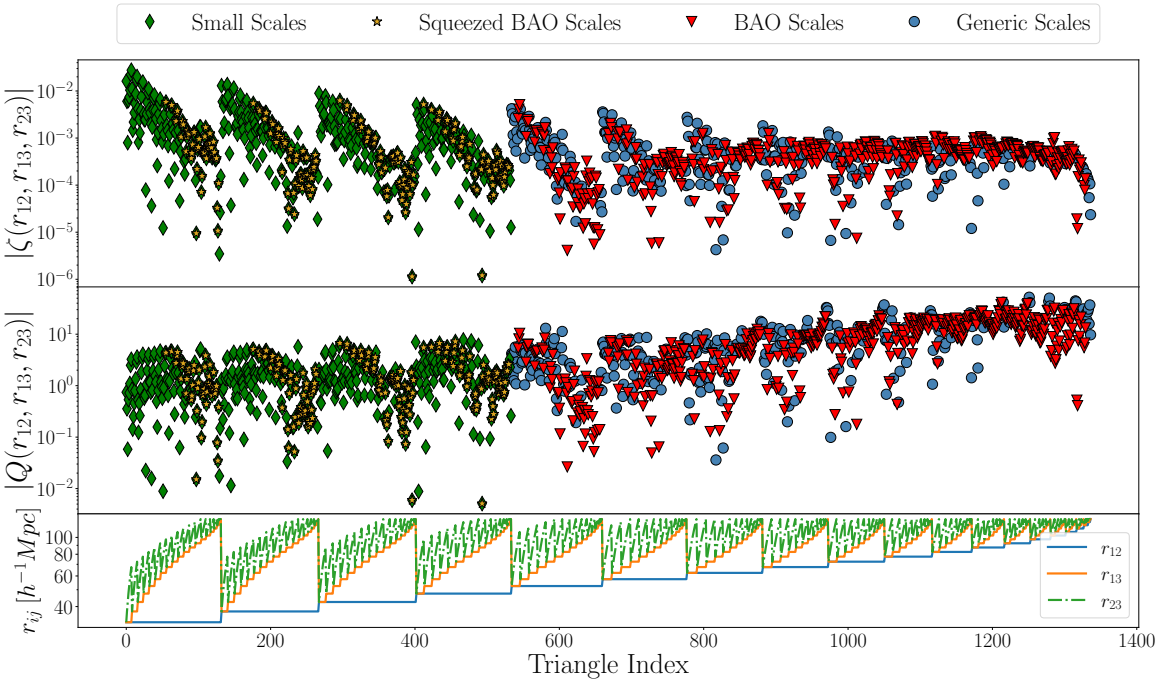

Figure 2.9: The measured NGC high-z, all-scales ζ and Q for all triangles between $r = 30 - 130\,h^{-1}$Mpc. Triangles are classified by varying criteria, described in the main text. At the bottom, the scale ordering (r_{12}, r_{13}, r_{23}) is shown.

2.4 Covariance matrices

The covariance matrices are estimated numerically from the 2PCF and 3PCF multipoles measured in 2000 MULTIDARK-`patchy` mock catalogues (Kitaura et al. 2016). Each has been tuned to match the footprint, redshift distribution, and halo occupation distribution of BOSS samples. We therefore expect the clustering of the mock samples to match the BOSS measurements. The `patchy` algorithm is based on extended Lagrangian perturbation theory and a stochastic halo biasing scheme calibrated on high-resolution N-body MULTYDARK simulations run for a ΛCDM cosmology with the following fiducial parameters:

$$\{\, \Omega_m = 0.307, \quad \Omega_\Lambda = 0.692, \quad \Omega_b = 0.048, \quad \sigma_8 = 0.828, \quad h = 0.677 \,\}$$

This is done for all four of the BOSS data sets, yielding a total of 8.000 mock catalogues. The covariance matrix is estimated as follows:

$$\hat{C}_{ij} = \frac{1}{N_m - 1} \sum_{k=1}^{N_m} \left(d_i^k - \bar{d}_i\right) \left(d_j^k - \bar{d}_j\right), \tag{2.4.1}$$

where N_m is the number of mocks, while d_i^k represents the data vector (i.e., the measured multipoles of the 2PCF or 3PCF), with indexes i and k identifying the bin and the mock sample, respectively and $\bar{d}_i = (1/N_m) \sum_{k=1}^{N_m} d_i^k$. From $\hat{C}_{ij}$, the corresponding correlation matrix $\hat{c}_{ij}$ is estimated as

$$\hat{c}_{ij} \equiv \frac{\hat{C}_{ij}}{\sqrt{\hat{C}_{ii}\hat{C}_{jj}}}. \tag{2.4.2}$$

The correlation matrix for the 2PCF multipoles of the NGC high-z data set is shown in Fig. 2.10. We separated the 2PCF multipoles with a black line; thus, the blocks on the diagonal represent correlations within the multipole itself and show a higher degree of correlation with respect to off-diagonal blocks, which represent the cross-correlation between multipoles ($\xi_0\xi_2$, $\xi_2\xi_4$, $\xi_4\xi_0$).

The multipole expansion of the observed 2PCF is a useful way of data compression, as often used in the literature, and can characterize anisotropies in the 2PCF with a relatively low number of multipoles (the optimal solution would be to evaluate 2PCF multipoles up to infinite orders, which is infeasible in practice). In addition, the covariance of the 2PCF multipoles is easier to estimate, e.g., requiring a smaller number of mock catalogs, due to the dimension reduction of the data vector.

For the 3PCF the measured multipoles do not represent the anisotropy in the clustering with respect to the l.o.s. rather they are the estimated quantity of the SHD algorithm. The length of the data vector of the 3PCF multipoles amounts to $N_\ell \times N_{bins}^2$; in our case $N_\ell = 11$, therefore yielding a data vector of size $\sim 10^4$. Exploiting the full potential of these multipoles requires a large number of realizations to get a reliable estimate of the covariance matrix, which we simply lack.

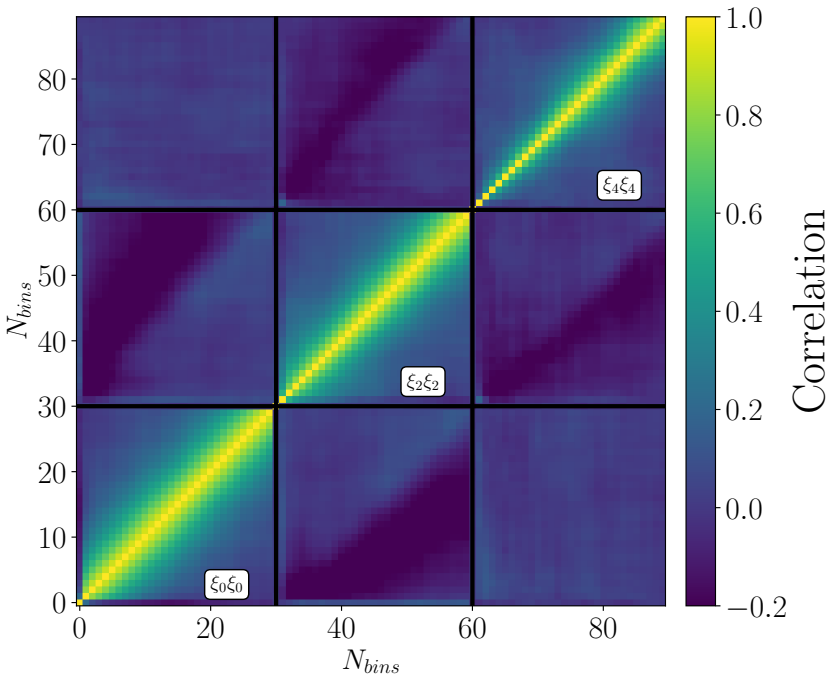

Figure 2.10: The correlation matrix of the 2PCF multipoles for scales between 0 and $150\,h^{-1}Mpc$, of the NGC high-z data set. Each multipole is separated by a black line; the 3 diagonal blocks are correlations within the same multipole and the 3 off-diagonal blocks are correlations between multipoles.

To overcome this issue, we re-sum the multipoles using Eq. 2.3.1 and 2.3.2. The estimated quantities are called the single-scale 3PCF obtained by fixing two triangle sides and varying the third one. The covariance of single-scale 3PCF is estimated from the covariance of the multipoles as:

$$\hat{C}_{\theta_i\theta_j}(r_{12}, r_{13}) = \sum_{\ell,\ell'=0}^{\ell_{\max}} \mathcal{L}_\ell(\cos\theta_i)\, \hat{C}_{\ell,\ell'}(r_{12}, r_{13})\, \mathcal{L}_{\ell'}(\cos\theta_j) \tag{2.4.3}$$

where the quantity $\hat{C}_{\ell,\ell'}(r_{12}, r_{13})$ is the estimated covariance of the multipoles $\zeta_\ell(r_{12}, r_{13})$. The covariance of all scale 3PCF is obtained from the covariance of the 3PCF multipoles through the following expansion in Legendre polynomials:

$$\hat{C}_{ij} = \sum_{\ell,\ell'=0}^{\ell_{\max}} \mathcal{L}_\ell(r_{12}, r_{13}, r_{23})\, \hat{C}_{\ell,\ell'}(r_{12}, r_{13}; r'_{12}, r'_{13})\, \mathcal{L}_{\ell'}(r'_{12}, r'_{13}, r'_{23}). \tag{2.4.4}$$

where the quantity $\hat{C}_{\ell,\ell'}(r_{12}, r_{13}; r'_{12}, r'_{13})$ denotes the cross-covariance between the multipoles $\zeta_\ell(r_{12}, r_{13})$ and $\zeta_{\ell'}(r'_{12}, r'_{13})$, and i, j represent the triangle indices associated to the triplets (r_{12}, r_{13}, r_{23}) and $(r'_{12}, r'_{13}, r'_{23})$.

The correlation matrices of the all-scales ζ and Q for all 1898 triangles are shown in Fig. 2.11. The pattern that appears is due to the choice of scale ordering; each new block

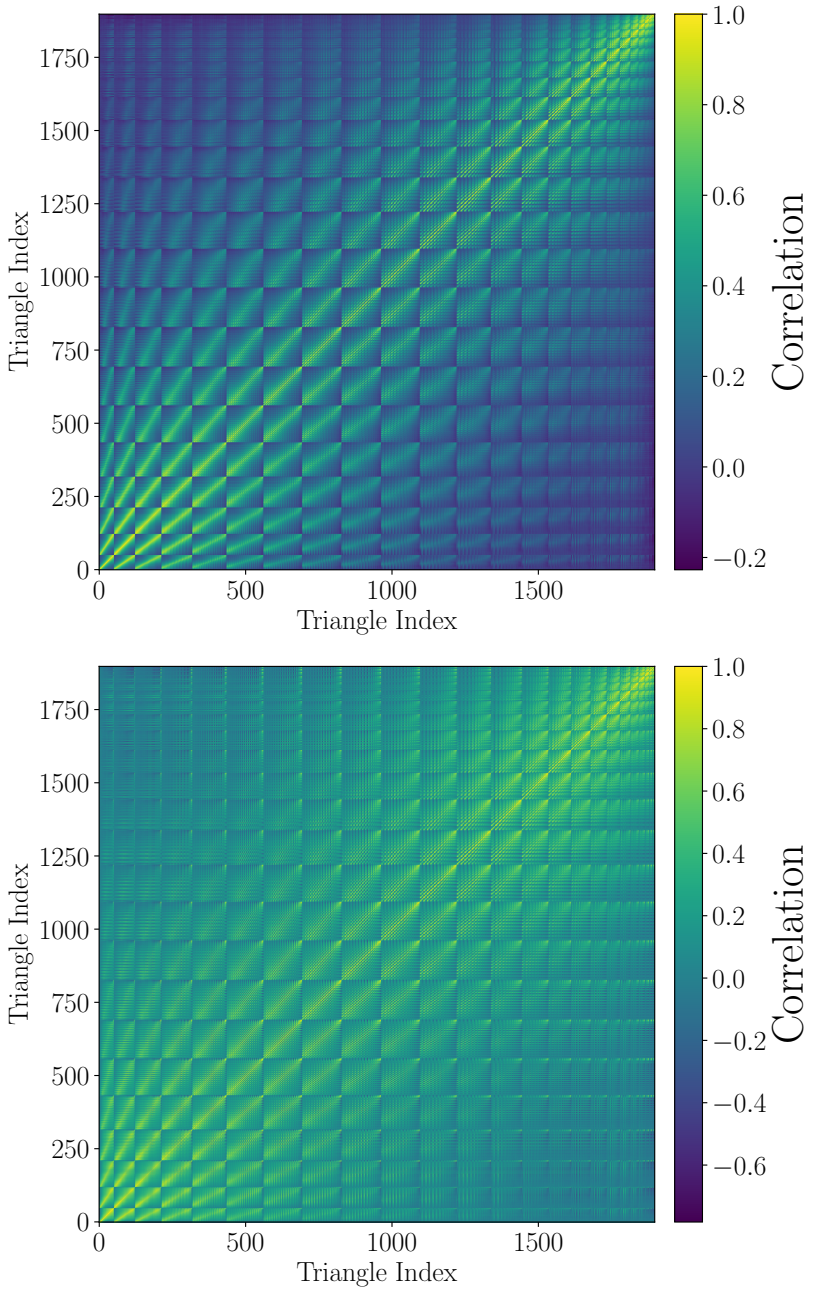

Figure 2.11: The correlation matrices of the connected 3PCF (top) and reduced 3PCF (bottom) for all triangles (1898) between 0 and 130 h^{-1}Mpc, of the NGC high-z data set.

contains a different minimum scale r_{min}. The correlation matrix of Q shows a higher degree of correlation at all scales, which is to be expected since Q is defined by the mixing of scales between the 2PCF and 3PCF.

In fact, the covariance of the single-scale and all-scales reduced 3PCF Q, must be estimated by using error propagation (Cowan 1998). It is known that if $\boldsymbol{x} = (x_1, \ldots, x_n)$ is a set of n random variables whose estimated means and covariance matrix are respectively $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_n)$ and $\hat{V}$, then the covariance matrix $\hat{C}$ of a set of m functions $y_1(\boldsymbol{x}), \ldots, y_m(\boldsymbol{x})$ of $\boldsymbol{x}$ can be estimated to first order as

$$\hat{C} = J\hat{V}\,J^T, \tag{2.4.5}$$

where

$$J_{ij} \equiv \left(\frac{\partial y_i}{\partial x_j}\right)_{\boldsymbol{x}=\boldsymbol{\mu}}. \tag{2.4.6}$$

Applying the procedure to our case, we construct the $\hat{V}$ matrix from the data vector

$$\boldsymbol{x} \equiv (\zeta(\theta_1), \ldots, \zeta(\theta_n), \xi_0(\theta_1), \ldots, \xi_0(\theta_n)),$$

for the single scale Q evaluated at the angles $\theta_1, \ldots, \theta_n$. We have omitted the dependence on r_{12} and r_{13} in ζ since the sides are fixed and we have indicated with $\xi_0(\theta)$ the 2PCF monopole evaluated at a separation equal to the length of the side opposite to the θ angle. In the all-scales approach, if the triangles are ordered with indices $1, \ldots, n$ and the centers of the radial bins are $r_1, \ldots, r_{max}$, we define $\boldsymbol{x}$ to be

$$\boldsymbol{x} \equiv (\zeta_{\text{all},1}, \ldots, \zeta_{\text{all},n}, \xi_0(r_1), \ldots, \xi_0(r_{max})),$$

The $\hat{V}$ matrix constructed from this vector is the covariance matrix we use to fit the joint 2- and 3-point correlation functions (with the added dependence on ξ_2 and ξ_4). Combining the two statistics this way provides the minimum amount of scale mixing. The non-trivial dependence of Q on ξ_0 and ζ mixes scale dependencies, increasing uncertainties over all scales. In practice, the error propagation requires the computation of a Jacobian matrix, encoding the dependence of Q on ξ_0 and ζ. The relatively simple expression of Q as a function of ζ and ξ_0 makes it possible to compute analytically the derivatives needed to construct the Jacobian matrix J. Given all scales $Q_{\text{all},1}, \ldots, Q_{\text{all},n}$ where each triangle t is characterized by triangle side lengths (r_i, r_j, r_k), the jacobian matrix is constructed as

$$\begin{aligned}
\frac{\partial Q_t}{\partial \zeta_t} &= \frac{1}{\xi(r_i)\xi(r_j) + \xi(r_j)\xi(r_k) + \xi(r_k)\xi(r_i)}, \\
\frac{\partial Q_t}{\partial \xi(r_i)} &= -\frac{\zeta_t[\xi(r_j) + \xi(r_k))]}{\{\xi(r_i)[\xi(r_j) + \xi(r_k)] + \xi(r_j)\xi(r_k)\}^2}, \\
\frac{\partial Q_t}{\partial \xi(r_w)} &= 0
\end{aligned} \tag{2.4.7}$$

for all t triangles from 1 to n.

The ultimate goal of this work is to extract relevant cosmological parameters from a Bayesian analysis. Uncertainty estimates are drawn from the inverse covariance, and the

invertibility of a matrix is dependent on the number of realizations from which it has been estimated. It can be demonstrated that if the number of mock catalogues from which the covariance is estimated is smaller than the data vector, the resulting matrix is singular (Hartlap et al. 2007). However, even if $N_{mocks} > N_{bins}$, the inverse covariance matrix is biased; that is beacause the inverse of an unbiased estimator for the covariance matrix is, in general, not unbiased, leading to the underestimation of confidence regions. To account for it, we apply the Hartlap factor to inverse covariances:

$$\hat{C}^{-1} = \frac{N_{mocks} - N_{bins} - 2}{N_{mocks} - 1}\hat{C}_*^{-1} \qquad \text{if} \quad N_{bins} < N_{mocks} - 2 \tag{2.4.8}$$

It is likely that even when this condition is satisfied ($N_{mocks} > N_{bins}$), finite realizations for the estimation of covariance lead to modes (eigenvalues) that are uninformative, being close to 0 but not quite so. Inverting them would yield a larger noise than we might expect. To overcome this issue we adopt a pseudo-inversion procedure that inverts only modes that bring significant information, leading to a more stable approximation of the inverse covariance. The byproduct of this procedure is a more accurate estimation of the data vector size that we use for the computation of Harlap factors and goodness of fit by means of the reduced χ^2.

We compute the Moore-Penrose pseudo-inverse matrix by means of Singular Value Decomposition (SVD). It is possible to prove that if A is any matrix, it can be decomposed as

$$A = U\Sigma V^T$$

than the pseudo-inverse A^+ can be written as

$$A^+ = V\Sigma^+ U^T$$

where V and U are orthogonal matrices, Σ is a diagonal matrix consisting of A's so-called singular values (the eigenvalues for a square matrix), followed, typically, by zeros. Then Σ^+ is simply the diagonal matrix consisting of the reciprocals of A's singular values again, followed by zeros.

We compute the pseudo-inverse by keeping only the eigenvalues that have reached a certain cumulative variance threshold. If λ_i are the eigenvalues of the covariance matrix, from largest to smallest, then the cumulative variance of the first k modes is

$$\mathrm{Var_k} = \frac{(\sum_{\mathrm{i=1}}^{\mathrm{k}} \lambda_\mathrm{i})}{\sum_{\mathrm{i=1}}^{\mathrm{N_{bins}}} \lambda_\mathrm{i}}.$$

By only keeping eigenvalues that contribute to 99.9% of the total variance, we effectively remove badly conditioned modes. The number of eigenvalues that do not get cut by this threshold is the effective length of the data vector we use to compute Hartlap factors and degrees of freedom for the goodness of fit (χ^2 test).

Chapter 3

Development and validation of a Bayesian analysis framework

In this chapter we describe the framework we constructed to perform a Bayesian analysis of the BOSS two- and three-point correlation functions within the template-fitting approach. We also present a series of tests designed to assess the stability of the analysis pipeline and the ability of the data to constrain the cosmological parameters of interest, with the ultimate goal of extracting cosmological information from the BOSS measurements. Sec. 3.1 describes the statistical methodology and the construction of the analysis pipeline, which combines the measured clustering statistics with the corresponding theoretical models. In Sec. 3.2 we introduce and motivate the use of synthetic data sets for validation purposes and describe the procedure adopted to fit them. Sec. 3.3 discusses the parameter degeneracies that arise in the clustering statistics when fitting synthetic data, while Sec. 3.4 presents strategies to mitigate these degeneracies. Unless otherwise stated, throughout this chapter we adopt a fiducial cosmological model consistent with Planck Collaboration: Aghanim et al. (2020) for the computation of the theoretical templates (see Section 1.4).

3.1 Bayesian framework for posterior sampling

In the previous chapters we described how the BOSS correlation functions are measured, how their covariance matrices are estimated, and which physical model is adopted to generate theoretical predictions for a template-fitting analysis. The final step is to connect these ingredients: namely, to define how the model predictions are compared with the observational data and their associated uncertainties in order to infer the parameter values that best describe the measurements.

We build a Bayesian framework for fitting BOSS 2-point and 3-point correlation functions where the unknown posterior distribution is sampled by means of a Markov Chain Monte Carlo (MCMC) algorithm. We sample the posterior using the `emcee`[1] code (Foreman-Mackey et al. 2013), implementing an affine-invariant ensemble sampler. This

[1] https://emcee.readthedocs.io/en/stable/

application of the MCMC algorithm is well-suited to high-dimensional parameter spaces and allows us to sample the posterior distribution by choosing the number of walkers, that is, the number of points that sample the posterior at any given step of the chain. The optimal number of walkers is dependent on the number of fitting parameters; it is advised to use a number of walkers always greater than 4 times the number of fitting parameters, but in general, the higher the number of walkers the better acceptance fraction and correlation times we can achieve (Foreman-Mackey et al. 2013). Unless otherwise specified we employ 64 walkers for our chains. The second parameter to set is the number of steps the chain performs; an optimal choice is determined by the autocorrelation time, which quantifies the number of independent samples. To ensure convergence we allow the chains to run for at least 50 autocorrelation-times (Goodman and Weare 2010).

The use of MCMC sampling in Bayesian analysis is of great usefulness when dealing with high-dimensional parameter spaces and when one may want to marginalize information to estimate expectation values of model parameters. While this is standard practice in cosmology, dealing with both high-dimensions and marginalization can create several problems which go under the category of volume and projection effects. We explain them while also discussing how we tried to overcome them in Sec. 3.4.

3.1.1 The pipeline

The pipeline developed to fit BOSS clustering statistics is organized into three main directories:

`/Data` This directory allows one to load the measured correlation functions from the BOSS catalogue as well as those from the MD-`patchy` mock catalogues. The covariance matrix is estimated from 2000 mock catalogues for each *bin* {low-z, high-z} and *cap* {NGC, SGC}.

By combining measurements of ξ_0 with $\zeta^{(all)}$, we construct the reduced 3PCF $Q^{(all)}$ and its covariance matrix, as described in Sec. 2.4. These quantities are loaded through dedicated `Python classes`, which allow the user to select the desired *cap*, *bin*, range of scales $r_{\rm min} \leq r \leq r_{\rm max}$, the minimum η value for the 3PCF ($\eta_{\rm min}$), and the multipoles of the 2PCF from the set $\{\xi_0, \xi_2, \xi_4, (\xi_0 - \xi_2 - \xi_4)\}$.

`/Model` This directory contains the implementation of the theoretical models used to compute predictions for the 2PCF multipoles and the 3PCF, described in Sec. 1.4. The theoretical predictions are generated using the emulator code `MElCorr` Guidi et al. (including L. Cavazzini) 2026, in prep. This code provides predictions for the anisotropic multipoles of the two-point correlation function (2PCF), ξ_0, ξ_2, and ξ_4, as well as for the isotropic connected three-point correlation function (3PCF).

`MElCorr` implements an emulator for the 2PCF multipoles, enabling rapid evaluation of many cosmological models within a Bayesian analysis. Alcock-Paczynski (AP) rescaling modifies the shape of the power spectrum. In our template-fit approach, this requires a re-computation of the 2PCF templates, which need to be Fourier transformed at every step. Although this rescaling increases the computational cost

of evaluating 2PCF predictions, the emulator reduces the evaluation time of a single cosmological model by a factor of 10^4.

The 3PCF model does not include AP parameters and therefore does not require emulation. Its Legendre coefficients are computed up to $\ell_{\max} = 10$, matching the SHD-based estimation performed on the BOSS catalogue (see Chap. 2). The coefficients ζ_ℓ are then resummed using Eq. 2.3.2 to obtain $\zeta^{(all)}$. Finally, $\zeta^{(all)}$ is combined with ξ_0 to produce theoretical predictions for the reduced 3PCF $Q^{(all)}$.

The user can specify the minimum and maximum scales, as well as the value of $\eta_{\min}$, for each statistic. The corresponding methods return the theoretical prediction for a given set of model parameters $\boldsymbol{\theta}$ over the selected range of scales.

`/Fit` This directory combines the measured data vectors with the theoretical predictions in order to perform posterior sampling. The code is organized around dedicated `Python classes` corresponding to the different correlation functions considered in the analysis: the 2PCF multipoles ξ_ℓ, the all-scales connected 3PCF $\zeta^{(all)}$, the single-scale connected 3PCF $\zeta^{r_{12},r_{13}}(\theta)$, the all-scales reduced 3PCF $Q^{(all)}$, the single-scale reduced 3PCF $Q^{r_{12},r_{13}}(\theta)$, and the joint analysis of 2PCF and all-scales 3PCF.

Each class can fit the BOSS correlation functions either for a single *cap* and *bin* (single mode) or for multiple *caps* and *bins* simultaneously (combined mode). In the latter case, the log-likelihoods of the individual datasets are summed at each step of the Markov chain:

$$\ln \mathcal{L}(\boldsymbol{\theta}_{\mathcal{J}}) = \sum_{i=1}^{N_s} \ln \mathcal{L}_{\text{sample}_i}(\boldsymbol{\theta_i}),$$

where the sum runs over the datasets in
$\{\text{NGC high-z}, \text{SGC high-z}, \text{NGC low-z}, \text{SGC low-z}\}$, or a chosen subset.

The parameters entering each likelihood can differ between datasets. In a combined analysis, one may choose to fit parameters independently for each dataset, enforce a common set of parameters across all datasets, or adopt an intermediate scheme (e.g., splitting parameters by redshift). This choice can be made independently for each free parameter.

For instance, fitting the linear bias parameter b_1 as a single shared value across all datasets would be incorrect, since linear bias depends on redshift. Instead, b_1 should be split by redshift, resulting in two distinct values when all BOSS datasets are combined. Parameters can also be split across the four individual datasets, as done by Ivanov et al. (2020). This approach is motivated by the fact that the NGC and SGC caps correspond to spatially separated regions of the sky that may probe slightly different galaxy populations. However, as discussed in Sec. 2.2.2, we adopt a common parameter set for NGC and SGC and therefore combine them in our analysis. All possible parameter-splitting configurations supported by the pipeline are illustrated in Fig. 3.1.

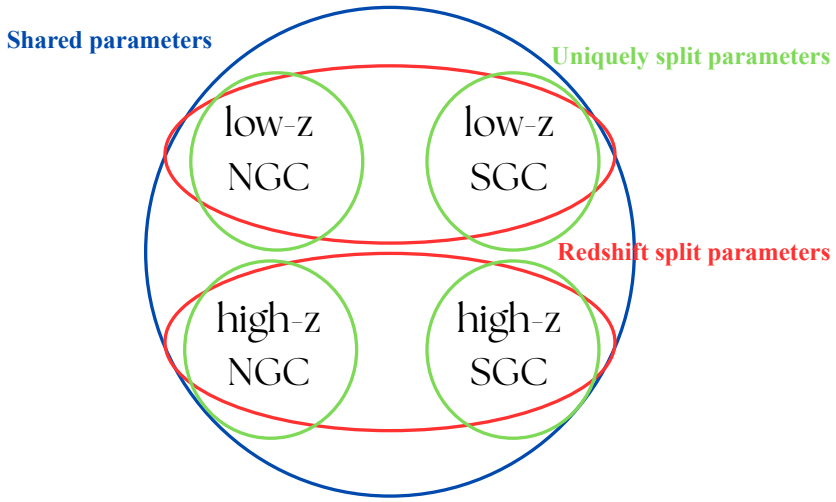

Figure 3.1: Possible choices for parameter splitting in the combined BOSS fitting procedure.

A third option supported by the pipeline is to fit **synthetic** datasets, by providing an external data vector and covariance matrix. The definition and role of these synthetic datasets in our analysis are discussed in the next section.

Finally, the pipeline allows for flexible choices of priors. Parameters may have *uniform* priors $\mathcal{U}$ to remain uninformative while excluding unphysical regions of parameter space, *Gaussian* priors $\mathcal{N}$ with specified mean and variance, or be *fixed* $\mathcal{F}$ if they are not included in the fit. In the latter case, the fixed parameter value must be provided.

An overview of the pipeline and its main features is shown in Fig. 3.2.
Our 2PCF and 3PCF models are characterized by different sets of free parameters that we infer, in the Bayesian framework, from comparisons with measurements. In pracice the parameter vector $\boldsymbol{\theta}$ is obtained by maximizing the posterior probability $\mathcal{P}(\mu(\boldsymbol{\theta})|\mathbf{d}) \propto \mathcal{L}(\mathbf{d}|\mu(\boldsymbol{\theta}))P(\boldsymbol{\theta})$, where $\mu(\boldsymbol{\theta})$ is the $\boldsymbol{\theta}$-dependent model prediction and $\mathbf{d} \equiv d(x_i)$ is the data vector. The quantity $P(\boldsymbol{\theta})$ represents the prior probability and $\mathcal{L}(\mathbf{d}|\mu(\theta))$ is the likelihood function. We adopt a Gaussian likelihood function for both the 2PCF and the 3PCF measurements. We have then

$$\ln \mathcal{L}_\xi(\boldsymbol{\theta}) = -\frac{1}{2}\chi^2_\xi(\boldsymbol{\theta}) \tag{3.1.1}$$

where the χ^2 for the 2PCF is given by

$$\chi^2_\xi(\boldsymbol{\theta}) = \sum_{i,j=1}^{N_d} [\hat{\xi}(r_i) - \xi(r_i, \boldsymbol{\theta})]\, C^{-1}_\xi(r_i, r_j)\, [\hat{\xi}(r_j) - \xi(r_j, \boldsymbol{\theta})] \tag{3.1.2}$$

with $\xi(r, \boldsymbol{\theta})$ and $\hat{\xi}(r)$ representing respectively the 2PCF theoretical prediction and its measurements. The sum runs over the separation bins r_i and r_j and N_d is the total number of bins, corresponding to the size of the data vector. Similarly, for the 3PCF, we have

$$\ln \mathcal{L}_\zeta(\boldsymbol{\theta}) = -\frac{1}{2}\chi^2_\zeta(\boldsymbol{\theta}) \tag{3.1.3}$$

with

$$\chi^2_\zeta(\boldsymbol{\theta}) = \sum_{i,j=1}^{N_d} [\hat{\zeta}(t_i) - \zeta(t_i, \boldsymbol{\theta})]\, C_\zeta^{-1}(t_i, t_j)\, [\hat{\zeta}(t_j) - \zeta(t_j, \boldsymbol{\theta})] \tag{3.1.4}$$

with the sum now extending over all triangular configurations. where each triangle is denoted by t_i being a unique triangle configuration with sides lenghts $\{r_{12}, r_{13}, r_{23}\}$. The same procedure applies to the reduced 3PCF; we do not report it for redundancy. We also perform a Bayesian analysis of the joint measurement of the 2PCF and 3PCF. We exploit the full covariance matrix, where the cross-covariance between ξ and ζ is included. The joint 2PCF+3PCF likelihood is

$$\ln \mathcal{L}_{\xi+\zeta}(\boldsymbol{\theta}) = -\frac{1}{2}\chi^2_{\xi+\zeta}(\boldsymbol{\theta}) \tag{3.1.5}$$

with

$$\chi^2_{\xi+\zeta}(\boldsymbol{\theta}) = \sum_{i,j=1}^{N_d} [\hat{d}(x_i) - d(x_i, \boldsymbol{\theta})]\, C_{\xi+\zeta}^{-1}(x_i, x_j)\, [\hat{d}(x_j) - d(x_j, \boldsymbol{\theta})] \tag{3.1.6}$$

where we used d to denote the data vector and model prediction vector, which is a combination of ξ data bins and ζ triangles. If the triangles are ordered with indices $1, \ldots, n$ and the centers of the radial bins are $r_1, \ldots, r_m$, we define $d(\boldsymbol{x})$ to be

$$d(\boldsymbol{x}) \equiv \{\zeta_{\mathrm{all},1}, \ldots, \zeta_{\mathrm{all},n}, \xi_0(r_1), \ldots, \xi_0(r_m), \xi_2(r_1), \ldots, \xi_2(r_m), \xi_4(r_1), \ldots, \xi_4(r_m)\}.$$

The $\boldsymbol{\theta}$ vector is a combination of the ξ and ζ parameters that are either shared by the theoretical models of both functions or not.
The Harlap factor in 2.4.8 does not correct for errors in the covariance matrix, which propagate through to errors on parameters estimated from the posterior probability. To account for this effect, we follow Percival et al. (2014) and apply the following correction factor on the parameters' variance:

$$m_1 = \frac{1 + B(N_d - N_p)}{1 + A + B(N_p + 1)} \tag{3.1.7}$$

where N_p is the number of fitting parameters and

$$A = \frac{2}{(N_m - N_d - 1)(N_m - N_d - 1)} \tag{3.1.8}$$

$$B = \frac{N_m - N_d - 2}{(N_m - N_d - 1)(N_m - N_d - 4)}. \tag{3.1.9}$$

Just like the factor present in the inverse covariance m_1 depends on the number of mocks used to estimate uncertainties in the data, an increase in N_m will decrease the bias in the estimated parameters.

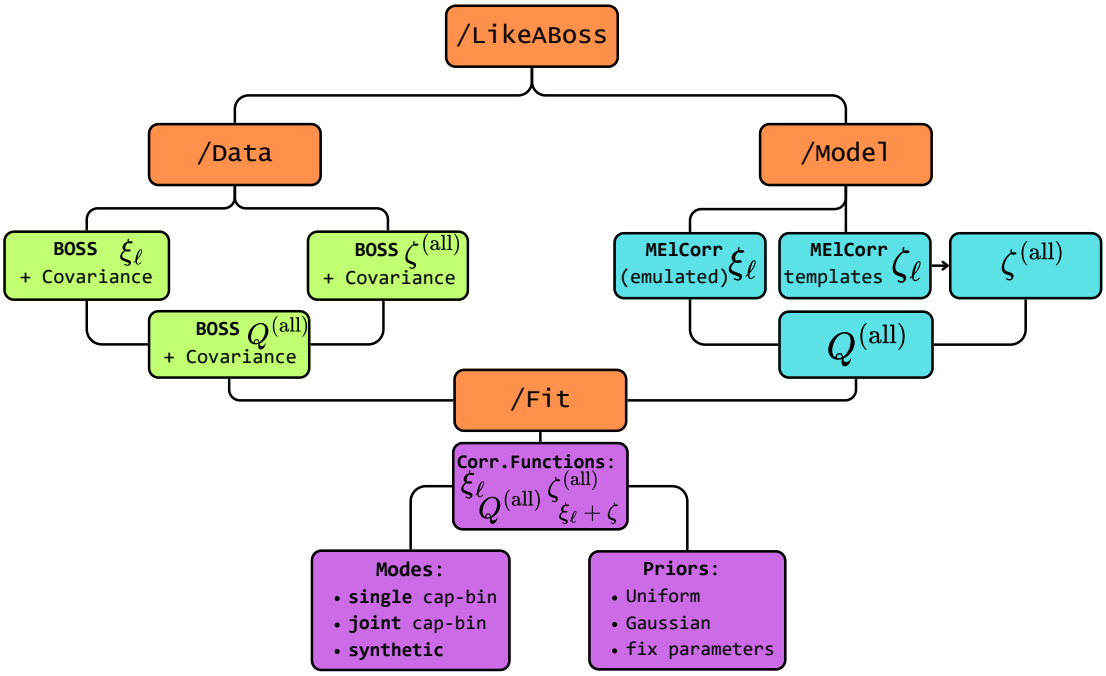

Figure 3.2: Flowchart of the pipeline used to fit BOSS correlation functions. Single-scales ζ and Q can also be fit, but we have omitted their presence in the figure, since our Bayesian analysis focuses on the all-scales ζ and Q.

3.2 Analysis of synthetic mock data

We define *synthetic data* as a data vector constructed from theoretical models of the clustering statistics, which is then used in the fitting procedure. In this case, the χ^2_ξ statistic is defined as

$$\chi^2_\xi(\boldsymbol{\theta}) = \sum_{i,j=1}^{N_d} \left[\xi_i(\hat{\boldsymbol{\theta}}) - \xi_i(\boldsymbol{\theta})\right] C^{-1}_{i,j} \left[\xi_j(\hat{\boldsymbol{\theta}}) - \xi_j(\boldsymbol{\theta})\right], \tag{3.2.1}$$

where $\hat{\boldsymbol{\theta}}$ denotes the parameter vector used to generate the theoretical prediction for ξ. In this framework, the model prediction $\xi_i(\hat{\boldsymbol{\theta}})$ replaces the measured BOSS two-point correlation function $\hat{\xi}(r_i)$. Since the synthetic data are generated from a known parameter vector, the "true" parameter values are known a priori. This allows us to perform controlled tests of the fitting procedure. The synthetic data vector is paired with a covariance matrix measured from one of the BOSS data-sets; unless otherwise specified, we adopt the NGC high-z covariance matrix. The motivation for this procedure is twofold:

- First, constructing a synthetic data vector for which the input parameters are known allows us to validate the pipeline. When fitting synthetic data with a covariance matrix artificially reduced by a factor of 10^{-3}, we expect the resulting constraints to be tightly centered on the "true" parameter values if the pipeline is functioning correctly.

- Second, the choice of adopting the BOSS covariance matrix is not arbitrary. Using a realistic covariance allows us to evaluate the constraining power of a given clustering statistic for a specific BOSS data set. Consequently, when fitting the actual BOSS measurements, we already have a clear expectation of the achievable parameter constraints.

In Fig. 3.4 we show the 2D-marginalised posteriors distribution coming from fitting a synthetic 2PCF multipoles data vector: $\{\xi_0 - \xi_2 - \xi_4\}$, adopting $r_{min} = 20\,h^{-1}$Mpc to $r_{max} = 150\,h^{-1}$Mpc. In this case we have used a BOSS covariance reduced by a factor 10^{-3}. The red lines in the figure show the value of the parameters used to generate the synthetic data set, corresponding to:

$$\{b_1 = 2.0,\ b_2 = 0.5,\ b_{\mathcal{G}2} = 0,\ b_{\Gamma 3} = 0,\ f = 0.78,\ \sigma_8 = 0.60,\ \alpha = 1.0,\ \epsilon = 0\}.$$

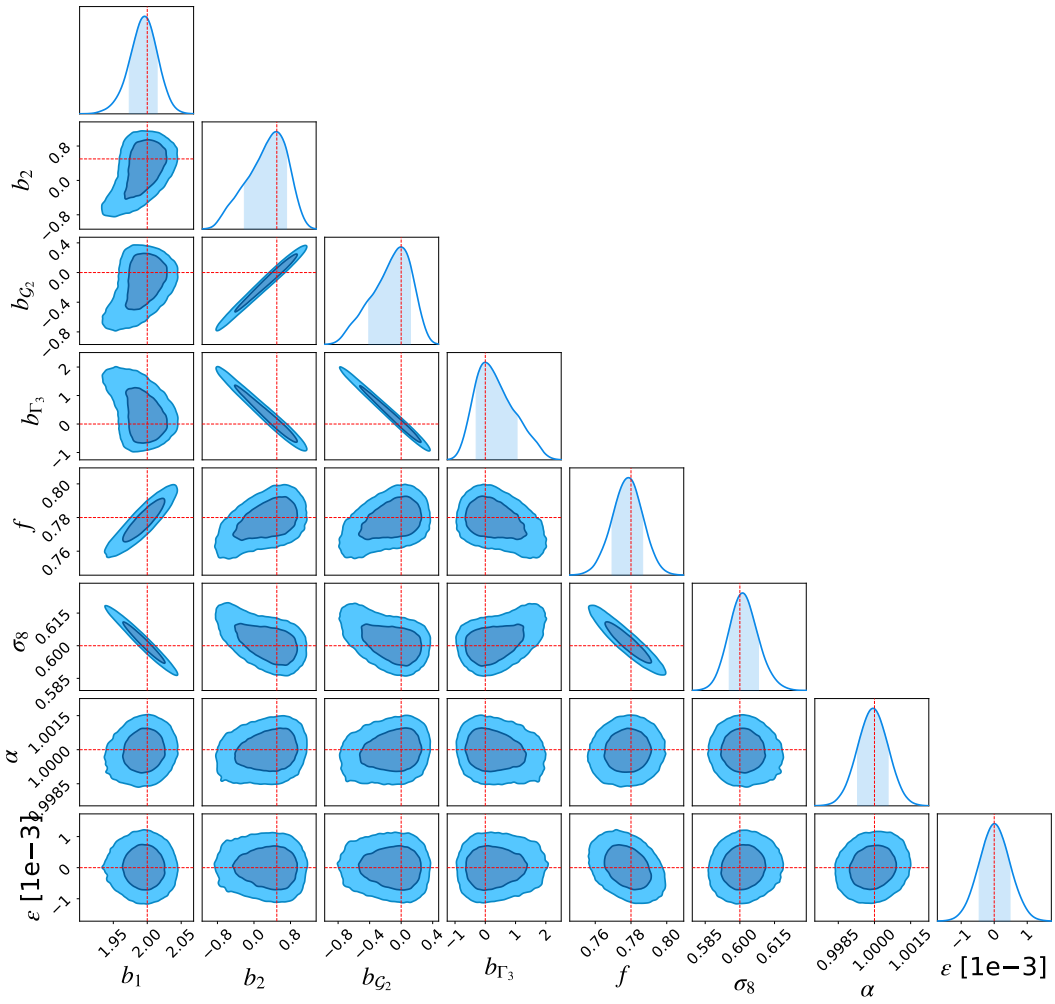

Figure 3.3: 2D marginalized posteriors of a 2PCF multipoles $\ell = 0, 2, 4$ synthetic data vector. The parameters used to generate the 2PCF multipoles are shown through the red line. The covariance matrix used is that of NGC high-z multiplied by a factor 10^{-3}. the minimum and maximum scales are $r_{min} = 20\,h^{-1}Mpc$ and $r_{max} = 150\,h^{-1}Mpc$. Contours show the 68% and 95% confidence regions.

The two-dimensional marginalized contours are centered on the true parameter values. Parameters that affect model predictions of ξ or ζ in a similar way are said to be *degenerate*, as different combinations of such parameters can yield the same posterior probability. This manifests in 2D contours as more or less tilted and stretched ellipses. If two parameters are positively correlated, increasing one parameter will result in the same probability when the other is also increasing (ellipses stretching from the lower left to upper right corners). If the two are negatively correlated, an increase in one will give the same probability for the decrease of the other (ellipses stretching from the upper left to lower right corners). In some cases, particularly for higher-order bias parameters, the contours exhibit curved or 'banana-shaped' features, reflecting non-Gaussianities in the marginalized posterior distributions.

Standard MCMC algorithms can struggle to efficiently sample highly non-Gaussian parameter spaces. However, the affine-invariant ensemble sampler adopted in this analysis is well suited to handling such skewed or stretched probability densities.

Focusing on the more elliptical contours of Fig. 3.3, we can notice two degeneracies for the 2PCF multipoles. The first involves the higher-order bias parameters b_2, $b_{\mathcal{G}2}$, and $b_{\Gamma 3}$, with a weaker coupling to b_1. The second one is between $b_1 - f - \sigma_8$. Although these degeneracies can be disentangled in the present analysis, it is important to note that the reduced covariance matrix used here leads to unrealistically tight constraints. As such, these results should not be taken as representative of the constraints expected when using the full BOSS covariance.
The same procedure is repeated for the 3PCF and the reduced 3PCF to ensure that the pipeline is working correctly.

3.3 The $b_1 - f - \sigma_8$ degeneracy

The use of synthetic data with the full covariance matrix allows us to establish what can and cannot be properly constrained. In Fig. 3.4 we show the 2D marginalized constraints of the synthetic data vector composed of the 2PCF multipoles $\{\xi_0 - \xi_2 - \xi_4\}$ from $r_{min} = 20\,h^{-1}$Mpc to $r_{max} = 150\,h^{-1}$Mpc. The blue contours have been obtained with synthetic data with the following parameters

$$\{b_1 = 2.0,\ b_2 = 0.5,\ b_{\mathcal{G}2} = 0,\ b_{\Gamma 3} = 0,\ f = 0.78,\ \sigma_8 = 0.60,\ \alpha = 1.0,\ \epsilon = 0\}$$

while the orange contours are obtained with the same set of parameters but σ_8 is not allowed to vary in the fitting procedure and is instead fixed to its true value ($\sigma_8 = 0.60$).

Both the blue and orange contours are able to find the true value of the higher-order bias parameters, with the orange ones being more peaked around it. Their 1D marginalized constraints deviate significantly from Gaussianity. Clearly b_2, $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$ are poorly constrained by the 2PCF alone. We will break this degeneracy by combining the 2PCF with the 3PCF, which contains shape information from triangles that is complementary to that of the 2PCF.

The AP parameters are largely independent of the other model parameters and exhibit no significant differences between the two fits. This behavior is expected because the AP

parameters primarily encode transverse geometric information (see Fig. 1.3), whereas parameters such as b_1 and σ_8 primarily affect the amplitude in a scale-independent way.

A more striking feature of Fig. 3.4 is the pronounced lack of constraints on b_1, f, and σ_8 for the blue contours. The degeneracy drives b_1 toward very large values while pushing σ_8 to correspondingly small values, until both parameters saturate their prior bounds. The f contours are likewise shifted toward higher values relative to expectations. For brevity, we do not display the results for the synthetic 3PCF, reduced 3PCF, or the joint 2PCF+3PCF analysis, but the same pathological degeneracy is present in all cases.

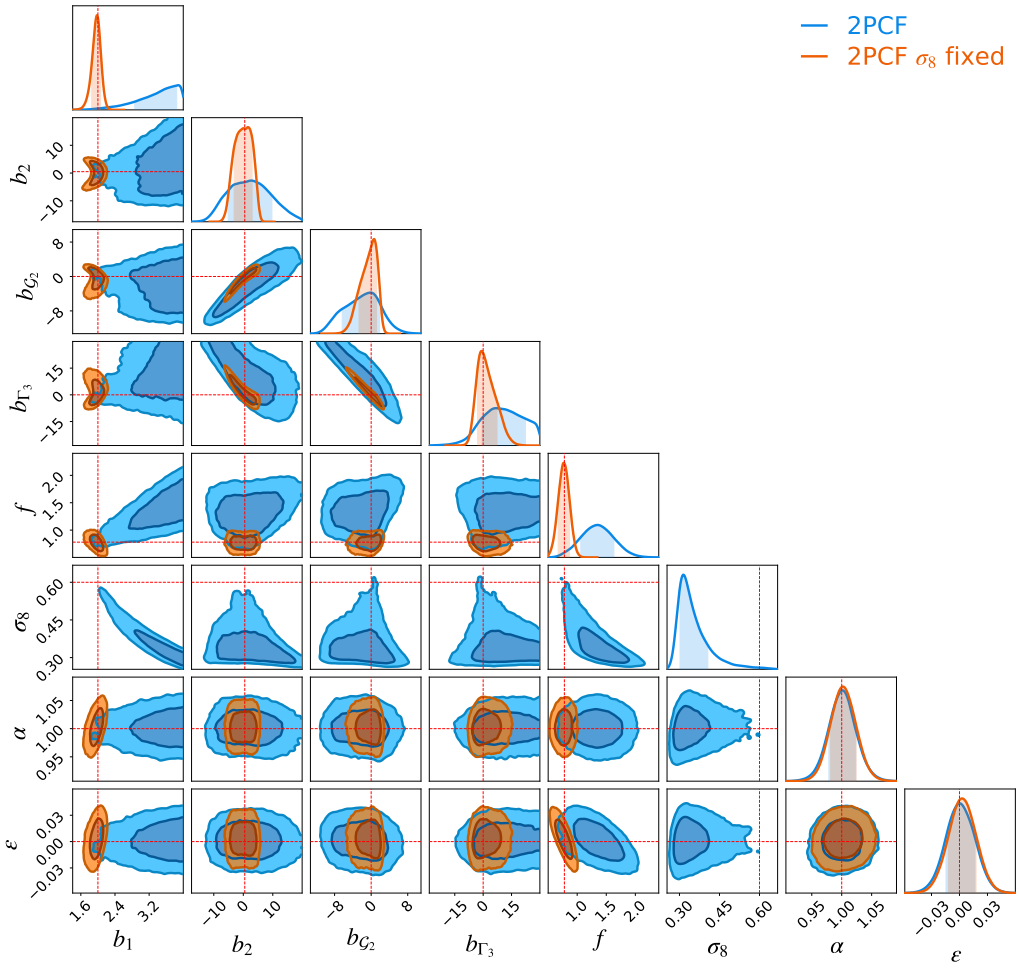

Figure 3.4: 2D marginalized posteriors of 2PCF multipoles $\ell = 0, 2, 4$ coming from a synthetic fit. The parameters used to generate the 2PCF multipoles are shown through the red line. The covariance matrix used is that of NGC high-z. The minimum and maximum scales are $r_{min} = 20\ h^{-1}$Mpc and $r_{max} = 150\ h^{-1}$Mpc. Contours show the 68% and 95% confidence regions. The blue contours are made while varying σ_8. In the orange contours σ_8 is fixed to the true value.

These findings motivate two distinct analyses of the BOSS samples. In the first, σ_8 is held fixed. In the second, σ_8 is treated as a free parameter, and additional strategies are introduced to lift the degeneracy. The rationale for these approaches and their

implementation are detailed in the following section.

3.4 Strategies to remove volume effects

Sampling algorithms such as `emcee` do not always behave as one might expect, and our interpretation of their output can change significantly when results are visualized through marginalization. The ultimate goal of MCMC methods is the estimation of expectation values. In practice, this is achieved by marginalizing the posterior distribution, i.e., integrating over subsets of parameters to project the probability density onto lower-dimensional spaces, typically 1D or 2D. However, marginalization can distort a complex, highly non-Gaussian N-dimensional posterior, leading to projection effects that are not representative of the true expectation values one aims to extract. The importance of these effects increases as the likelihood deviates from Gaussianity.

Projection effects can be quantified using profiling techniques (Raveri et al. 2024), but these are computationally expensive. Instead, we adopt a simpler approach by reporting expectation values both as posterior medians and as maximum-likelihood (ML) samples. For a multivariate Gaussian posterior, these two estimates coincide. When they differ significantly, this signals the presence of strong projection effects, indicating that marginalized posteriors cannot be trusted to provide reliable expectation values.

Another important aspect is related to the role of volume in high-dimensional parameter spaces. Sampling algorithms are intrinsically sensitive to the volume of the domain they explore. To illustrate this, consider a d-dimensional hypercube with side length l, whose volume scales as

$$V(l) = \prod_{i=1}^{d} l = l^d. \tag{3.4.1}$$

The differential volume element between l and $l + dl$ is then

$$dV(l) = d \cdot l^{d-1}\, dl. \tag{3.4.2}$$

This exponential scaling with dimensionality implies that, as d increases, volume becomes increasingly concentrated in thin shells located progressively farther from the center of the hypercube. While the numerical factors depend on the geometry, this exponential dependence on d is a generic feature of high-dimensional spaces. In practical terms, increasing the number of parameters leads to an exponential growth in the number of available parameter combinations that must be explored.

Because of how MCMC algorithms work, samples tend to concentrate in regions of high posterior mass, defined as the product of density and volume, rather than exclusively in regions of highest posterior density (Speagle 2019). As a consequence, a relatively flat (in terms of probability) region of the posterior may encompass a larger volume than the narrow region around the global maximum of the likelihood. Such regions, despite having lower density, can therefore dominate the sampled chains.

This behaviour is illustrated in Fig. 3.5, where the likelihood surface marginalized over b_1 is highly asymmetric. In this case, posterior sampling preferentially explores the

plateau of the likelihood surface, which corresponds to the region of strongest degeneracy at large values of b_1 and f. It is important to stress that the volume effects discussed here are distinct from prior volume effects commonly discussed in the literature (Raveri et al. 2024). The latter arise when posteriors are prior-dominated and marginalized distributions become distorted by the choice of prior. This is not the case in our analysis: we adopt uniform priors with sufficiently wide ranges such that the posterior is entirely likelihood dominated.

To mitigate the volume effects encountered when fitting synthetic data with the BOSS covariance matrix, we adopt several strategies, which we describe and test below. When analysing real data, however, the situation is more subtle, and the assumptions valid for synthetic data do not necessarily apply.

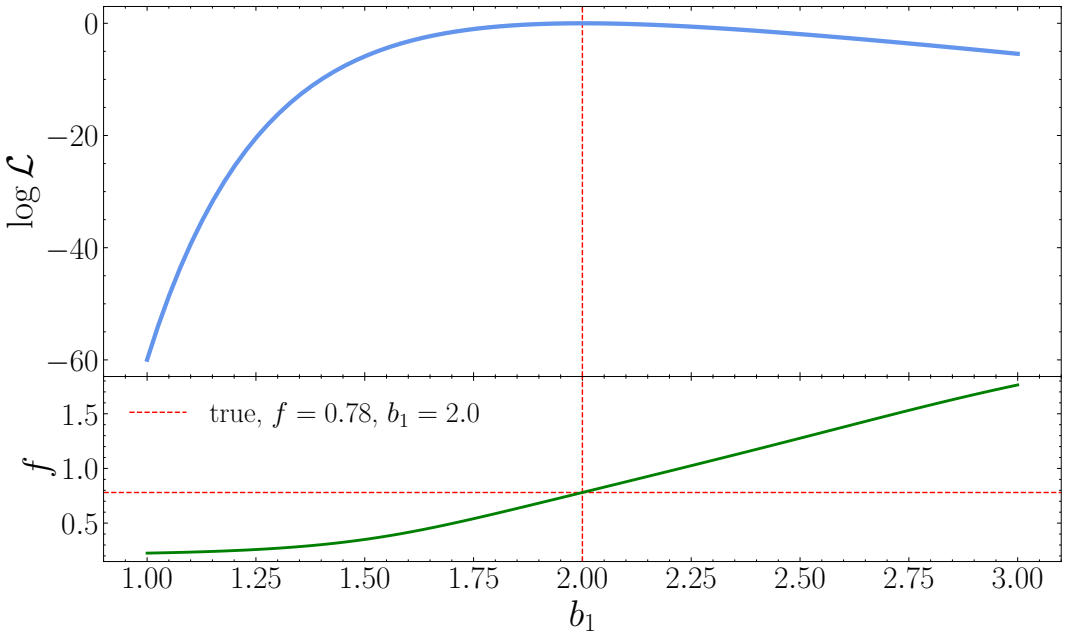

Figure 3.5: Logarithmic likelihood of a synthetic $Q^{(all)}$ data set as a function of f. The value of $\log \mathcal{L}$ is obtained by fixing b_1 to values from 1 to 3, and finding the maximum likelihood as a function of f. A BOSS covariance matrix is used to compute the likelihood. The values of the parameters used for the synthetic data are outlined in red. Other parameters of the model are kept to 0.

3.4.1 Reduction of the parameter space

In Sec. 3.3, we showed that for the 2PCF, the 3PCF, and joint 2PCF+3PCF, the $b_1 - f - \sigma_8$ degeneracy shifts marginalized posteriors to non-accurate values. At the level of the reduced 3PCF, this degeneracy persists even when σ_8 is fixed. A powerful way to suppress the associated volume effects is to reduce the dimensionality of the parameter space itself. We have already demonstrated that fixing σ_8 allows us to recover the true parameters. A natural extension is to keep σ_8 free while restricting the range or freedom of other parameters.

While highly effective, this approach is not always intuitive. Volume effects arise for two main reasons: first, the sheer size of high-dimensional spaces makes it increasingly difficult for sampling algorithms to efficiently identify narrow regions of high posterior density; second, parameter degeneracies and large uncertainties generate extended regions of moderate likelihood that, due to their large volume, can dominate the posterior mass. Higher-order bias parameters enter the 2PCF in a degenerate manner, and although the inclusion of the 3PCF helps to break this degeneracy, these parameters remain correlated with b_1, f, and σ_8.

We have explored fixing different combinations of higher-order bias parameters and find that b_2 is the most stable one, both for the reduced 3PCF and for the joint 2PCF+3PCF analysis, whereas $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$ exhibit the strongest degeneracies. Figure 3.6 shows the correlation matrix of all model parameters obtained from fitting synthetic joint 2PCF+3PCF data with fiducial values

$$\{b_1 = 2.0,\ b_2 = 0,5\, b_{\mathcal{G}2} = 0,\ b_{\Gamma 3} = 0,\ f = 0.78,\ \sigma_8 = 0.60\}.$$

Although the precise values of the correlation coefficients depend on the chosen fiducial parameters, we do not expect large qualitative variations. The matrix provides a useful diagnostic of which parameters are most strongly correlated and where volume effects are likely to arise.

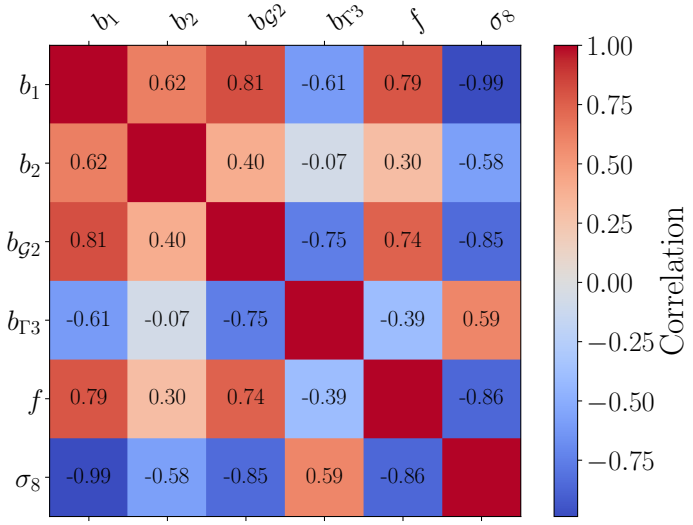

Figure 3.6: Correlation matrix of all model parameters. Obtained from fitting the joint 2PCF+3PCF with a synthetic data vector with scales $r_{min}^{2PCF} = 30h^{-1}$Mpc, $r_{min}^{3PCF} = 40h^{-1}$Mpc, $\eta_{min} = 3$.

For real data, the question of what value to fix parameters is non-trivial. Unlike synthetic data, the true parameter values are not known a priori. Fixing $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$ to zero would therefore be an arbitrary choice. We have considered fixing them to best-fit values obtained from a posterior where σ_8 is fixed to its Planck 2018 value, but this procedure is circular and effectively biases the results toward the chosen σ_8. In other words, it introduces an arbitrary assumption.

A more physically motivated approach exploits the fact that bias parameters are not independent, but are linked by relations derived from theoretical arguments or calibrated using numerical simulations. Such relations have been widely used in studies of three-point statistics to reduce the effective dimensionality of the parameter space (Euclid Collaboration: Guidi et al. 2025; Veropalumbo et al. 2022; Veropalumbo et al. 2021; Gil-Marín et al. 2016). In this work, we adopt the Lagrangian Local In Matter Density (LLIMD) prescription for the tidal bias parameters. In the Eulerian basis, these relations read

$$b_{\mathcal{G}2}(b_1) = -\frac{2}{7}(b_1 - 1), \tag{3.4.3}$$

and

$$b_{\Gamma 3}(b_1) = -\frac{1}{6}(b_1 - 1) - \frac{5}{2}b_{\mathcal{G}2} = \frac{23}{42}(b_1 - 1). \tag{3.4.4}$$

These are the simplest theoretical relations connecting b_1 to $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$. However, the LLIMD approximation is known to be in tension with high-precision simulation results, which provide clear evidence for the presence of tidal Lagrangian bias (Abidi and Baldauf 2018).

A less restrictive alternative is to impose Gaussian priors centered on the theoretical relations (Philcox and Ivanov 2022),

$$b_{\mathcal{G}2} \sim \mathcal{N}\left(-\frac{2}{7}(b_1 - 1),\ 1^2\right), \quad b_{\Gamma 3} \sim \mathcal{N}\left(\frac{23}{42}(b_1 - 1),\ 1^2\right),$$

which allows additional freedom while still assigning higher probability density to parameter combinations consistent with theoretical expectations.

The impact of these two strategies is illustrated using synthetic data with a BOSS covariance matrix in Fig. 3.7 for the reduced 3PCF and in Fig. 3.8 for the joint 2PCF+3PCF. The synthetic data vector is generated with parameters

$$\{b_1 = 2.0,\ b_2 = 0,\ b_{\mathcal{G}2} = -2/7,\ b_{\Gamma 3} = 23/42,\ f = 0.78,\ \sigma_8 = 0.60\}.$$

Remarkably, both approaches recover the true parameters with good precision, within 1σ. When the LLIMD relations are imposed exactly, volume effects are almost entirely eliminated. When Gaussian priors are applied, the degeneracy is only partially broken, leading to small shifts in the expectation values of b_1, f, and σ_8. Some volume effects also remain for $b_{\mathcal{G}2}$, which shows a tail towards negative values for Q and to positive values for 2PCF+3PCF.

We emphasize that results obtained from synthetic data should not be interpreted as a perfect proxy for real data analyses. Real data may exhibit different scatter and parameter dependencies, and in our synthetic tests, the data are constructed to exactly satisfy the bias relations under consideration. There is no guarantee that real observations will obey these relations with the same accuracy. Nevertheless, this approach yields robust results and is supported by several previous studies, making it a well-motivated and effective strategy for mitigating volume effects.

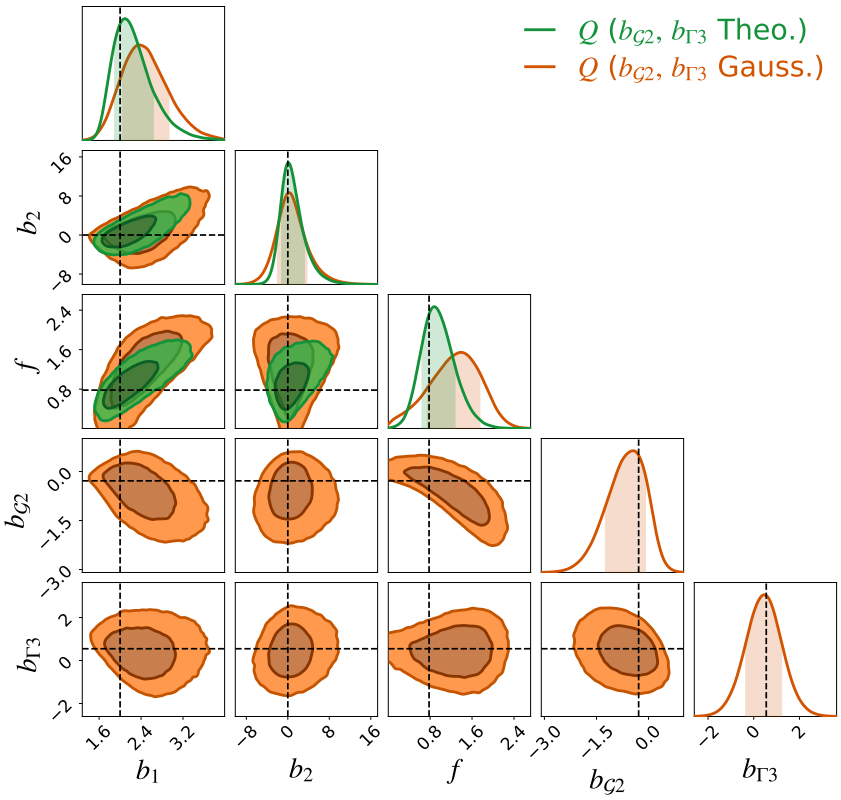

Figure 3.7: marginalized 2D constraints at confidence intervals of 68% and 95%. Obtained from fitting a synthetic reduced 3PCF with BOSS covariance. The black dashed line denotes the true parameters of the generated data. The synthetic data correspond to all triangles between $40 - 120\ h^{-1}$Mpc, $\eta_{min} = 3$. Green contours show the possterior samples of $b_1 - b_2 - f$ with $b_{\mathcal{G}2}$, $b_{\Gamma 3}$ fixed to the theoretical relation of Eq. 3.4.3 and 3.4.4. Orange contours show the posterior samples of $b_1 - b_2 - f - b_{\mathcal{G}2} - b_{\Gamma 3}$ where Gaussian priors are applied to tidal bias parameters. In both analyses σ_8 is fixed to the value of the synthetic data vector.

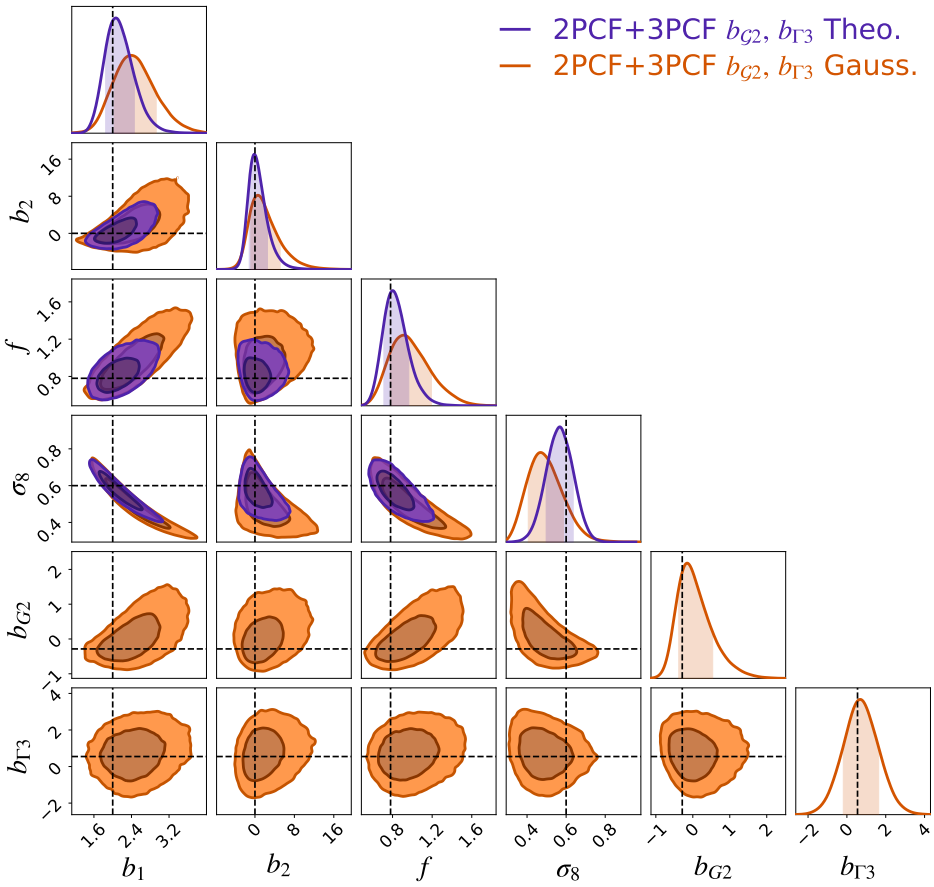

Figure 3.8: marginalized 2D constraints at confidence intervals of 68% and 95% for parameters b_1, b_2, f and σ_8. Obtained from fitting a synthetic joint 2CPF+3PCF with BOSS covariance. The black dashed line denotes the true parameters of the generated data. The synthetic data correspond to all triangles between $r^{\rm 2PCF} = 30 - 150\ h^{-1}{\rm Mpc}, r^{\rm 3PCF} = 40 - 130\ h^{-1}{\rm Mpc}$, $\eta_{min} = 3$. Purple contours show the posterior samples of $b_1 - b_2 - f - \sigma_8$ with $b_{\mathcal{G}2}$, $b_{\Gamma 3}$ fixed to the theoretical relation of Eq. 3.4.3 and 3.4.4. Orange contours show the posterior samples of $b_1 - b_2 - f - \sigma_8 - b_{\mathcal{G}2} - b_{\Gamma 3}$ where Gaussian priors are applied to tidal bias parameters.

3.4.2 Principal Component Analysis

A potential strategy to help the fit converge is based on Principal Component Analysis (PCA) (Dacunha et al. 2022). We attempt to dampen volume effects by projecting the original parameter space into a new basis in which the parameters are decorrelated. Notice that this transformation does not change the volume of the parameter space; however, it should, in principle, reduce the probability density associated with regions of moderately high likelihood. By moving to a parameter space where the plateau in the likelihood surface is damped, we hope to recover the correct parameters when fitting synthetic data, and then apply the same procedure to the real BOSS data.

PCA linearly transforms the original parameter space into a new basis where the parameters are uncorrelated. This is achieved by computing the eigenvectors of the covariance matrix of a set of posterior samples. Given a parameter space of dimension d,

PCA yields d principal components, which correspond to orthogonal directions capturing decreasing amounts of variance in the samples. These directions form an orthonormal basis in which the different dimensions of the data are linearly uncorrelated.

We attempt to fit the (synthetic) reduced 3PCF, $Q^{(all)}$, and the joint 2PCF+3PCF using PCA. For simplicity, we restrict the explanation to the reduced 3PCF but the same applies to the joint 2PCF+3PCF. We follow the steps below to generate a space of uncorrelated features:

1. We generate a set of $b_1 - f$ posterior samples by fitting a synthetic reduced 3PCF with fixed higher-order bias parameters $b_2 = 0$, $b_{\mathcal{G}2} = 0$, and $b_{\Gamma 3} = 0$, and using a BOSS covariance matrix rescaled by a factor of 10^{-3};

2. We center the values of each variable by subtracting the mean of the samples;

3. We compute the covariance matrix of the mean-centered $b_1 - f$ samples and determine its eigenvectors;

4. Since the parameter space is two-dimensional, the eigenvectors represent the direction of highest variance, corresponding to the degeneracy, and the direction orthogonal to it. From these eigenvectors we construct the linear transformation matrix, which allows us to move into a basis of uncorrelated parameters. A visual representation of this transformation is shown in Fig. 3.9.

This procedure can be extended to an d-dimensional parameter space, in which case the linear transformation matrix has dimension $d \times d$.

Once we have the ability to move between parameter spaces, we fit the reduced 3PCF again. This time, instead of fitting b_1 and f, we fit a new set of parameters, u and v, defined in the uncorrelated space. At each step of the chain, the proposed (u, v) samples are transformed back into (b_1, f) to compute the model predictions. Finally, the (u, v) samples obtained in this way are transformed back to yield the corresponding $b_1 - f$ posterior samples.

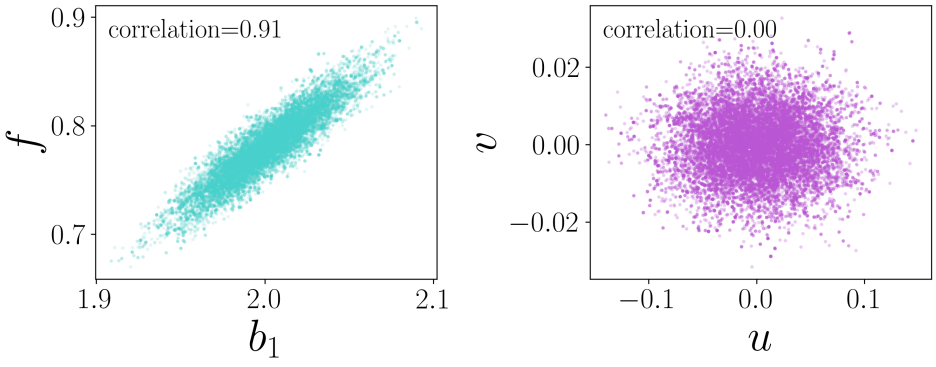

Figure 3.9: (left) Samples of the $b_1 - f$ parameter space, showing a high degree of correlation. These samples are obtained from fits to synthetic data with $b_1 = 2.0$, $f = 0.78$, and a BOSS covariance matrix reduced by a factor of 10^{-3}. (right) The same samples transformed into a decorrelated parameter space, $u - v$.

There is an important choice to be made when following this procedure, namely, which basis (i.e., linear transformation) should be used to move from one parameter space to another. The possible choices are four: constructing the basis from synthetic data with full or reduced covariance, and defining the basis using all model parameters or only a subset of them. If the goal is for the basis to provide the best possible representation of parameter degeneracies, which are intrinsic properties of the model rather than of any particular dataset, then constructing the basis from synthetic data with a strongly reduced covariance matrix, as shown in Fig. 3.9, is the most appropriate option. Conversely, if one wishes to incorporate features specific to a given dataset, a basis constructed using the full covariance matrix is better suited.

From testing all four choices, we find no clear evidence that a particular mapping performs better than the others. In fact, using this procedure, we are not able to recover the true values of the parameters $b_1 - f$ when all model parameters are allowed to vary freely.

PCA fails in this context for several reasons. As discussed above, the need to choose a single global basis significantly limits its applicability, as this choice may not be valid across the full parameter space. In other words, PCA only captures a single 'slice' of the degenerate parameter space. Moreover, while the degeneracy between b_1 and f is linear (as shown in Fig. 3.9), this is not generally true for correlations involving higher-order bias parameters, which can be non-linear. In such cases, the core assumption of PCA is violated.

Given these limitations, we do not rely on this procedure for the fitting of real data.

Chapter 4

Clustering analysis of the BOSS data

In this chapter, we present the results from Bayesian analysis of the 2-and 3-point correlation functions from the BOSS data sets.

In Sec. 4.1, we perform in-depth analyses on real BOSS measurements to assess the ranges of validity of our model, to determine in particular the best minimum scale r_{min}, maximum scale r_{max}, and η_{min} to be applied in all our final analyses. in Sec. 4.2 we summarize the results from these tests and detail the choices made for the fitting of combined BOSS samples.

We divide the analysis of the main results into five separate sections. In Sec. 4.3, we present the Bayesian analysis of the BOSS 2PCF multipoles. The main results are drawn from the expectation values of the Alcock-Paczynski parameters, from which the corresponding cosmological distance measurements are derived. In Sec. 4.4, we present the analysis on the joint 2PCF+3PCF with fixed σ_8, focusing in particular on the constraints on the bias and f parameters; the main goal of this section is to compare the results of the joint 2PCF+3PCF with the constraints obtained from the individual fits of the 2PCF and 3PCF, to assess how much the combination can help to break degeneracies and improve the accuracy on the parameters. In Sec. 4.5 we present the results from fitting jointly the 2PCF and 3PCF with a uniform prior on σ_8; thanks to a marginalization over tidal bias parameters, $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$, we are able to break the degeneracy found, and derive values of f and σ_8 independently. In Sec. 4.6, we explore and discuss the results obtained from the reduced 3PCF. Finally, in Sec. 4.7 we summarize our main findings on cosmological and bias parameters, comparing them with current constraints from other surveys.

4.1 Determining the optimal analysis regimes

We perform a series of analyses to assess the sensitivity of our results to the choice of the minimum separation scale, r_{min}, and η_{min}. The goal of these tests is to identify the range of validity of our models and the optimal scale ranges for a combined analysis in which the four BOSS samples are jointly fitted, to maximize the statistical significance of the parameters of interest.

We choose to perform these tests directly on the BOSS measurements rather than on synthetic data. This approach allows us to fully capture the intrinsic sensitivity of the

data and to quantify variations from sample to sample. Using synthetic data for such tests would not provide a proper assessment of the goodness of fit relative to the observed data, and the resulting confidence regions would depend on the assumed parameters of the synthetic data vector.

The motivation for excluding specific scales and triangle configurations is driven by two considerations: modelling accuracy and survey volume. Modelling limitations determine the smallest scales that can be reliably used, while the survey volume constrains the largest scales that can be meaningfully included. Perturbative modelling of clustering statistics breaks down at a non-linear scale that depends on several factors:

- gravitational collapse acts progressively, scales that are evolving linearly at high redshift might not evolve linearly at a lower redshift;
- in redshift space, perturbative modelling breaks down at larger scales than in real space;
- while we have a one-loop model for the 2PCF, our model for the 3PCF is only at tree-level, therefore we expect 3PCF prediction to break down at larger scales than 2PCF;
- the choice of maximum multipole affects how well isosceles triangles are reconstructed by the resummation. As described in Sec. 2.3, having $\ell_{max} = 10$ limits our ability to correctly model isosceles triangles. For this reason, we will have to cut those configurations with the η_{min} parameter.

On the other hand, the effective volume of a survey determines the precision with which clustering statistics can be measured. Larger volumes reduce uncertainties across all scales and allow the estimation of correlation functions at larger separations. While there is no intrinsic drawback to including the largest scales, they provide diminishing returns. The signal-to-noise ratio of the 3PCF above 130 h^{-1}Mpc is effectively zero. For this reason, we limit the maximum scale of the 3PCF to $r_{max} = 130\ h^{-1}$Mpc, while retaining $r_{max} = 150\ h^{-1}$Mpc for the 2PCF, where a significant signal is still present.

In practice, to determine which scales are safe to use, we fit the 2PCF multipoles and the all-scale 3PCF of each BOSS dataset as a function of the minimum separation scale $r_{min} = \{20, 30, 40, 50, 60\}\ h^{-1}$Mpc and, for the 3PCF, also as a function of $\eta_{min} = \{1, 2, 3\}$. In all cases, $\sigma_8(z)$ is fixed to its fiducial value at the corresponding redshift. From the posterior samples, we extract the maximum-likelihood point (denoted by a star), the median, and the 16th to 84th percentiles of the one-dimensional marginalized distributions of $\{b_1,\ b_2,\ b_{\mathcal{G}2},\ f\}$. To assess the goodness of fit, we compute the reduced χ^2 for each posterior using the median parameters and the associated number of degrees of freedom. Reduced χ^2 distributions are obtained from the samples, and their 68% and 95% confidence intervals are shown as gray bands around one.

To identify whether a given scale cut introduces significant systematic effects, we follow a set of general criteria. Since the physical values of the parameters are not expected to depend on scale, a significant shift in best-fit values when changing the scale cut is indicative of modelling failures at too small scales or of excessive signal loss when too

many scales are removed. Increasing r_{min} should therefore result in larger uncertainties, but should not systematically shift best-fit values. The presence of projection effects provides further diagnostic power. Such effects may arise from parameter degeneracies that render marginalized distributions non-Gaussian, causing the median to differ from the maximum-likelihood value (MLE). They can also appear when the removal of informative scales induces new degeneracies among parameters that were previously well constrained. When projection effects are present at all scales, we do not consider the affected parameter to guide the choice of r_{min}. Another important consideration is the potential underestimation of confidence intervals in regimes where the model loses descriptive power. In numerical simulations, where cosmological parameters are known a priori, fits of the measured correlation functions are expected to recover their true values. Although the true cosmology is not known in our case, previous studies (Veropalumbo et al. 2022, Guidi et al. 2023, Euclid Collaboration: Guidi et al. 2025) guide the scales over which tree-level models remain reliable. Finally, reduced χ^2 values should not be significantly larger than unity, as the χ^2 test quantifies the consistency between the model and the data, given the uncertainties.

The results of the scale tests for the 2PCF are shown in Fig. 4.1. The b_2 and $b_{\mathcal{G}2}$ parameters exhibit projection effects at all scales, as expected given that the 2PCF alone does not strongly constrain them. Systematic effects are therefore better assessed through the b_1 and f posteriors. The NGC high-redshift sample shows a clear scale dependence in both b_1 and f when moving from $r_{min} = 20\,h^{-1}$Mpc to larger scale cuts. Since this sample has the highest signal-to-noise ratio, this behaviour alone motivates the adoption of a more conservative cut of $r_{min} = 30\,h^{-1}$Mpc. The reduced χ^2 values show substantial variation across datasets, with particularly large values for the SGC high-redshift sample. Uncertainty intervals are consistently larger for the SGC subsamples, as expected from their lower signal-to-noise ratio compared to the NGC subsamples.

The corresponding analyses for the 3PCF are shown in Fig. 4.2. The reduced χ^2 exhibits a clear dependence on both r_{min} and η_{min}. For $\eta_{min} = 1$ and 2, the posterior-averaged χ^2 is incompatible with unity, especially for the NGC subsamples. Combined with the extremely small variance of the χ^2 distributions, this motivates the rejection of triangles with $\eta_{min} \leq 2$ and the adoption of $\eta_{min} = 3$ for the combined analysis. Posterior medians for the parameters are generally unstable and show projection effects for b_1 and f. The degeneracy between b_1 and f remains mildly present in the 3PCF, with best-fit values of f consistent with zero across all scales. A more reliable diagnostic of systematic effects is provided by the b_2 and $b_{\mathcal{G}2}$ posteriors, which remain stable in the NGC samples up to $r_{min} = 40\,h^{-1}$Mpc and begin to exhibit significant projection effects at larger scale cuts. Removing triangles at smaller scales leads to a substantial loss of signal-to-noise ratio and noticeably affects posterior sampling, while including those scales reduces uncertainties at the expense of increased systematic errors. Results from the literature (Veropalumbo et al. 2022, Euclid Collaboration: Guidi et al. 2025) indicate that tree-level 3PCF models lose their descriptive power below $r_{min} = 40\,h^{-1}$Mpc. Given these considerations, we find that a good compromise between volume and modelling is $r_{min} = 40\,h^{-1}$Mpc.

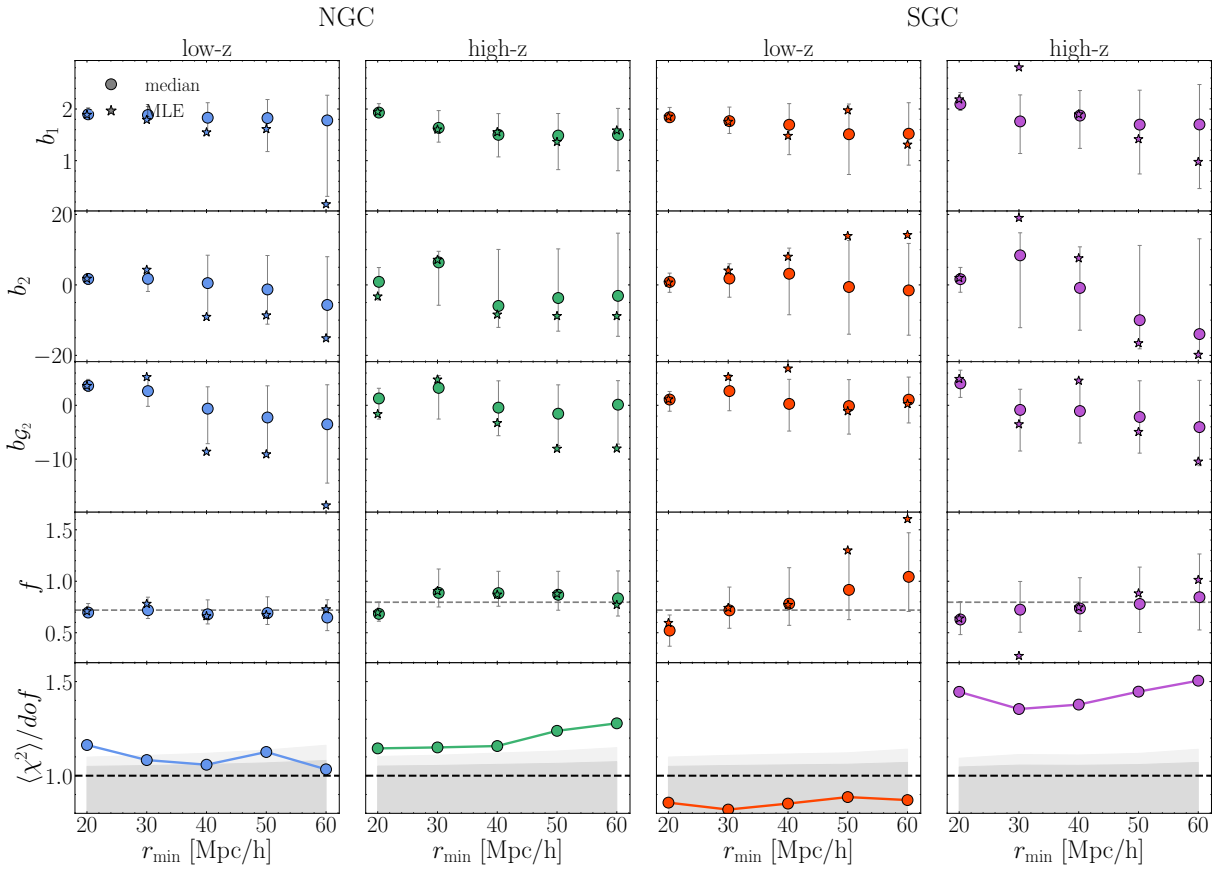

Figure 4.1: Marginalized constrains on bias parameters and f for the BOSS 2PCF. The posterior median, along with 16th and 84th percentiles, is displayed as a function of the minimal scale r_{min}. The star marker corresponds to the posterior sample of the associated maximum likelihood. The maximum scale r_{max} is fixed at $150\,h^{-1}$Mpc. The gray dashed line in the f posteriors refers to the Planck 2018 fiducial value for the corresponding redshift. The last panel shows posterior-averaged χ^2 divided by the number of degrees of freedom (d.o.f.) while the gray bands display the 68% and 95% confidence intervals for the χ^2 distribution.

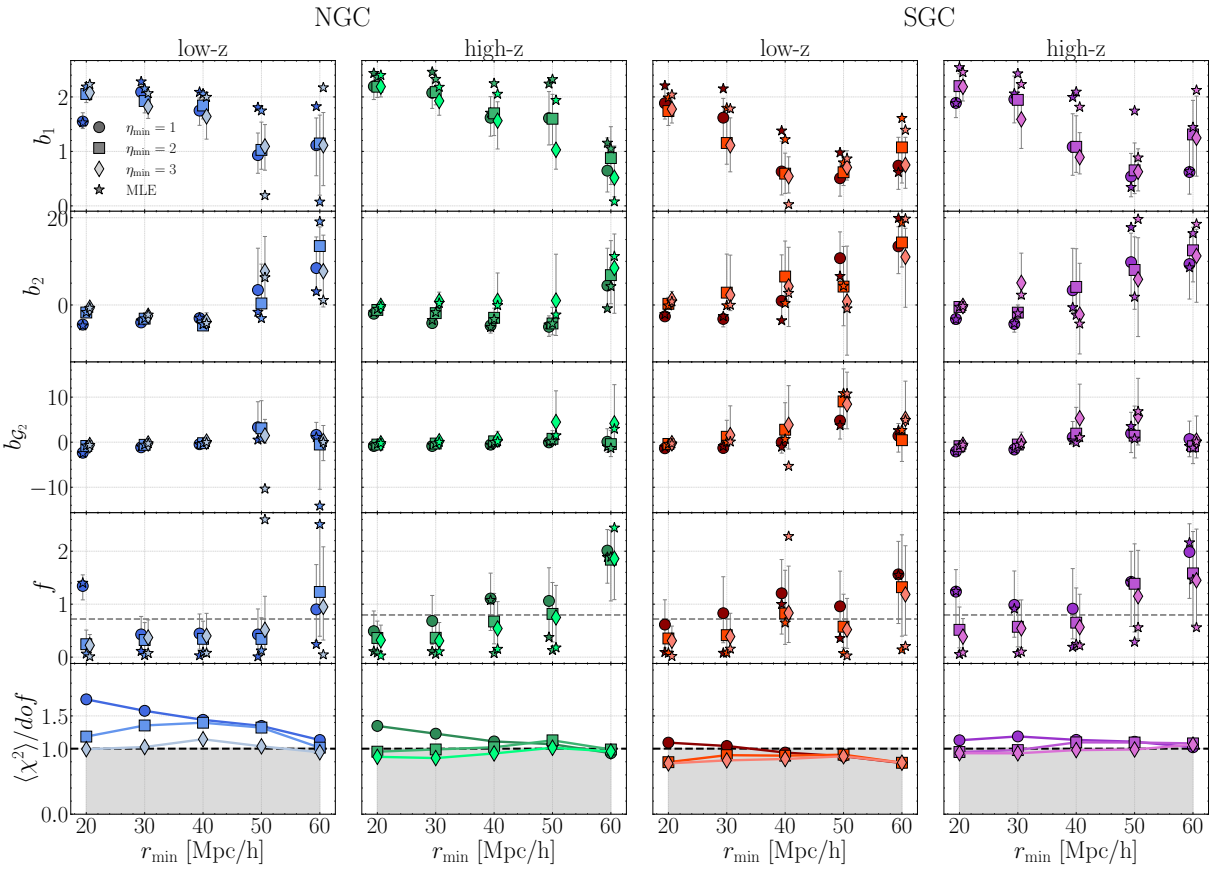

Figure 4.2: Marginalized constrains on bias parameters and f for the BOSS 3PCF. The posterior median, along with 16th and 84th percentiles, is displayed as a function of the minimal scale r_{min} and η_{min}. The star marker corresponds to the posterior sample of the associated maximum likelihood. The maximum scale r_{max} is fixed at $130\,h^{-1}$Mpc. The gray dashed line in the f posteriors refers to the Planck 2018 fiducial value for the corresponding redshift. The last panel shows posterior averaged χ^2 divided by the number of degrees of freedom (d.o.f.) while the gray bands display the 68% and 95% confidence intervals for the χ^2 distribution, refereed to the $\eta_{min} = 1$ case.

4.2 The final configuration

In Sec. 3.4 we have performed tests on synthetic data to quantify their power to constrain models. We showed that when fitting the correlation functions with BOSS covariance matrices, it is impossible to retrieve unbiased values of the parameters. We have attributed this behaviour to volume effects, which are ultimately rooted in both parameter degeneracies and the uncertainties of the BOSS data. We have proposed several strategies for dealing with this problem and have found that the best choice is to limit the dimensionality of the parameter space by fixing higher order bias parameters $b_{\mathcal{G}2}$, $b_{\Gamma 3}$ to theoretical relations (see Eq. 3.4.3 and 3.4.4) which vary as a function of b_1. In Sec. 4.1 we performed tests on the BOSS 2PCF and 3PCF to choose the optimal scale and triangle cuts r_{min} and η_{min} in a way that minimizes modelling systematics and maximizes volume.

Here, we summarize the final configurations we have explored in our work. We perform four separate Bayesian analyses of the BOSS combined samples. Each evaluation is made by combining the likelihood of each of the four BOSS samples at every step of the chain. In this procedure, the NGC and SGC samples share the same set of parameters, while the low-z and high-z samples have separate dedicated parameters as they belong to different redshift bins. This means that every parameter of the theory model appears twice (one for each redshift bin). We have described this procedure in detail in Sec. 3.1. The final setups for our analysis are summarized here:

2PCF σ_8 fixed We fit the 2PCF multipoles $\xi_0 - \xi_2 - \xi_4$ from $r_{min} = 30\ h^{-1}$Mpc to $r_{max} = 150\ h^{-1}$Mpc. The total length of the data vector is $24 \times 3 = 72$, where 24 is the number of bins given $\Delta r = 5\ h^{-1}$Mpc and 3 is the number of multipoles. The main goal of this analysis is to estimate the values of AP parameters for the distance measurement.

3PCF σ_8 fixed We fit the all-scales 3PCF $\zeta^{(all)}$ from $r_{min} = 40\ h^{-1}$Mpc to $r_{max} = 130\ h^{-1}$Mpc and $\eta_{min} = 3$. The total number of triangles is 658. The main goal of this analysis is to compare the confidence regions on both bias and cosmological parameters with the 2PCF constraints. For reference, the number of triangles from a specific scale cut of the all-scales 3PCF is given in Tab. 4.1.

2PCF+3PCF σ_8 fixed We fit the joint 2PCF+3PCF of BOSS where the same scale cuts of the individual 2PCF and 3PCF are applied. The total length of the data vector is $72 + 658 = 730$ bins. The main goal of this analysis is to assess the constraints that can be achieved when combining 2PCF with the 3PCF, and compare this with their independent results.

2PCF+3PCF σ_8 free We fit the joint 2PCF+3PCF of BOSS where the same scale cuts of the individual 2PCF and 3PCF are applied. In this case, we let the σ_8 parameter free, using a uniform prior. We explore in this way the potential of combining 2PCF with 3PCF measurements to disentangle the $b_1 - f - \sigma_8$ degeneracy. We limit the parameter space available by fixing $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$ to theoretical relations in Eq. 3.4.3 and 3.4.4. This allows us to break the degeneracy, without introducing significant systematic biases.

The priors used for each of the four chains are described in Tab. 4.2. We generally choose uninformative priors to allow parameters to vary within physical limitations. In the first three fits, all parameters vary as freely as possible within physical limits, with the only exception of σ_8, which is kept fixed at the fiducial Planck 2018 cosmology. Regarding the bias parameters, we limit the linear bias b_1 only to have positive values. Higher-order biases are fully free to vary in the range $\mathcal{U}[-20, 20]$. The linear growth rate f is treated similarly, as it is not expected to have negative values. The priors on AP parameters allow for a large exploration of cosmological models, but we do not expect them to vary significantly from the fiducial Planck 2018 cosmology, as also shown in other BOSS studies. In the fourth fit, we marginalize over tidal bias parameters by fixing them to LLIMD theoretical relations while keeping σ_8 free to vary within the physical limitation of non-negative values.

	$r_{\rm min}\,[h^{-1}\,{\rm Mpc}]$				
$\eta_{\rm min}$	20	30	40	50	60
1	1399	1169	919	673	455
2	1232	1015	782	557	364
3	1078	874	**658**	454	286

Table 4.1: Number of traingles of the all-scales 3PCF as a function of η_{min} and r_{min}. The maximum scale is kept fixed at $r_{max} = 130\,h^{-1}$Mpc

	2PCF	**3PCF**	**2PCF+3PCF**	**2PCF+3PCF** $+\sigma_8$
Bias parameters				
b_1	$\mathcal{U}[0,4]$	$\mathcal{U}[0,4]$	$\mathcal{U}[0,4]$	$\mathcal{U}[0,4]$
b_2	$\mathcal{U}[-20,20]$	$\mathcal{U}[-20,20]$	$\mathcal{U}[-20,20]$	$\mathcal{U}[-20,20]$
$b_{\mathcal{G}2}$	$\mathcal{U}[-20,20]$	$\mathcal{U}[-20,20]$	$\mathcal{U}[-20,20]$	$\mathcal{F}$[theo.]
$b_{\Gamma 3}$	$\mathcal{U}[-20,20]$		$\mathcal{U}[-20,20]$	$\mathcal{F}$[theo.]
Cosmology				
f	$\mathcal{U}[0,4]$	$\mathcal{U}[0,4]$	$\mathcal{U}[0,4]$	$\mathcal{U}[0,4]$
σ_8	$\mathcal{F}$[fid.]	$\mathcal{F}$[fid.]	$\mathcal{F}$[fid.]	$\mathcal{U}[0,4]$
α	$\mathcal{U}[0.9,1.1]$		$\mathcal{F}$[1.]	$\mathcal{F}$[1.]
ϵ	$\mathcal{U}[-0.1,0.1]$		$\mathcal{F}$[0.]	$\mathcal{F}$[0.]

Table 4.2: Priors used for the final posteriors. $\mathcal{U}$ denotes a *uniform* prior; $\mathcal{F}$ denotes a *fixed* prior. "fid." refers to fiducial values for the Cosmological model assumed throughout the analysis (Planck Collaboration: Aghanim et al. 2020). "theo" refers to the use of theoretical relations to fix a particular parameter.

4.3 Distance measurements from the AP test

We have performed a BOSS 2PCF Bayesian analysis with the goal of obtaining a reliable measure of AP parameters. We exploit information on the 2PCF monopole, quadrupole and hexadecapole; by applying the scale cuts described in Sec. 4.1 we have $r_{min} = 30\,h^{-1}$Mpc, $r_{max} = 150\,h^{-1}$Mpc. The length of the data vector is $24 \times 3 = 72$, given a binsize of $\Delta r = 5\,h^{-1}$Mpc. The priors on the parameters are described in 4.2 and are equal for low-z and high-z samples. At each step of the chain, the log-likelihood from the 4 BOSS datasets is summed to fully exploit the BOSS survey volume and get the best possible constraints. The theoretical predictions for the one-loop 2PCF are fast thanks to the emulation process described in Sec. 3.1. In this estimation, the σ_8 parameter is fixed to the Planck 2018 cosmology for both redshift bins. The free parameters are

$$\{b_1^{lowz}, b_2^{lowz}, b_{\mathcal{G}2}^{lowz}, b_{\Gamma 3}^{lowz}, f^{lowz}, \alpha^{lowz}, \epsilon^{lowz},$$
$$b_1^{highz}, b_2^{highz}, b_{\mathcal{G}2}^{highz}, b_{\Gamma 3}^{highz}, f^{highz}, \alpha^{highz}, \epsilon^{highz}\}.$$

The two-dimensional marginalized constraints and expectation values of the AP parameters are shown in Fig. 4.3. The convergence of the chain is ensured by employing a number of walkers 4 times larger than the number of parameters, 128 in this case, and terminating the chain after more than 50 (mean) auto-correlation times.

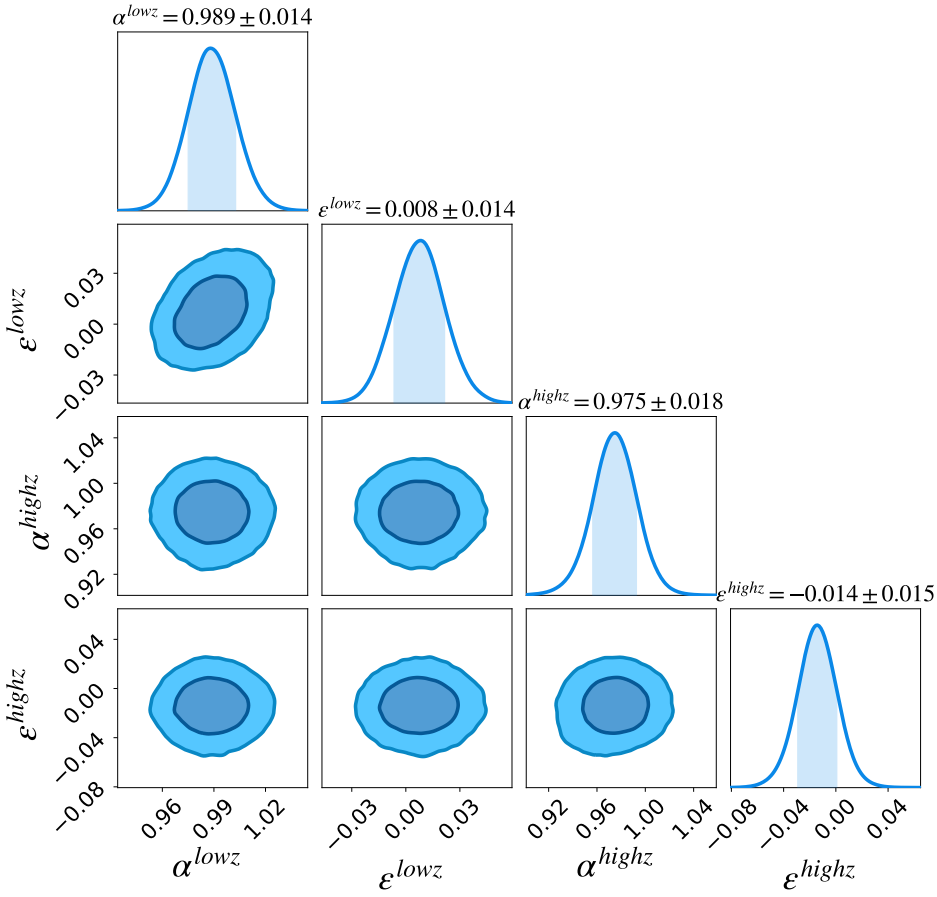

Figure 4.3: 2D marginalized posterior distributions on AP parameters from the BOSS 2PCF Bayesian analysis. The contours for both the low-z ($z = 0.38$) and high-z ($z = 0.61$) samples are present. Above each 1D distribution is the corresponding median (50th percentile) and 16th to 84th percentiles.

4.3.1 Comparison with the literature

AP parameters measure the deviation from the fiducial cosmology used to generate theoretical predictions of the 2PCF. Denoting with "fid" the quantities computed at the fiducial cosmology, we can write the coming transverse distance and the Hubble parameter as

$$\begin{aligned} \frac{D_M(z)}{r_d} &= \frac{\alpha}{1+\epsilon}\frac{D_{M,\mathrm{fid}}(z)}{r_{\mathrm{d,fid}}} \\ H(z)r_d &= \frac{1}{\alpha(1+\epsilon)^2}H_{\mathrm{fid}}(z)r_{\mathrm{d,fid}} \end{aligned} \tag{4.3.1}$$

Where $D_M(z)$ is defined as

$$D_M(z) = \begin{cases} D_H \frac{1}{\sqrt{\Omega_k}} \sinh\left[\sqrt{\Omega_k} D_C/D_H\right], & \text{for } \Omega_k > 0 \\ D_C, & \text{for } \Omega_k = 0 \\ D_H \frac{1}{\sqrt{|\Omega_k|}} \sin\left[\sqrt{|\Omega_k|} D_C/D_H\right], & \text{for } \Omega_k < 0 \end{cases} \tag{4.3.2}$$

which, in the case where the universe is spatially flat, is simply the previously defined l.o.s. comoving distance from Eq. 1.3.1 and is related to the angular diameter distance by $D_M = (1+z)D_A$. Notice that what is constrained by the AP rescaling is actually the distances in units of the sound horizon at decoupling time r_d. Expressed in these terms, the results are independent of the fiducial cosmologies assumed.

Our constraints correspond to a 3.5−3.8% uncertainty on $H(z)r_d/r_{d,\text{fid}}$ and a 1.6−2.4% uncertainty on $D_M(z)r_{d,\text{fid}}/r_d$, for the low-$z$ and high-z bins, respectively. Other studies have made similar measurements using BOSS DR12 data. In general, small deviations are to be expected, given the differences in the analysis. Since we provide an independent measurement of distance from the AP test on the 2PCF of BOSS DR12, we compare our results with the *BOSS BAO consensus*. As explained in Alam et al. (2017), the posterior distributions covered from the different analysis methodologies applied to BOSS are well described by Gaussian multivariate distributions. To maximize the constraining power, their posterior distributions can be combined. We report, in Tab. 4.3, the combined one-dimensional marginalized intervals of $D_M(z)$ and $H(z)$ from 'BAO' studies which constrain distance information by means of the AP test after the reconstruction of the density field (the so-called BOSS BAO consensus). In the table, we also report the mean and standard deviation of $D_M(z)r_{d,\text{fid}}/r_d$ and $H(z)r_d/r_{d,\text{fid}}$ extrapolated by Planck Collaboration: Aghanim et al. (2020) at the corresponding BOSS redshifts.

The uncertainties of the BAO consensus in comoving transverse distance are 1.6% across the two redshift bins, while for the Hubble parameter, a $2.9 - 2.5\%$ uncertainty is achieved for low-z to high-z. Our low redshift result is remarkably similar to the BOSS BAO consensus, while the high redshift result is more discrepant; a difference that is primarily attributable to the use of density-field reconstruction, which improves the precision of distance measurements.

In Fig. 4.4 we compare our measurements of $D_M(z)r_{d,\text{fid}}/r_d$ and $H(z)r_d/r_{d,\text{fid}}$ at the two BOSS redshifts (low-z and high-z) with the parameter space allowed by a flat ΛCDM model with parameters determined by Planck Collaboration: Aghanim et al. (2020), as well as 2D marginalized contours at the 68% and 95% confidence regions of the BOSS BAO consensus. To better visualise how our results could contribute to the H_0 tension, we have included values of the Hubble constant up to $H_0 \sim 74$ Km/s/Mpc which correspond to those inferred by local universe experiments that exploit SNe type Ia as standard candles. Clearly, our marginalised distributions favour CMB-derived solutions for the values of H_0. Our low-redshift result is within 1σ of the Planck 2018 prediction of $H_0 = 67.3$Km/s/Mpc. Our high-redshift result barely overlaps with the fiducial Planck 2018 prediction and is just outside the 1σ region.

name	redshift	$D_M(z)r_{\rm d,fid}/r_{\rm d}$ (Mpc)	$H(z)r_d/r_{\rm d,fid}$ (km s^{-1} Mpc^{-1})
This work	0.38	1506 ± 24	82.5 ± 2.9
This work	0.61	2287 ± 54	100.6 ± 3.8
Planck 2018	0.38	1535 ± 10	82.8 ± 0.2
Planck 2018	0.61	2312 ± 13	95.2 ± 0.3
BOSS consensus	0.38	1506 ± 24	80.7 ± 2.4
BOSS consensus	0.61	2309 ± 37	98.9 ± 2.5

Table 4.3: Our $D_M(z)$ and $H(z)$ constraints from the BOSS 2PCF multipoles compared with Planck 2018 constraints extracted from TT, TE, EE.

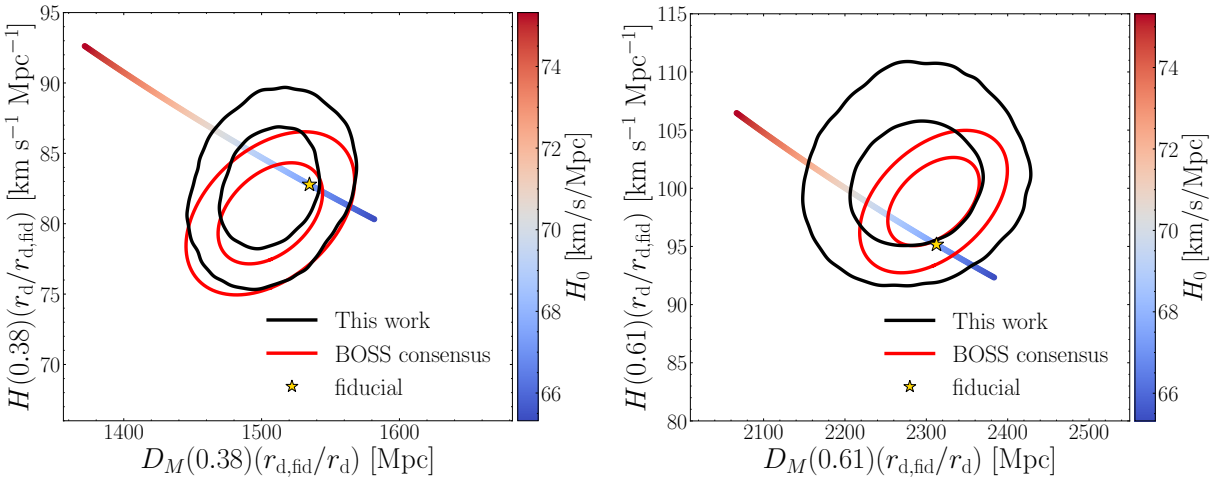

Figure 4.4: The allowed 1 and 2σ regions (black ellipses) in the Hubble parameter, H, and the comoving transverse distance, D_M determined from AP rescaling of the 2PCF, and those extracted from Alam et al. (2017) for the BOSS BAO consensus (red ellipses). The left panel shows the result for the low-z sample, while the right panel for the high-z one. The colored points represent a range of values in the $D_M - H$ space allowed by a flat ΛCDM cosmology with parameters from Planck Collaboration: Aghanim et al. (2020). Different colours represent the value of H at $z = 0$ (as indicated by the colour bar on the right). The star marker denotes values of $D_{M,\rm fid}(z)$ and $H_{\rm fid}(z)$.

4.4 Cosmological constraints from the joint analysis of 2PCF and 3PCF

We performed a joint 2PCF+3PCF Bayesian analysis on the BOSS combined samples, assuming a fixed value of σ_8. The primary goal of this study is to assess the additional constraining power provided by the 3PCF when combined with the 2PCF, compared to using each statistic independently. Figure 4.5 shows the 2D marginalized constraints obtained from three different chains. The blue contours correspond to the 2PCF multipoles (monopole, quadrupole, and hexadecapole) with a minimum scale cut of $r_{\rm min}^{2PCF} = 30\,h^{-1}$Mpc. The orange contours are derived from the all-scales 3PCF, using $r_{\rm min}^{3PCF} = 40\,h^{-1}$Mpc and $\eta_{\rm min} = 3$. Finally, the dark brown contours represent the joint 2PCF+3PCF analysis, adopting the same scale cuts as in the individual cases. The total data vector consists of 72 radial bins from the 2PCF and 658 triangle configurations from the 3PCF, for a total length of 730. The motivation for the adopted scale cuts is discussed in Sec. 4.1.

The convergence of the chain is ensured by employing a number of walkers 4 times larger than the number of parameters (128 walkers for the 2PCF and 64 walkers for the other ones), and terminating the chain after more than 50 auto-correlation times.

Parameter expectation values are extracted from the one-dimensional marginalized posterior distributions, taking the 50th percentile as the central value and the 16th and 84th percentiles as the quoted uncertainties. These results are reported in Tab. 4.4. We also report, as MLE values, the parameter set corresponding to the maximum posterior probability.

The 2PCF-only posterior is unable to properly constrain the higher-order bias parameters, which remain strongly degenerate with one another. Their one-dimensional marginalized posteriors exhibit bimodal distributions for the high-redshift sample, as expected since these parameters enter the model quadratically and are sampled under largely uninformative priors. In contrast, the linear bias b_1 and the growth rate f are well constrained by the 2PCF alone. The inclusion of the quadrupole, and to a lesser extent the hexadecapole, plays a crucial role in constraining redshift-space distortion anisotropies.

The situation differs for the 3PCF-only posterior, where the $b_1 - f$ degeneracy is not fully broken, and the posterior for f extends to the lower prior boundary at zero. We expect that including anisotropic components of the 3PCF would improve constraints on f. For the higher-order bias parameters, the two-dimensional marginalized contours from the 3PCF show probability distributions that are nearly complementary to those obtained from the 2PCF. This complementarity is key: when the 2PCF and 3PCF are combined, degeneracies present in both statistics are effectively broken. In particular, the 3PCF provides crucial information to constrain higher-order bias parameters, which in turn improves the constraints on other parameters such as b_1, f, and $b_{\Gamma 3}$, yielding overall tighter confidence regions.

Slightly tighter constraints are obtained for the low-redshift sample. Unlike the original BOSS LOWZ and CMASS samples, which differed significantly in survey volume, the optimally binned samples used in this analysis have nearly equal volumes. While the

survey volume primarily determines the noise level, the clustering signal is expected to be stronger at lower redshifts. This alone can account for the observed difference between the low-z and high-z confidence regions.

Quantitatively, the 2PCF alone yields uncertainties of $7-15\%$ on b_1, while the 3PCF alone provides constraints at the $19-22\%$ level (from low-z to high-z). In the joint 2PCF+3PCF analysis, the uncertainty on b_1 is reduced to $\sim 4\%$. For the growth rate f, the 2PCF alone constrains it at the $12-16\%$ level, whereas it is essentially unconstrained by the 3PCF alone; combining the two statistics reduces the uncertainty to $\sim 11\%$.

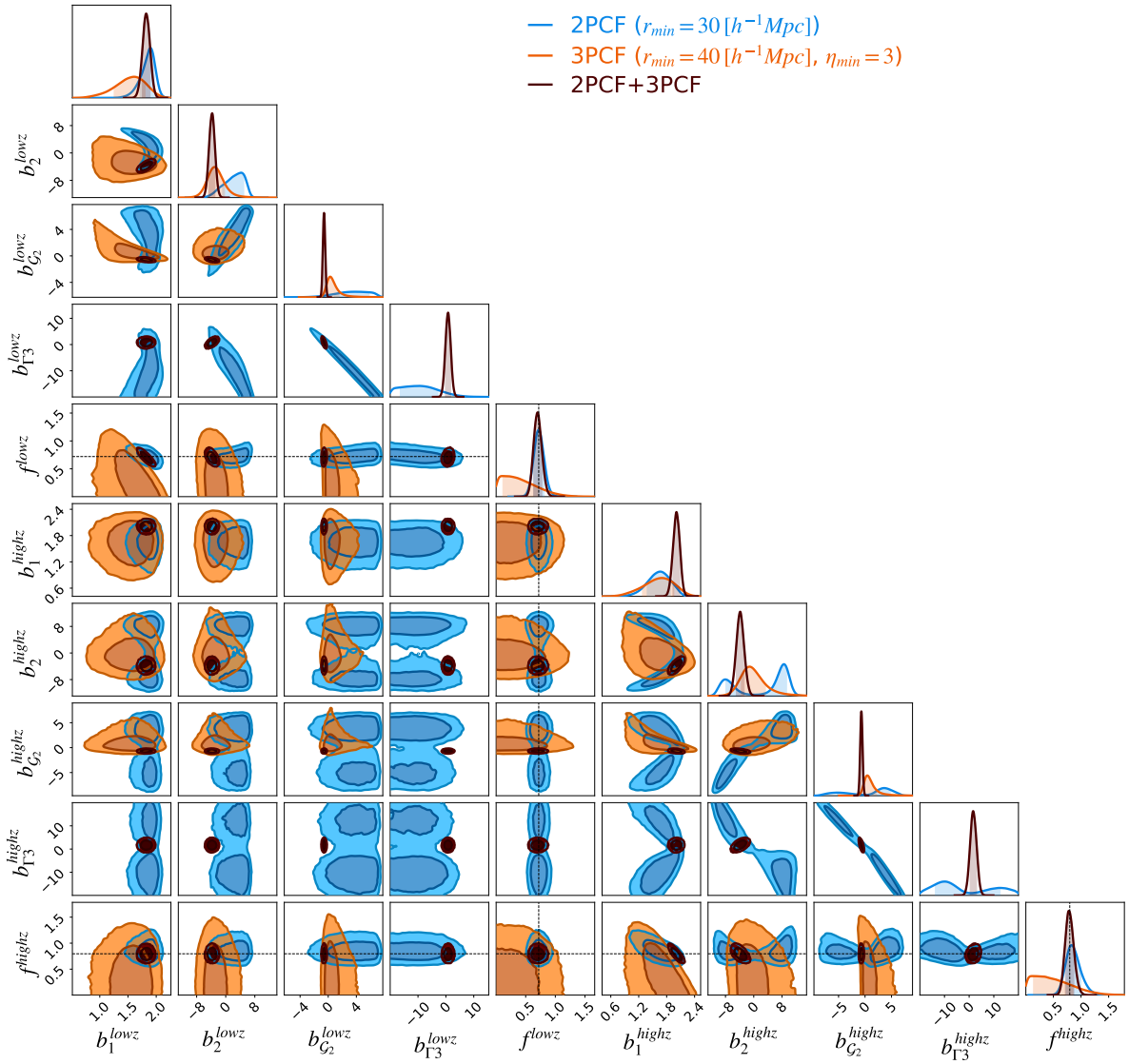

Figure 4.5: 2D marginalized constraints on three different chains. Contours show the 68% and 95% confidence regions. In blue, the contours of the 2PCF multipoles. In orange, the contours of the 3PCF. In dark brown, the contours of the joint analysis (2PCF+3PCF). Orange contours are not present for the $b_{\Gamma 3}$ parameter since it is not part of the 3PCF model. The black dashed lines indicate the fiducial values of f at $z = 0.38$ for low-z and $z = 0.61$ for high-z.

	2PCF		3PCF		2PCF+3PCF	
parameter	MLE	median	MLE	median	MLE	median
b_1^{lowz}	1.87	$1.87^{+0.10}_{-0.15}$	1.81	$1.55^{+0.27}_{-0.32}$	1.95	$1.82^{+0.07}_{-0.07}$
b_1^{highz}	1.62	$1.64^{+0.24}_{-0.26}$	2.08	$1.64^{+0.34}_{-0.39}$	2.09	$2.01^{+0.08}_{-0.08}$
b_2^{lowz}	3.40	$2.52^{+2.46}_{-3.39}$	-3.48	$-2.82^{+2.66}_{-2.26}$	-2.93	$-3.78^{+0.87}_{-0.88}$
b_2^{highz}	7.90	$6.42^{+2.87}_{-14.50}$	-1.03	$-0.56^{+4.20}_{-3.16}$	-1.15	$-3.84^{+1.23}_{-1.21}$
$b_{\mathcal{G}2}^{lowz}$	4.72	$3.54^{+2.31}_{-2.62}$	0.24	$0.50^{+1.11}_{-0.65}$	-0.73	$-0.59^{+0.19}_{-0.18}$
$b_{\mathcal{G}2}^{highz}$	4.20	$2.52^{+2.23}_{-7.87}$	0.24	$0.76^{+1.53}_{-0.86}$	-0.60	$-0.69^{+0.24}_{-0.23}$
$b_{\Gamma3}^{lowz}$	-13.57	$-9.91^{+7.09}_{-6.61}$			0.43	$0.78^{+1.03}_{-1.03}$
$b_{\Gamma3}^{highz}$	-11.56	$-6.31^{+19.14}_{-7.48}$			2.01	$1.69^{+1.45}_{-1.41}$
f^{lowz}	0.68	$0.72^{+0.09}_{-0.08}$	0.21	$0.36^{+0.38}_{-0.27}$	0.62	$0.70^{+0.08}_{-0.07}$
f^{highz}	0.85	$0.85^{+0.15}_{-0.12}$	0.07	$0.44^{+0.45}_{-0.33}$	0.72	$0.79^{+0.09}_{-0.08}$

Table 4.4: Constraints on bias and cosmological parameters from the σ_8 fixed analysis on BOSS 2PCF, 3PCF and 2PCF+3PCF. The expectation values are given in terms of the median (50th percentile), where the associated uncertainties are the 16th to 84th percentiles. MLE values denote the sample associated with the maximum likelihood.

The inclusion of the 3PCF sharpens the constraints on higher-order bias parameters by more than a factor of three. In particular, $b_{\mathcal{G}2}$ and $b_{\Gamma3}$ transition from being completely unconstrained in the 2PCF-only analysis to relative uncertainties of $40-80\%$ in the joint case.

The expectation values of f obtained from the joint 2PCF+3PCF analysis are fully consistent with the fiducial Planck 2018 cosmology,

$$f_{\rm fid}(z_{\rm eff}=0.38)=0.70, \qquad f_{\rm fid}(z_{\rm eff}=0.61)=0.79\,.$$

This agreement is not surprising given the strong prior imposed on σ_8,

$$\sigma_{8,\rm fid}(z_{\rm eff}=0.38)=0.66, \qquad \sigma_{8,\rm fid}(z_{\rm eff}=0.61)=0.59\,.$$

We stress that the goal of this analysis is not to obtain the most accurate measurement of the cosmological parameters, but rather to compare the confidence regions obtained by combining the 2PCF with the 3PCF.

The tightness of the constraints is further illustrated by the best-fit models shown in Figs. 4.6 and 4.7. The black curves correspond to models evaluated at the median parameter values listed in Tab. 4.4, while the gray bands indicate the 16th to 84th percentile ranges. These bands are obtained by computing 2PCF multipoles and all-scales 3PCF models for a subset of posterior samples, and then extracting the 16th and 84th

percentiles of the resulting distribution. Both the best-fit 2PCF monopole and the all-scales 3PCF exhibit a deficit of power on large scales relative to the BOSS measurements. This behaviour was already found in Sec. 2.3, where we compared the mean of the mock catalogues to the BOSS data. It is, in fact, a well-established feature of BOSS clustering measurements (Ross et al. 2017).

Although the overall amplitude of the best-fit model may appear puzzling, particularly because it fails to match the low-uncertainty bins of the 2PCF monopole, we stress that the errors reported in the figure represent only the diagonal variance of the data. A reliable assessment of goodness of fit requires accounting for the full covariance matrix, which properly captures bin-to-bin correlations. Indeed, the reduced χ^2 values are closer to unity for the SGC (low-z) data set than for the NGC sample, despite the opposite impression that might be drawn from the visual comparison. As demonstrated in Appendix B, when the posterior is sampled only considering the diagonal component of the covariance, the resulting best-fit models preferentially adjust the overall normalization to better match the bins with the smallest quoted uncertainties.

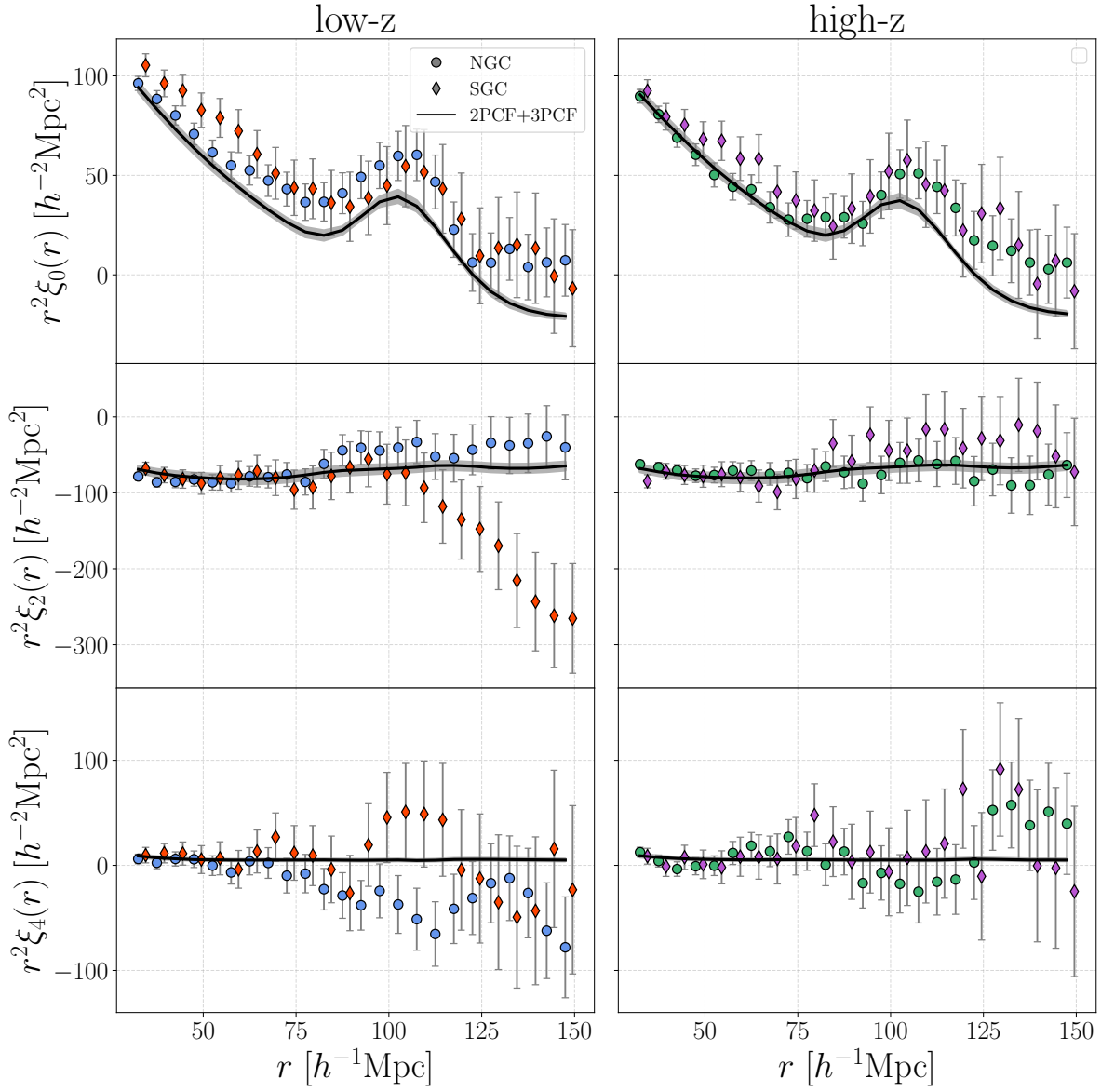

Figure 4.6: From top to bottom the BOSS 2PCF multipoles (monopole, quadrupole and hexadecapole from $r_{min} = 30\ h^{-1}Mpc$ to $r_{max} = 150\ h^{-1}Mpc$). Alongside best-fit models extracted from median parameters and 16th to 84th percentiles of the σ_8 fixed 2PCF+3PCF fit. The set of parameters that defines the model is the same for the NGC and SGC sub-samples. Error bars are computed as the square root of the covariance matrix diagonal and so represent the standard deviation associated with each data point.

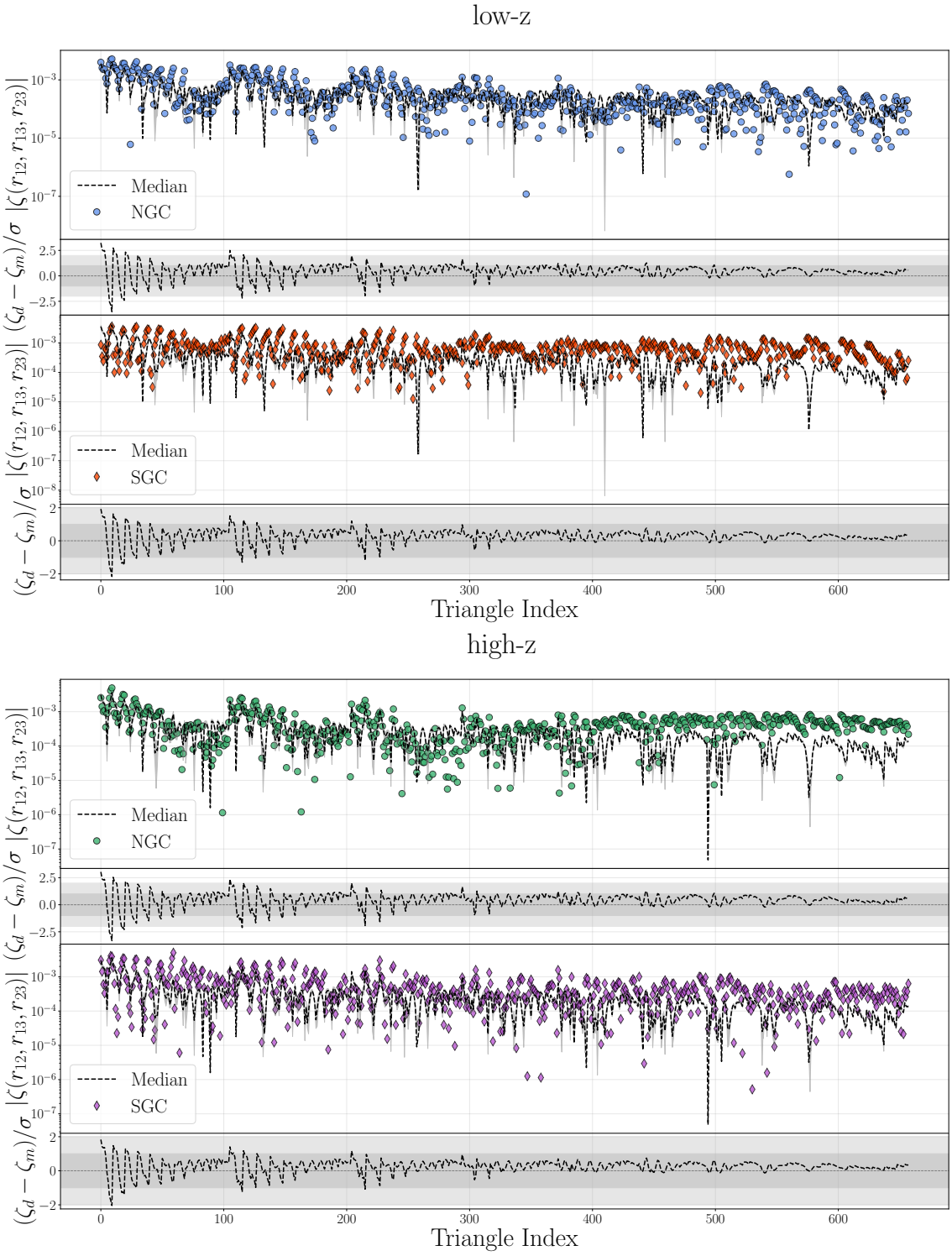

Figure 4.7: From top to bottom the BOSS all-scales 3PCF (all triangles from $r_{min} = 40h^{-1}Mpc$ to $r_{max} = 130\ h^{-1}Mpc$ and $\eta_{min} = 3$) of low-z and high-z samples. Alongside best-fit models extracted from median parameters and 16th to 84th percentiles of the σ_8 fixed 2PCF+3PCF fit. Under each of them is the residual plot showcasing the difference between models and data per unit σ, the standard deviation associated with each data point.

4.5 Extending the analysis to free σ_8

As a final step, we perform a joint analysis of the 2- and 3-point correlation functions (2PCF and 3PCF) measured from the combined BOSS samples, this time allowing the parameter σ_8 to vary freely. As discussed in Sec. 3.4, we mitigate the degeneracies that arise in the analysis by reducing the dimensionality of the parameter space, imposing theoretical relations among the bias parameters $b_{\mathcal{G}2}$ and $b_{\Gamma 3}$. The main objective of this analysis is to evaluate the extent to which the combined 2PCF+3PCF can simultaneously constrain f and σ_8, a task that is not feasible when either statistic is considered in isolation. In Fig. 4.8, we show the resulting two-dimensional marginalized contours from this fit, alongside those obtained from the previous 2PCF+3PCF analysis in which σ_8 was kept fixed.

The marginalized constraints for the main parameters are reported as the median of the posterior distributions, with uncertainties given by the 16th and 84th percentiles. We also quote the MLE values corresponding to the sample with the highest posterior probability. A comparison between the expectation values obtained with σ_8 free and those from the σ_8-fixed analysis is provided in Tab. 4.5.

As expected, allowing σ_8 to vary leads to broader uncertainty regions. Nevertheless, the joint analysis successfully breaks the degeneracy, enabling separate constraints on f and σ_8. The linear bias is constrained at the $13-16\%$ level (from low-z to high-z), compared to the $\sim 4\%$ uncertainties obtained when σ_8 is fixed. The growth rate f is measured with a $13-15\%$ uncertainty, slightly degraded with respect to the $\sim 11\%$ uncertainty of the σ_8-fixed case. In addition, we obtain an independent measurement of σ_8 with an associated uncertainty of $11-14\%$. The shift of the median values of b_1 and σ_8 (and, also, b_2) away from their maximum-likelihood estimates indicates that projection effects are still present, implying that the degeneracy is not completely removed. Projection effects are unavoidable given the BOSS covariance, but they could be mitigated in future surveys probing larger volumes.

The uncertainties on the quadratic bias parameter b_2 show no significant variation between the two analyses, with the marginalized posteriors remaining consistent within 1σ. When σ_8 is free, b_2 is constrained at the $27-34\%$ level, compared to $21-29\%$ in the σ_8-fixed case. This confirms that b_2 is largely decoupled from the other parameters in the joint 2PCF+3PCF analysis.

In Fig. 4.9, we present the best-fit model compared to the BOSS 2PCF measurements. The model curves are evaluated at the median parameter values reported in Tab. 4.5, shown for both the σ_8-fixed and σ_8-free posterior distributions. The best-fit models for NGC and SGC subsamples constitute the same prediction as they are jointly fit, but the corresponding χ^2 is not and is reported in Tab. 4.6. Both models produce extremely similar predictions and therefore result in similar χ^2. It might seem, from a visual comparison, that predictions for the low-z SGC sample should be associated with a significantly higher reduced χ^2, but in reality, the simple representation of uncertainties in the plot doesn't show the full picture, which is instead captured when considering the full covariance matrix. By inspecting the covariance matrix of low-z SGC and NGC subsamples, it is clear that the SGC one has larger uncertainty than NGC, and this alone can

explain the difference in χ^2. In Appendix B, we show best-fit models and corresponding reduced χ^2 of posterior distributions obtained by considering only the variance associated with the 2PCF radial bins.

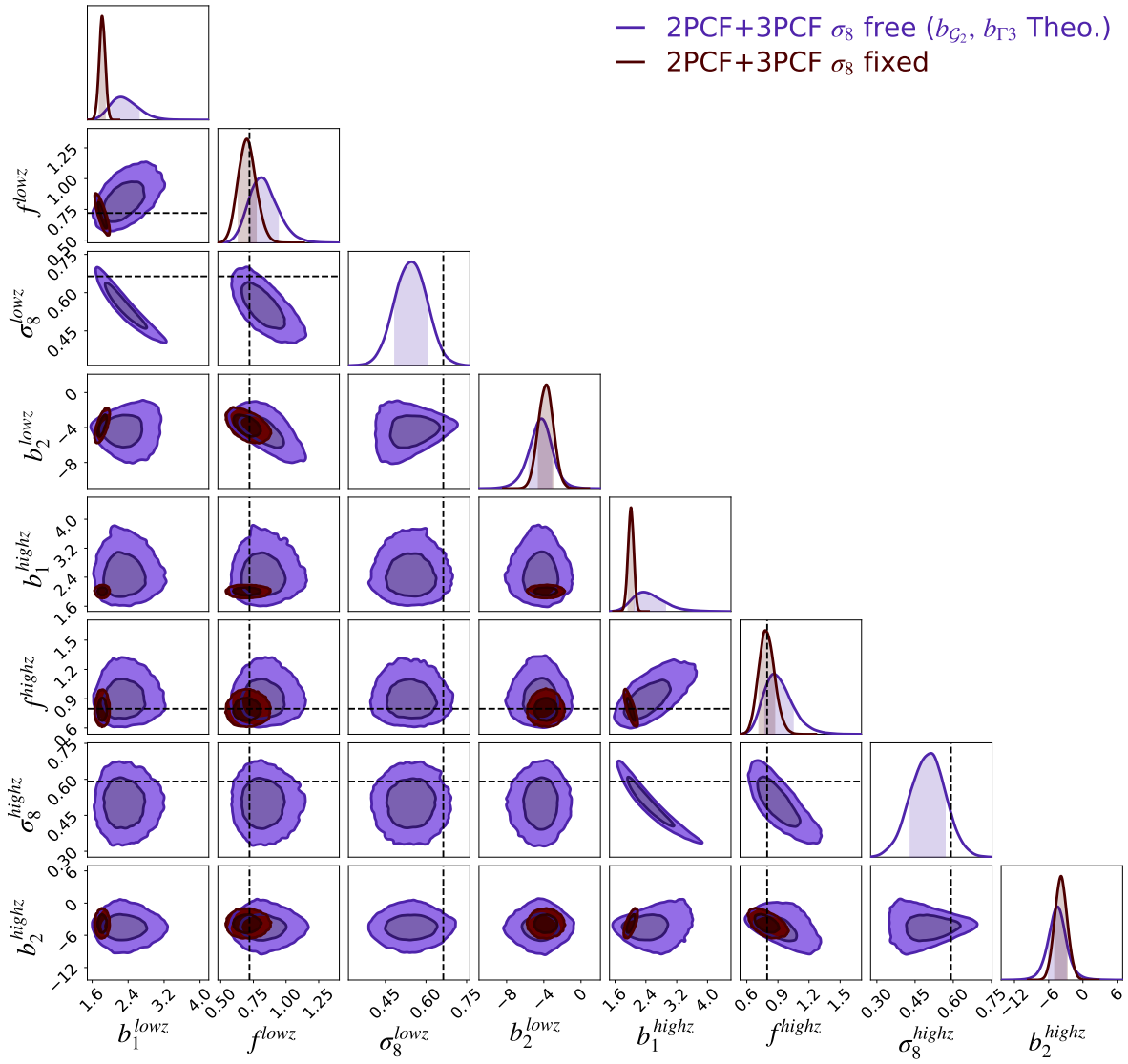

Figure 4.8: 2D marginalized constraints of two different chains. Contours show the 68% and 95% confidence regions. In purple, the contours of the 2PCF+3PCF fit from on $b_1 - f - \sigma_8 - b_2$ parameters. In dark brown, the contours of the 2PCF+3PCF analysis are presented at a fixed value of σ_8, as also shown in the previous section in Fig. 4.5. Black dashed lines indicate the fiducial values of f, σ_8 at $z = 0.38$ for low-z and $z = 0.61$ for high-z.

The BOSS 3PCF best-fit models shown in Fig. 4.10 display an excess of power at large scales relative to the model predictions. The residuals exhibit a clear pattern: triangles with the highest signal, which appear at the beginning of each sequence of $r_{\rm min}$ values, also tend to show the largest residuals. This is particularly evident at low triangle indices, corresponding to smaller scales. Such behaviour is expected, since these configurations are

associated with smaller values of η and probe smaller physical scales. While the statistical uncertainties are lower in this regime, the perturbative modelling is also expected to be less accurate. The reduced χ^2 values are consistent with unity, and their variation with respect to the σ_8 posterior occurs only at the second or third decimal place.

	2PCF+3PCF σ_8 free		**2PCF+3PCF σ_8 fixed**	
parameter	MLE	median	MLE	median
b_1^{lowz}	2.13	$2.30^{+0.34}_{-0.27}$	1.87	$1.82^{+0.07}_{-0.07}$
b_1^{highz}	2.27	$2.45^{+0.46}_{-0.35}$	1.96	$2.01^{+0.07}_{-0.07}$
σ_8^{lowz}	0.58	$0.54^{+0.06}_{-0.06}$	$\mathcal{F}[0.66]$	
σ_8^{highz}	0.53	$0.50^{+0.07}_{-0.07}$	$\mathcal{F}[0.59]$	
f^{lowz}	0.78	$0.82^{+0.12}_{-0.10}$	0.61	$0.70^{+0.07}_{-0.07}$
f^{highz}	0.86	$0.89^{+0.15}_{-0.12}$	0.81	$0.79^{+0.08}_{-0.07}$
b_2^{lowz}	-4.43	$-4.33^{+1.14}_{-1.24}$	-2.03	$-3.78^{+0.79}_{-0.81}$
b_2^{highz}	-4.91	$-4.48^{+1.50}_{-1.53}$	-4.55	$-3.85^{+1.11}_{-1.09}$

Table 4.5: Constrains on bias and cosmological parameters from the σ_8 free analysis on BOSS 2PCF+3PCF and 2PCF+3PCF with σ_8 fixed. The expectation values are given in terms of median (50th percentile) and 16th to 84th percentiles. While MLE values refer to the sample with the associated maximum likelihood.

2PCF		χ^2**/d.o.f**	
		σ_8 free	σ_8 fixed
low-z	NGC	1.12	1.19
	SGC	0.89	0.90
high-z	NGC	1.29	1.24
	SGC	1.47	1.54

3PCF		χ^2**/d.o.f**	
		σ_8 free	σ_8 fixed
low-z	NGC	1.142	1.149
	SGC	0.850	0.857
high-z	NGC	0.927	0.933
	SGC	0.979	0.982

Table 4.6: Posterior reduced χ^2 from median parameters of the "σ_8" free analysis on BOSS 2PCF+3PCF and 2PCF+3PCF with σ_8 fixed. The χ^2 is computed for the 2PCF multipoles (monopole-quadrupole-hexadecapole) with number of degrees of freedom $d.o.f = 68$ for the σ_8 fixed fit and $d.o.f = 67$ for the σ_8 free fit. The χ^2 is computed for the all-scales 3PCF with a number of degrees of freedom corresponding to the difference between the effective number of data points from pseudo-inversion of the covariance matrix and the number of free parameters in the fitting procedure.

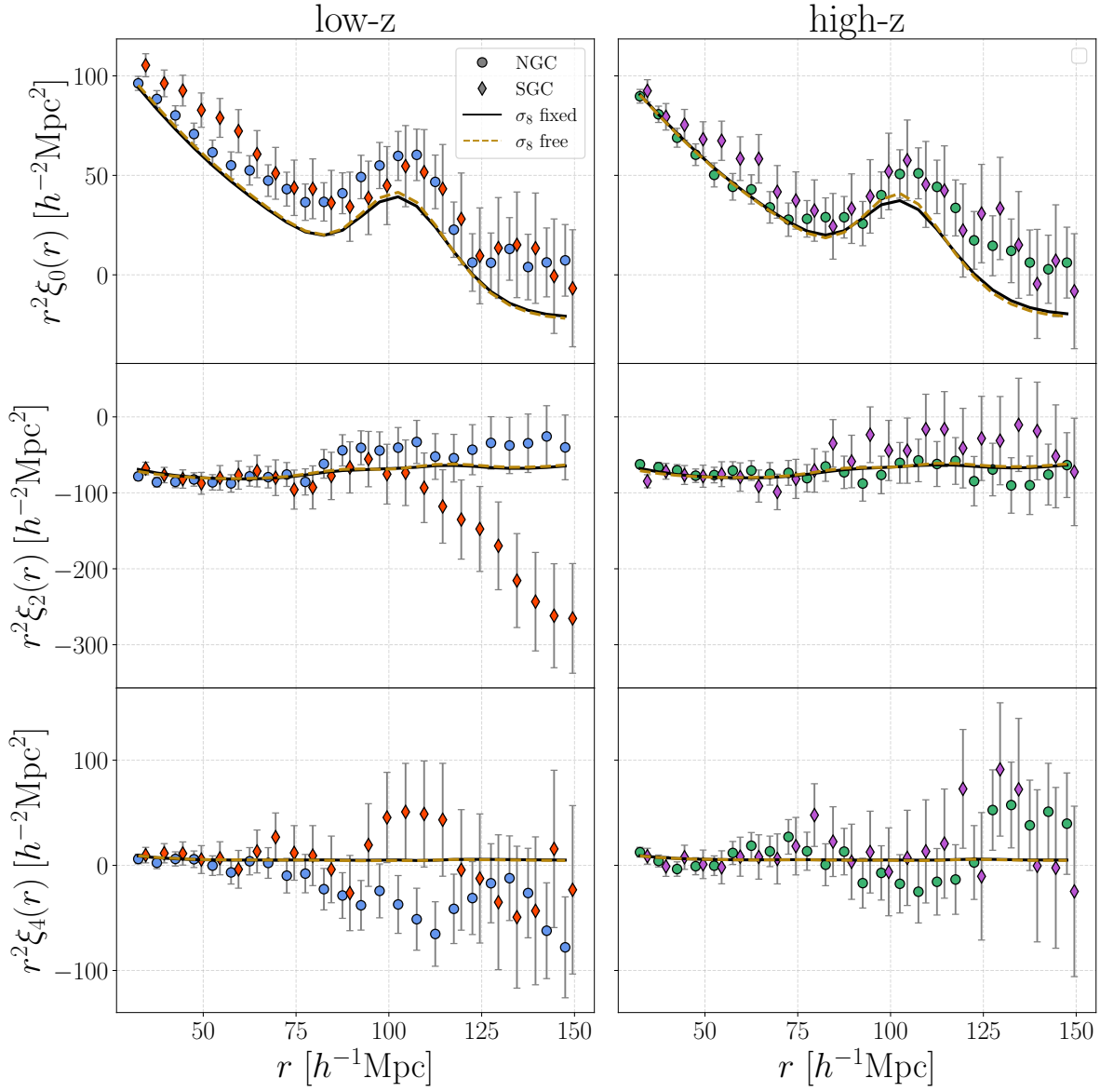

Figure 4.9: From top to bottom the BOSS 2PCF multipoles (monopole, quadrupole and hexadecapole from $r_{min} = 30h^{-1}Mpc$ to $r_{max} = 150h^{-1}Mpc$) alongside best-fit models extracted from median parameters of the σ_8 fixed and σ_8 free fitting of 2PCF+3PCF. The set of parameters that defines the model is the same for the NGC and SGC sub-samples. Error bars are computed as the square root of the covariance matrix diagonal and so represent the standard deviation associated with each data point.

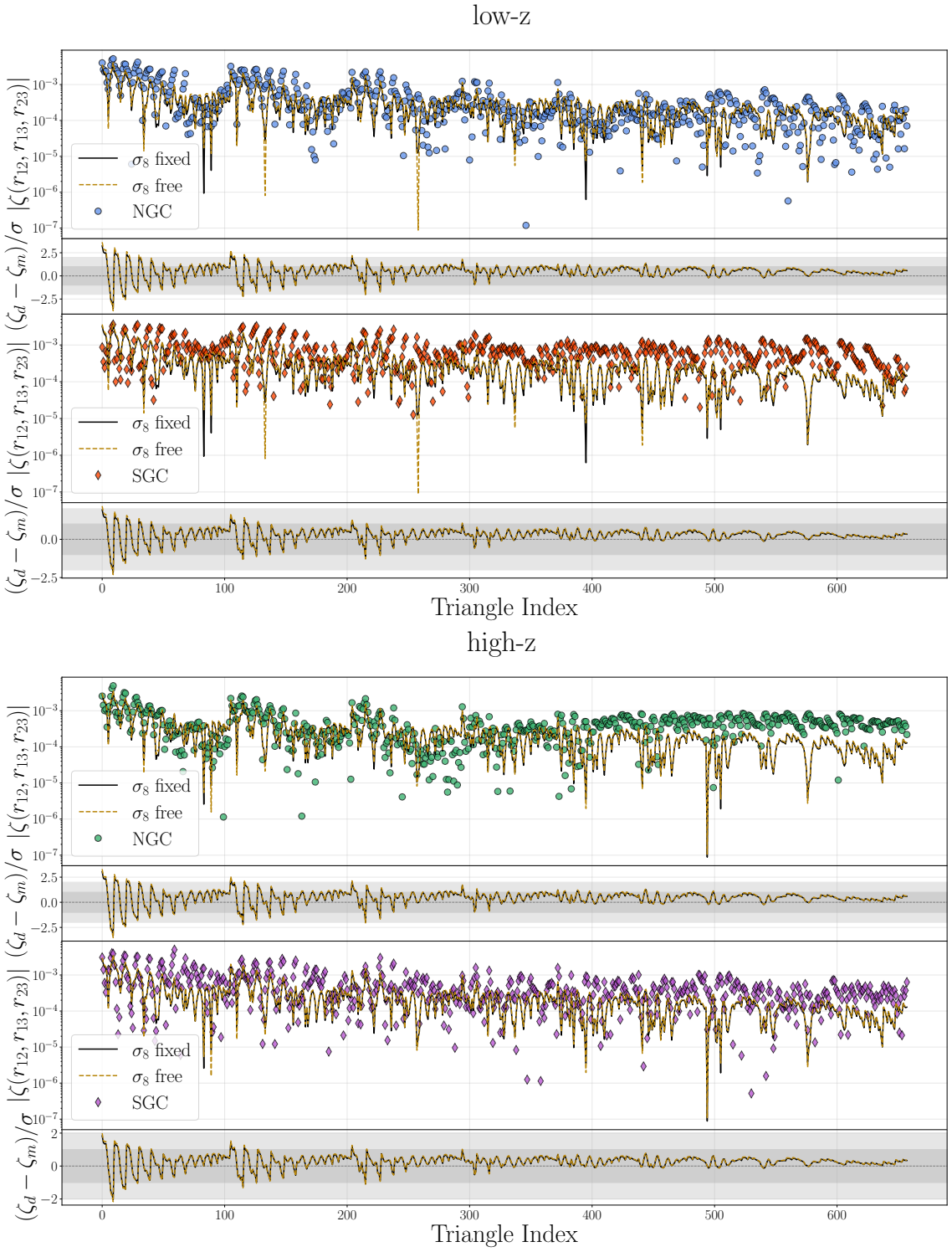

Figure 4.10: From top to bottom the BOSS all-scales 3PCF (all triangles from $r_{min} = 40\,h^{-1}Mpc$ to $r_{max} = 130\,h^{-1}Mpc$ and $\eta_{min} = 3$) of low-z and high-z samples. Alongside best-fit models extracted from median parameters of the σ_8 fixed and σ_8 free fitting of 2PCF+3PCF. Under each 'best-fit' is the residual plot showcasing the difference between models and data per unit σ, the standard deviation associated with each data point.

4.6 Problematics related to the BOSS reduced 3PCF

Finally, we explored the possibility of directly fitting the BOSS reduced 3PCF, which by definition combines 2PCF and 3PCF information, but did not obtain satisfactory results. In particular, even imposing the theoretical relations among the tidal bias parameters and fixing σ_8, in contrast with what is found from the analysis on the synthetic data vector, leads to unphysical expectation values, with $b_1 \approx 8$.

In Fig. 4.11, we compare the all-scales Q measured from the BOSS NGC high-z sample with three theoretical models corresponding to different values of the linear bias, $b_1 = \{2,\ 5,\ 8\}$, while keeping the remaining parameters fixed to

$$\{b_2 = 0,\ b_{\mathcal{G}2} = 0,\ b_{\Gamma 3} = 0,\ f = 0.78,\ \sigma_8 = 0.60\}.$$

From the residuals, models with extreme values of b_1 appear favoured, as they exhibit smaller differences with respect to the data. This effect is particularly pronounced for large-scale triangles ($r_{\rm min} \geq 67.5, h^{-1}$Mpc). The well-known tendency of BOSS correlation functions to display an excess of large-scale clustering relative to theoretical expectations becomes especially problematic for $Q^{(\rm all)}$. A closer inspection of this behaviour leads to the following considerations:

1. The full 3PCF scales as $\zeta \propto b_1^3$, whereas the reduced 3PCF obeys a different dependence. Since $Q \propto \zeta/\xi^2$ and $\xi \propto b_1^2$, one obtains $Q \propto b_1^{-1}$. In differential terms, $d\zeta^{(\rm all)}/db_1 \propto b_1^2$, while $dQ^{(\rm all)}/db_1$ depends only logarithmically on b_1. Consequently, ζ is significantly more sensitive to variations in b_1 than Q. Producing a substantial change in Q requires a much larger shift in b_1 than is necessary for ζ. As a result, a particularly noisy measurement of Q can drive the fit toward extreme and unphysical values of b_1.

2. The signal-to-noise ratio of Q remains relatively high for large-scale triangles (still below $120\,h^{-1}$Mpc), precisely where this issue is most evident. The enhanced large-scale correlation in Q further pushes the posterior toward increasingly large values of b_1.

For these reasons, we are unable to recover physically meaningful constraints on b_1 and f from the BOSS reduced 3PCF, $Q^{(\rm all)}$. This behaviour reflects the limitations of the covariance matrix, which is too noisy to yield robust constraints with the current BOSS data. We therefore defer a comprehensive analysis of the reduced 3PCF to future surveys probing substantially larger volumes.

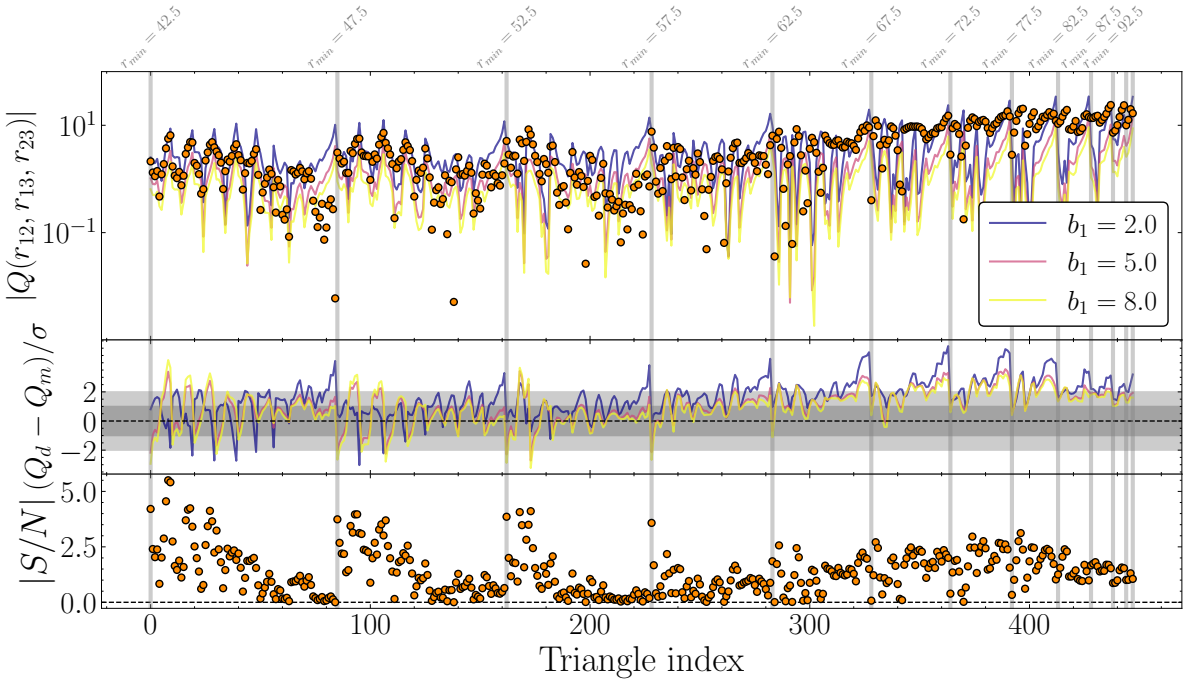

Figure 4.11: The upper panel shows the absolute value of NGC high-z reduced 3PCF for all triangles with scales between $40 - 120\,h^{-1}$Mpc, $\eta_{min} = 3$ (orange points). The data is compared with three models with changing b_1, while the other parameters are fixed $b_2, b_{\mathcal{G}2}, b_{\Gamma 3} = 0$, $f = 0.78$, $\sigma_8 = 0.60$. The middle panel shows the residuals for each model, i.e., for each triangle (data - model) over the uncertainty σ computed as the square root of the diagonal of the covariance matrix. The lower panel shows the 'per triangle' signal-to-noise ratio, where noise is the square root of the diagonal of the covariance matrix. Vertical gray bars denote the minimum scale for the side of each triangle.

4.7 Discussion

We compare our results on bias and cosmological parameters with other BOSS studies and other LSS surveys to provide a more comprehensive overview of current cosmological constraints from LSS. Most studies on clustering statistics use either 2PCF or power spectrum measurements to constrain a combination of the linear growth rate and the clustering amplitude smoothed over scales of $8\,h^{-1}Mpc$, i.e., they constrain the product $f\sigma_8$ down to some redshift z associated with the surveys measured statistic. We report the $f\sigma_8(z)$ values obtained from BOSS data at redshifts $z = 0.38$ and $z = 0.61$, using constraints from the σ_8-free analysis of the 2PCF+3PCF. However, since we constrain f and σ_8 separately, we must apply error propagation:

$$\delta(f\sigma_8) = \sqrt{(\sigma_8\delta f)^2 + (f\delta\sigma_8)^2 + 2f\sigma_8\, Cov(f,\sigma_8)} \tag{4.7.1}$$

This is not standard practice; in fact, most studies directly measure the $f\sigma_8$ product. The results are shown in Fig. 4.12, and are compared with other surveys and BOSS results. In our analysis, we obtain a $12-13\%$ uncertainty on $f\sigma_8(z)$ at both redshifts, in line with other BOSS studies like Satpathy et al. (2017), which reaches a $11-13\%$ uncertainty.

The real power of our analysis comes from having a separate estimate of both f and σ_8. This has only been done once for BOSS DR12, on the joint power spectrum and bispectrum (Gil-Marín et al. 2016), and in general, a small number of separate measurements are present in the literature. We report a $11-14\%$ uncertainty on σ_8 (low-z to high-z) and a $13-15\%$ uncertainty on f (low-z to high-z). Our σ_8 estimate is roughly 2σ away from the Planck prediction at $z = 0.38$ and 1.2σ at $z = 0.61$. This shift from the fiducial values may, however, be attributed to a residual degeneracy with b_1, whose best-fit value likewise moves toward higher values relative to the σ_8-fixed case.

Our f estimate is within 1σ from the Planck 2018 model prediction. Uncertainty regions are slightly larger than the other BOSS measurement (Gil-Marín et al. 2016) that was only made on the CMASS sample. This small difference can be attributed to the larger volume on CMASS rather than any of the single bins we are using. The other study that separates f and σ_8, using the joint 2- and 3-point correlation functions, comes from Veropalumbo et al. (2021). They measure the 2PCF and 3PCF of the VIPERS catalogue but exploit only a small number of triangle configurations from $r_{min} = 15\,h^{-1}$Mpc to $r_{max} = 40\,h^{-1}$Mpc, due to limitations in survey volume and modelling systematics. They provide an estimate for σ_8 at $\sim 20\%$ uncertainty and f at $65-75\%$ uncertainty. We also report the SDSS DR7 estimates of f and σ_8 down to redshift $z = 0.1$, obtained by Shi et al. (2018), which combines measurements of the 2PCF with galaxy-galaxy weak lensing on the same set of SDSS galaxies. Using this method, they break the $b_1 - f - \sigma_8$ degeneracy and report an estimate of f with a 9% uncertainty and σ_8 with a 13% uncertainty. Another measurement of σ_8 is provided by Marín et al. (2013), which measures the 3PCF of the WiggleZ survey. To obtain a reliable result, they marginalize over bias parameters by measuring them in a separate set of N-body simulations calibrated to match the survey properties. They estimate σ_8 at redshift $z = 0.35$, $z = 0.55$, $z = 0.68$ with an uncertainty of $\sim 10-20\%$.

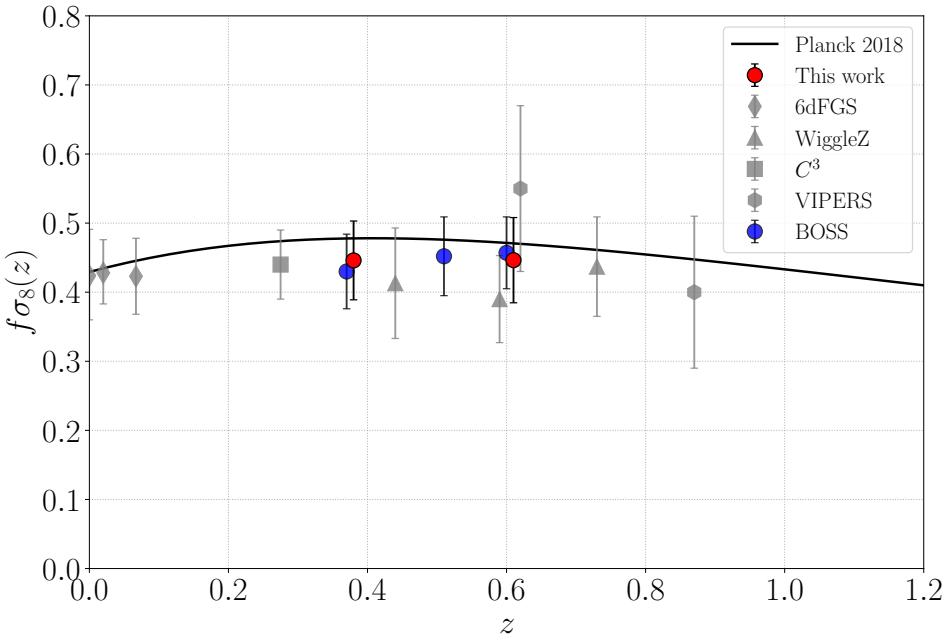

Figure 4.12: Comparison of our estimate (red circles) of the growth rate of structure, parameterised by the product $f\sigma_8(z)$, with previous measurements (using two-point statistics) from the literature: 6dFGS (Beutler et al. 2012,Huterer et al. 2017,Adams and Blake 2017); WiggleZ (Blake et al. 2012); "C^3 - Cluster clustering cosmology" (Marulli et al. 2021); VIPERS (Pezzotta et al. 2017); BOSS (Satpathy et al. 2017). The black solid line shows the ΛCDM + GR Planck Collaboration: Aghanim et al. (2020) model prediction.

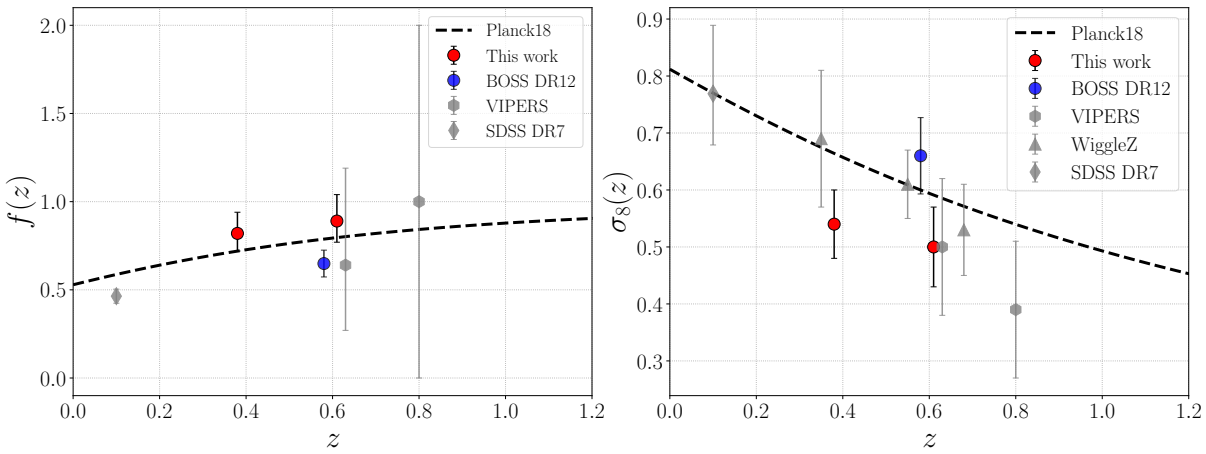

Figure 4.13: Values for $f(z)$ (left panel) and $\sigma_8(z)$ (right panel), as estimated from our joint 2PCF and 3PCF analysis (red circles), again compared to other measurements. BOSS: (blue) Gil-Marín et al. (2016) using a joint power spectrum and bispectrum measurement. VIPERS: Veropalumbo et al. (2021) with a joint 2PCF and 3PCF analysis. SDSS DR7: Shi et al. (2018) with a 2PCF+lensing measurment. WiggleZ: Marín et al. (2013) with a joint 2PCF and 3PCF analysis. The black dashed line shows the ΛCDM + GR Planck Collaboration: Aghanim et al. (2020) model prediction.

4.7.1 Bias parameters

Bias parameters are a key ingredient of clustering statistics models, and it is therefore worthwhile to discuss our results in the context of other surveys and theoretical expectations. Figure 4.14 shows the linear bias as a function of redshift from this work, together with several measurements from the literature. The measurements of b_1 are compared to theoretical expectations computed numerically using the `CosmoBolognaLib` library (Marulli et al. 2016).

We compute the effective large-scale bias of dark matter haloes as

$$b_{\rm eff}(z) = \frac{\int_{M_{\rm min}}^{M_{\rm max}} {\rm d}M\, b(M,z)\, \Phi(M,z)}{\int_{M_{\rm min}}^{M_{\rm max}} {\rm d}M\, \Phi(M,z)}, \tag{4.7.2}$$

where $\Phi(M,z)$ is the halo mass function (Tinker et al. 2008), and $b(M,z)$ is the bias of haloes of mass M at redshift z, computed using the prescription of Tinker et al. (2010).

Since different surveys probe different galaxy populations, we do not expect measurements of the linear bias to match across surveys. Nevertheless, all measurements show the same overall trend: an increase of bias with redshift. This behaviour is consistent with expectations for the evolution of haloes of fixed mass: massive objects are rarer at earlier cosmic times. They are therefore more strongly biased tracers of the underlying matter distribution.

All bias parameters reported here are derived from measurements of the 3PCF, or from joint analyses of the 2PCF and 3PCF in the respective surveys. Marín et al. (2013) show that the WiggleZ emission-line galaxies (ELGs) constitute a clearly less biased galaxy population. Veropalumbo et al. (2022) measures the linear bias of VIPERS galaxies, which do not correspond to a well-defined galaxy population, as the sample is selected through a magnitude cut applied to bright galaxies in the intermediate redshift range $0.5 < z < 1.2$. Marín (2011) measures the bias of the DR7 SDSS Luminous Red Galaxies, which represent a highly biased galaxy population.

Slepian et al. (2017) use the 3PCF alone to constrain the linear bias of BOSS DR12 CMASS galaxies ($z = 0.57$). Fixing σ_8 to a fiducial ΛCDM model, similarly to our analysis, they find a $\sim 4\%$ uncertainty on b_1. The scale cuts adopted in these surveys typically reach minimum separations of $r_{\rm min} = 15 - 20\,h^{-1}$Mpc. In contrast, we adopt more conservative scale cuts of $r_{\rm min} = 30 - 40\,h^{-1}$Mpc, while still achieving an excellent precision in the measurement of the linear bias of BOSS galaxies.

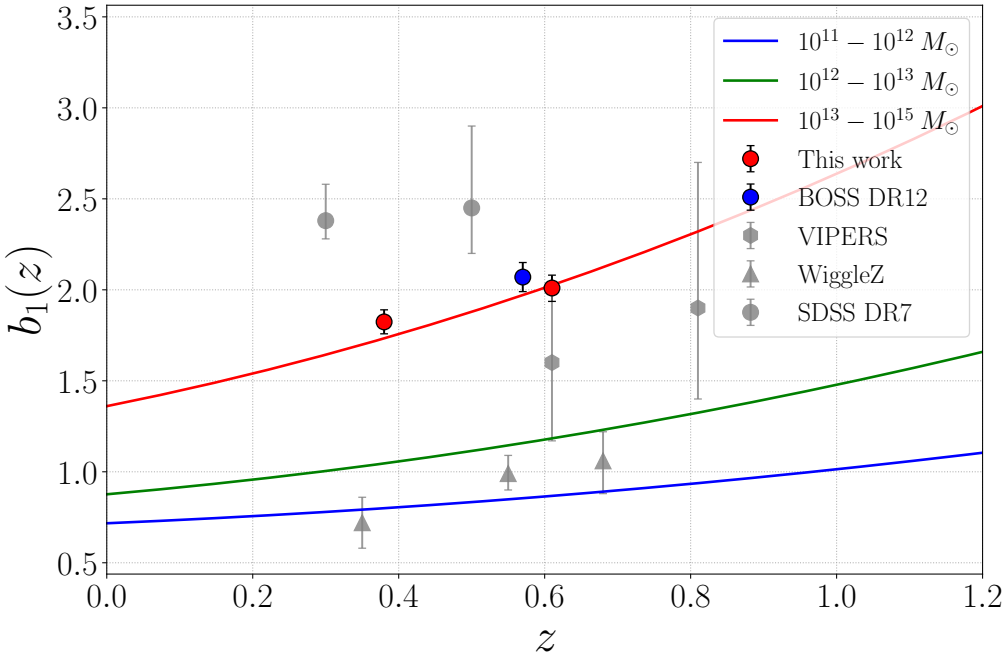

Figure 4.14: Linear bias as a function of redshift. Red markers are extracted from our 1D marginalized posteriors of the BOSS joint 2PCF and 3PCF, with the associated error bars representing the 16th to 84th percentiles. Other works are included for comparison: BOSS DR12-3PCF, Slepian et al. (2017); VIPERS, Veropalumbo et al. (2021); WiggleZ, Marín et al. (2013); SDSS, Marín (2011). The colored curves are linear bias relations, obtained from different mass ranges and computed with the Tinker mass function and bias prescription.

It is important to relate measurements to theoretical predictions to refine the physical models that describe bias. Adding the 3PCF to LSS analysis helps break degeneracy between bias parameters, and being able to achieve larger precision will help investigate these models.

In fig. 4.15 we show the tidal bias parameters as a function of linear bias as extracted from 1D marginalized posteriors of the BOSS 2PCF+3PCF with σ_8 fixed. These results are compared with theoretical relations. Namely, the Lagrangian Local In Matter density relations (see Eq. 3.4.3 and 3.4.4) and the quadratic fit obtained by Eggemeier et al. (2020) to the excursion-set predictions of Sheth et al. (2013)

$$b_{\mathcal{G}2}^{\text{ex-set}} = 0.524 - 0.547\, b_1 + 0.046\, b_1^2, \tag{4.7.3}$$

where the equivalent $b_{\Gamma 3}$ prediction is obtained by plugging the quadratic fit into Eq. 3.4.4. The LLIMD approximation is known to be in tension with high-precision simulation results, which provide clear evidence for the presence of tidal Lagrangian bias (Lazeyras and Schmidt 2018, Abidi and Baldauf 2018). This discrepancy arises because LLIMD relies on the spherical collapse model and therefore neglects tidal fields at the time of halo formation, but gravitational evolution induces a non-zero tidal field even if it is not present at the time of halo formation. This relation has been used in this work to limit the parameter space of the σ_8 free 2PCF+3PCF fit, to ensure its convergence. The error bars

for $b_{\Gamma 3}$ show a within 1σ agreement with our measurement, while $b_{\mathcal{G}2}$ is almost $\sim 2\sigma$ away from LLIMD. Our measurements suggest the presence of a large Lagrangian tidal bias. However, these results have been obtained with σ_8 fixed to the Planck 2018 cosmology and are therefore underestimating confidence intervals. At the level of precision of BOSS data, significant deviations from LLIMD cannot be detected (Philcox and Ivanov 2022), though this will likely change with the advent of *DESI* and *Euclid*.

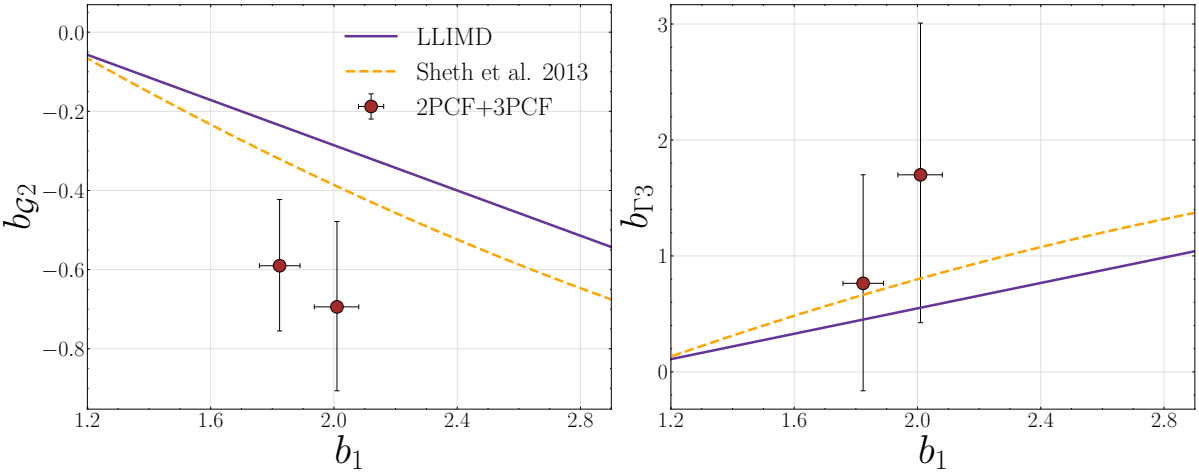

Figure 4.15: Tidal biases as a function of linear bias, extracted from 1D marginalized posteriors of the BOSS 2PCF+3PCF with σ_8 fixed. The associated error bars represent the 16th to 84th percentiles. The purple curves represent the LLIMD relation, and the dashed yellow line is the excursion-set prediction of Sheth et al. 2013.

Chapter 5

Conclusions and future prospects

The current cosmological paradigm, the ΛCDM model, while being extremely successful in describing many observables in the Universe, still presents open and pressing questions regarding the physical origin of its most prominent components, dark matter and dark energy. With the advent of Stage IV surveys (Euclid Euclid Collaboration et al. 2025, DESI DESI Collaboration et al. 2016, Nancy Grace Roman Space Telescope Doré et al. 2019) and the prospect of Stage V surveys (e.g., WST Mainieri et al. 2024), the Universe will be mapped over an unprecedented volume with unparalleled statistical power, providing a significant increase of precision in the determination of cosmological parameters that will test the assumptions of this model. One of the biggest revolutions in cosmology has been the observation that the expansion of the Universe is accelerating (Riess et al. 1998, Perlmutter et al. 1999, Eisenstein et al. 2005). Whether this acceleration is caused by a cosmological constant, a dynamical process or a modification to the laws of General Relativity remains to be seen.

A measurement of the expansion history via the distance-redshift relation provides a direct constraint on the equation of state of dark energy. In parallel, a detailed measurement of the growth rate as a function of redshift, $f(z)$, can also shed light on the nature of dark energy and the underlying theory of gravity. Both can be constrained by LSS statistics such as the two-point correlation function or its Fourier counterpart, the power spectrum, by means of geometric distortions, quantified by the Alcock Pacinsky (AP) test, and Redshift Space Distortions (RSD), respectively. While 2-point statistics have been widely used in the literature, higher-order correlation functions, starting from the 3-point correlation function and bispectrum, have been far less used. These statistics are sensitive to the non-Gaussianity that develops as a consequence of gravitational instability. Furthermore, they are sensitive to galaxy biasing in a complementary way to 2-point statistics and can thus be used to break degeneracies between them.

Measuring the 3PCF and bispectrum in conjunction with the 2PCF and power spectrum can probe both, and adds significantly to the constraints by breaking the degeneracy between f and σ_8 (Marín et al. 2013, Gil-Marín et al. 2016). Also, the 3PCF has recently seen detections of BAO features (Slepian et al. 2017, Moresco et al. 2021) usable as a standard ruler. While in practice, density field reconstruction is already used to push back information from higher-order correlations to 2-point statistics, the 3PCF remains

a complementary probe of BAO as it is methodologically independent.

The information content of the 3PCF is formally equivalent to that of its Fourier-space counterpart, the bispectrum. However, being defined in configuration space, the 3PCF accounts more naturally for survey geometric effects. Its physical interpretation is also more straightforward, as it probes correlation over distances rather than Fourier modes. Despite this potential, a very limited number of studies have been carried out so far, and these have focused only on the bispectrum (in BOSS: Gil-Marín et al. 2016, Philcox and Ivanov 2022). On the other hand, measuring and modelling the 3PCF is significantly more computationally expensive. Estimating the 3PCF requires direct triplet-counting and theoretical models are first evaluated in Fourier space, requiring an anti-Fourier transform of the bispectrum model.

In this thesis, we have studied for the first time the full potential of the 3PCF ζ and reduced 3PCF Q on the BOSS DR12 catalogue. The number of triangle configurations used is unprecedented. We have used measures, taken by Guidi et al. (including L. Cavazzini) (2026, in prep.), of the isotropic 3PCF multipoles up to $\ell_{max} = 10$ and 2PCF monopole, quadrupole and hexadecapole, of the BOSS catalogue and for a set of 2 000 MD-`Patchy` mock catalogues (Kitaura et al. 2016). The measures were taken for the optimally binned BOSS samples constructed for the DR12 in two spatially separate redshift bins at $z_{\rm eff} = 0.38$ (low-z) and $z_{\rm eff} = 0.61$ (high-z). These samples have similar effective volumes ($\sim 4\,$Gpc) and are made to provide uniform constraints across the two redshift bins (Alam et al. 2017). The measures were taken with the novel code `MeasCorr` (Farina et al. 2026) which implements the SHD algorithm (Slepian and Eisenstein 2015) for fast computation of the 3PCF multipoles. We combined the measured 2PCF and 3PCF to construct the reduced 3PCF and derived its associated covariance matrix for all available triangles.

We computed theoretical predictions of correlation functions with `MElCorr` (Guidi et al. (including L. Cavazzini) 2026, in prep.) which implements an emulator for the 2PCF, required for the fast evaluation of ξ templates when AP parameters are free to vary. Soon, this code will implement an emulator for the 3PCF allowing for AP parameterization and *full-shape* analyses. Once again, we combine theoretical predictions of the 2PCF and 3PCF to construct a model of the reduced 3PCF. This work focuses on the *template fitting* approach in which the power spectrum shape parameters are fixed to a fiducial cosmology (Planck Collaboration: Aghanim et al. 2020) and cosmological information is extracted by means of RSD and AP parameters. In practice, the terms that make up the 3PCF are a linear combination of pre-computed templates and fitting parameters.

Results

A fundamental step was the development of a robust and versatile Bayesian framework to fit the BOSS correlation functions that includes different features. Within this framework, it is possible to analyze the BOSS correlation functions, combining different datasets and adopting different parameter combinations as described in Sec. 3.1. This code allows to analyze either separately or jointly 2PCF and 3PCF, estimating the models for the derived

reduced 3PCF, creating synthetic data vectors, and sampling the posterior considering different parameters and priors. We used this framework to fit the 2PCF multipoles and all scales 3PCF with synthetic data to validate the robustness of the code and the validity of results, and also to explore possible parameter degeneracies. We have found that the most effective strategy to deal with the $b_1 - f - \sigma_8$ degeneracy is to fix tidal bias parameters ($b_{\mathcal{G}2}$, $b_{\Gamma 3}$) to theoretical relations. We also run an extensive campaign of analyses to assess the ranges of validity of models in the 2PCF and 3PCF that would minimize modelling systematics while preserving the most signal-to-noise ratio. We find that a good compromise between the two is for the 2PCF $r_{min} = 30\,h^{-1}$Mpc, $r_{max} = 150\,h^{-1}$Mpc; for the all-scales 3PCF $r_{min} = 40\,h^{-1}$Mpc, $r_{max} = 130\,h^{-1}$Mpc; for the reduced all-scales 3PCF $r_{min} = 40\,h^{-1}$Mpc, $r_{max} = 120\,h^{-1}$Mpc. We used the same scale cuts for the joint 2PCF+3PCF as the individual analyses. These results are consistent with methodological analyses of the 3PCF (Veropalumbo et al. 2022, Guidi et al. 2023).

We summarize below the key results obtained from fitting the four BOSS data sets in a combined Bayesian analysis.

Measurement of distance with the 2-point correlation function. The results from fitting the 2PCF multipoles yield constraints for the AP parameters that are used for the measurement of distance. We derived constraints on $D_M(z)(r_d/r_{d,\rm fid})$ and $H(z)(r_d/r_{d,\rm fid})$ and find a good agreement with flat ΛCDM, preferring CMB-related constraints rather than local ones. We also compared our results with the BOSS BAO consensus, which, by means of reconstruction of the density field, was able to get to the $1-2\%$ precision goal. We also achieved a $1-2\%$ uncertainity on $D_M(z)$ with values closely centered on the BOSS BAO consensus; our values of $H(z)$ have a $\sim 3-4\%$ uncertainty.

Fit of the joint 2- and 3-point correlation functions with σ_8 fixed. We constrained bias and cosmological parameters for the 2PCF multipoles, all scales 3PCF and the joint 2PCF+3PCF with σ_8 fixed to fiducial Planck 2018 values. While the 2PCF alone constrains b_1 and f with good accuracy, the higher-order bias parameters are wildly unconstrained. We find that combining the 3PCF with the 2PCF successfully breaks the degeneracy among higher-order bias parameters and that overall leads to tighter constraints on b_1 and f. The marginalized posterior on b_1 shrinks by about $40-70\%$ while f shrinks by $17-40\%$ (low-z to high-z) from the 2PCF to the joint 2PCF+3PCF. In Fig. 5.1, the constraints for b_1 and f from the four main posteriors made for the BOSS combined likelihoods are summarized.

Fit of the joint 2- and 3-point correlation functions with σ_8 free. We conducted a Bayesian analysis of the joint 2- and 3-point correlation of the BOSS combined samples with a uniform prior on σ_8. By fixing tidal bias parameters to theoretical relations, we successfully broke the degeneracy among $b_1 - f - \sigma_8$. This is not possible with the use of the 2PCF alone since it constrains the product $b_1\sigma_8$. Standard analyses model the 2PCF with parameters multiplied by σ_8, thus constraining the $f\sigma_8$ product. We report a $11-14\%$ uncertainty on σ_8 (low-z to high-z) and a $13-15\%$ uncertainty on f (low-z to high-z). Our σ_8 estimate is roughly 2σ away from the

Planck prediction at $z = 0.38$ and 1.3σ at $z = 0.61$. Our f estimate is within 1σ from the Planck 2018 model prediction. We have shown that the quadratic bias b_2 is essentially decoupled from the other parameters in the joint 2PCF+3PCF. We have adopted the Local In Matter Density (LLIMD) theoretical relation to link the linear bias to tidal parameters $b_{\mathcal{G}2}$, $b_{\Gamma 3}$. These parameters are not fully decoupled in the joint 2PCF+3PCF, and their presence in the fit results in large volume effects. We have shown that uncertainties for these parameters are too large to detect any meaningful departure from the LLIMD theoretical relations. Future surveys that achieve larger volumes will better characterize theoretical relations of bias.

Analysis of the reduced 3-point correlation function. We find that at the current level of precision, the all-scales reduced 3PCF $Q^{(all)}$ is too noisy to get any reliable constraints. A large degeneracy between b_1 and f persists even when fixing the tidal biases to theoretical relations. Theoretical models of Q are particularly insensitive to changes of b_1 (since $Q \propto b_1^{-1}$); this dependency is responsible for the shift of b_1 posterior to extremely high values ($b_1 \sim 8$). This behaviour is enhanced by a power excess at large scales that is typical of BOSS correlation functions.

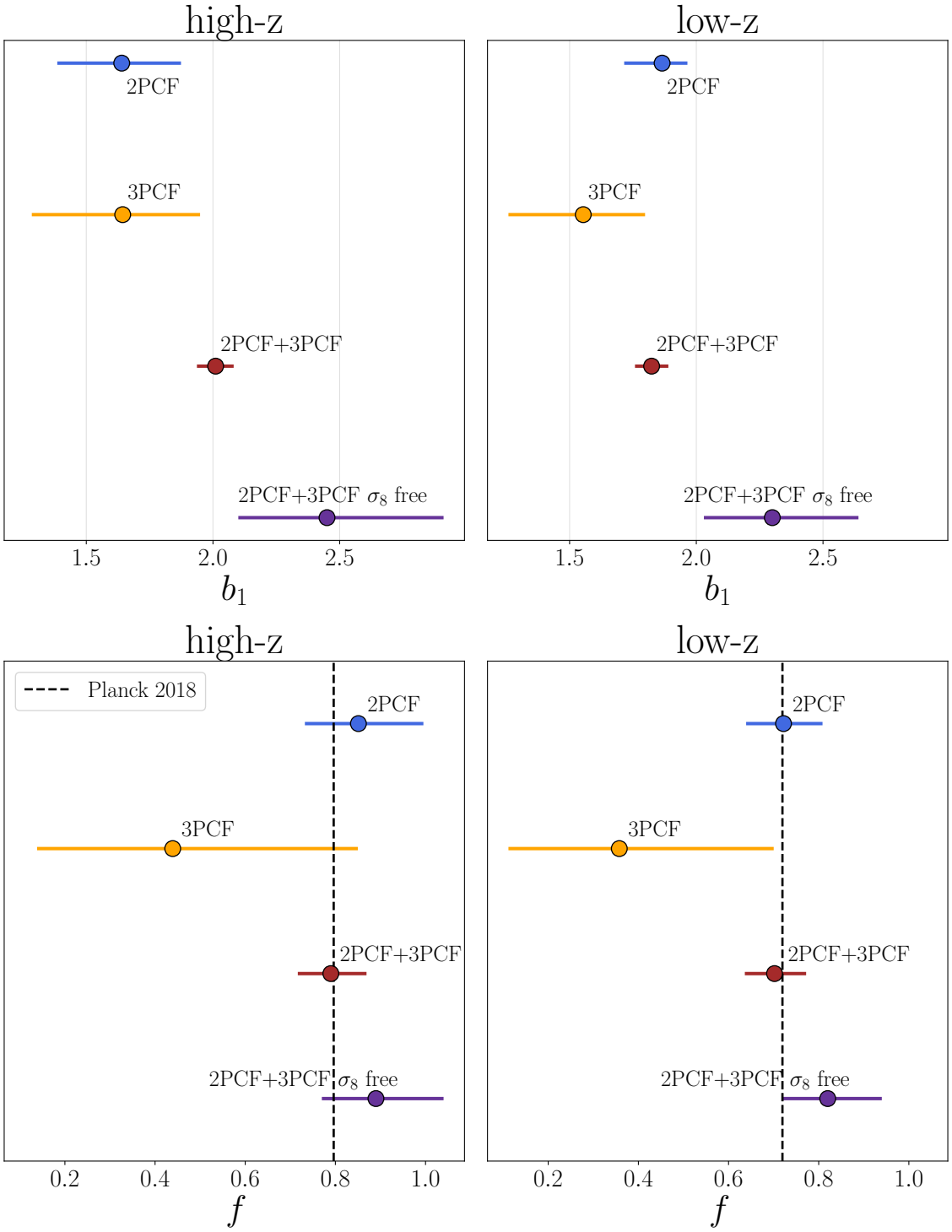

Figure 5.1: Summary of expectation values for the parameters b_1 (top) and f (bottom) from four different marginalized posterior distributions. The black dashed line indicates the value of f for Planck 2018 cosmology (Planck Collaboration: Aghanim et al. 2020) at redshift $z_{\text{eff}} = 0.38$ low-z and $z_{\text{eff}} = 0.61$ high-z.

Future Prospects

Building on the analysis presented in this Thesis, several future directions can be pursued. We outline below the potential extensions of this study.

Anisotropic 3PCF. In this thesis, we have considered the isotropic part of the 3PCF. In redshift space, the 3PCF also has an anisotropic component, because the action of RSDs depends on the orientation of triplets with respect to the line of sight. An extension of this work aiming to maximize the extraction of information in redshift space would therefore naturally include the measurement of the anisotropic multipoles of the 3PCF. The code `MeasCorr` provides an implementation of the algorithm for estimating the anisotropic 3PCF multipoles presented in Slepian and Eisenstein (2018), which generalizes the isotropic case while retaining the same $\mathcal{O}(N^2)$ scaling. Methodological analysis of the anisotropic 3PCF (Farina et al. 2025) has shown that including anisotropic components tightens constraints on the growth rate f, with the largest improvement achieved by including AP parameters.

Full AP analysis of the 3PCF. In the thesis, we provide a measurement of distance by the inclusion of AP parameters in the 2PCF model. The scale dependence of the theoretical model must be rescaled to compare the theoretical model calculated using the true cosmological parameters with the measurement using the 'fiducial' cosmological parameters. This fact makes it unsuitable for the template fitting approach which involves pre-building a table of the results of the 2PCF calculation. We have mitigated this problem by the use of an emulator, `MElCorr`, which allows for the fast estimation of theoretical predictions even on a few CPUs. A natural progression of this work would be to allow AP parameterisation of the 3PCF model. Sugiyama et al. (2021) come up with an approximate method for rescaling the wavevectors of the anisotropic bispectrum templates. However, the method still has a high computational cost for the evaluation of configuration space templates and the most promising results ultimately come from emulators (Euclid Collaboration: Guidi et al. 2025).

Full-shape analysis of the BOSS 3PCF. The complementary work to this thesis is Guidi et al. (including L. Cavazzini) (2026, in prep.) which makes the first *full-shape* study of the isotropic 3PCF on the BOSS catalogue. We have already outlined the challenges involved in the modelling of the 3PCF and the need of emulation. The first *full-shape* study of the 3PCF in real space has been made on numerical simulations by Euclid Collaboration: Guidi et al. (2025) were it was found that using the 3PCF complementarily to the 2PCF improves the constraints on shape parameters and breaks the degeneracy between A_s and b_1.

Beyond ΛCDM analysis. Configuration space studies of higher-order statistics can be used to improve understanding of a variety of phenomena. BAO studies that exploit complementary information on the 3PCF can be used to discard dynamical dark energy models (Aubourg et al. 2015) or modified gravity models. The departure from Gaussianity in the primordial field of fluctuations, often referred to as

Primordial Non-Gaussianity (PNG) is a phenomenon particularly well suited for higher order statistics and studies in configuration space are still lacking. The same can be said regarding neutrino mass constraints; their free streaming manifests as a loss of power at small scales, but this signal is degenerate with galaxy bias and the amplitude of perturbations. Adding 3-point statistics helps remove degeneracies and therefore enhances neutrino mass constraints. So far, most studies have been made in Fourier space, focusing on the bispectrum (Kamalinejad and Slepian 2025b, Kamalinejad and Slepian 2025a) while the 3PCF is still catching up (Labate et al. 2025).

As a final remark, the 10-fold increase in volume that *Euclid* and *DESI* provide makes higher-order statistics a crucial ingredient to be included, since so far, their usefulness has been hampered by the limited volume of past surveys. This work demonstrates the importance of their implementation in pipelines for cosmological analysis as a complementary probe of the newly incoming data. Of particular interest will be the study of the reduced 3PCF, which has been severely limited by the estimation of the covariance matrix. Furthermore, future surveys (Stage V) will map an even greater volume: WST (Mainieri et al. 2024) will obtain medium to high resolution spectroscopic data up to very high redshift ($z \sim 7$) and will provide a plethora of data in an era of the universe that has so far been unexplored.

Appendix A

Perturbation theory diagrammatics

Perturbation Theory (PT) provides a natural way of drawing the contributions of clustering statistics in a similar way of Feynman diagrams in quantum field theory. In this approach, contributions to $N-$point cumulants of the density contrast $\delta(\boldsymbol{k})$ (or any other cosmic field) come from connected diagrams with N external (solid) lines and $r = N-1, N\ldots$ internal (dashed) lines. The r index defines the order that is being adopted and goes from:

- leading order, called *tree-level (TL)*
- next to leading order, called *one-loop (1L)*
- next to one loop, called *two-loop (2L)*, and so on.

In essence, what the diagrammatic representation does is to order the contributions irrespective of the statistical quantity being considered. For example, it is not consistent to consider the evolution of the power spectrum in second-order PT since there is a contribution of the same order coming from third-order PT. Instead, one should consider the evolution of the power spectrum to "one-loop" PT. If, in fact, we consider an expansion of the density field up to second order in PT and construct the power spectrum, it is easy to demonstrate that all odd terms $\mathcal{O}(\delta^n)$ with n odd vanish by consequence of Wick theorem. If we want the next-to-leading order power spectrum, we must expand the density field up to third-order in PT, where all terms that contribute are $\mathcal{O}(\delta^4)$.

In this approach (Goroff et al. 1986, Scoccimarro and Frieman 1996), the **r**-order contribution to the **N**-point spectrum of the overdensity field is obtained through the following rules:

- Draw all distinct graphs consisting of N vertices, N external lines, each labeled $\{k_1, ..., k_N\}$, and connected by r internal lines, in turn labeled $\{q_1, ..., q_r\}$. Distinct graphs are defined as diagrams that are not continuously deformable into one another. For each of these diagrams:
  1. Each internal line corresponds to an integration $\int \frac{d^3q_j}{(2\pi)^3} P(q_j)$, where q_j is the relative wave vector.
  2. Assign a factor $(2\pi)^3 \delta_D(k_i + q_j + ... + q_{j+n}) F_n(q_j, ..., q_{j+n})$ to each vertex of order n, i.e., from which n internal lines branch, $\{q_j, ..., q_{j+n}\}$. The convention is to assign a positive sign to the wave vectors associated with outgoing internal lines and a negative sign to those associated with internal lines entering the vertex.
  3. Multiply by the symmetry factor, i.e., the number of permutations of the linear overdensities $\delta^{(1)}$ that leave the diagram unchanged.
  4. Add over all the distinct labellings of the external lines, thus generating $\frac{N!}{n_1!...n_{2r-N+1}!}$ diagrams, where n_i denotes the number of vertices of order i.
- Add the resulting expressions from each diagram.

In what follows we derive the contributions of the *tree-level* power spectrum, *one-loop* power spectrum and *tree-level* bispectrum.

Tree-level power spectrum. N = 2, r = N − 1 = 1

The corresponding diagram, shown in Fig. A.1 has two external labellings $\{\boldsymbol{k}_1, \boldsymbol{k}_2\}$ connected by one internal line $\boldsymbol{q}_1$; the order in the expansion of δ_n that is required is given by how many internal lines join a node which for the tree-level power spectrum is only first order in δ for each node.

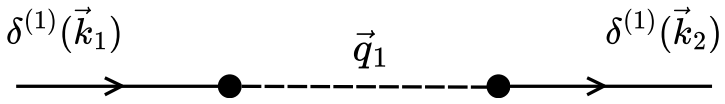

Figure A.1: diagrammatic representation of the tree-level power spectrum.

$$\langle\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{TL} = \int \frac{d^3\boldsymbol{q}_1}{(2\pi)^3} P_{lin}(q_1)(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{q}_1)F_1(\boldsymbol{q}_1)(2\pi)^3\delta_D(\boldsymbol{k}_2-\boldsymbol{q}_1)F_1(-\boldsymbol{q}_1)$$

since $F_1 = 1$ the expression reduces to

$$\langle\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{TL} = (2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)P_{lin}(k_1) \tag{A.0.1}$$

which is the definition of linear power spectrum obtained in Sec. 1.4.1.

One-loop power spectrum. N = 2, r = 2
The two topologically distinct diagrams of the one-loop power spectrum are shown in Fig. A.2. Each of them is made with two external lines $\{\boldsymbol{k}_1, \boldsymbol{k}_2\}$ and connected by two internal lines $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$.

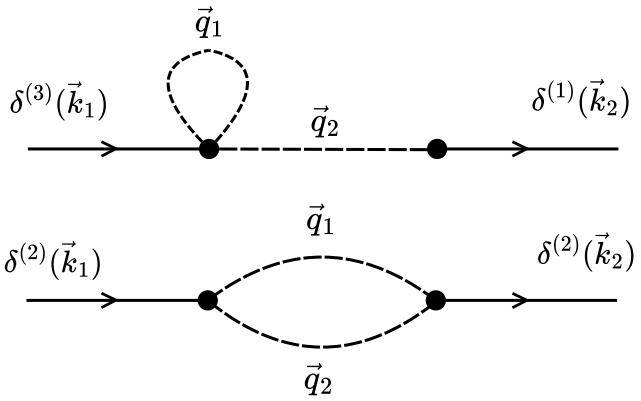

Figure A.2: diagrammatic representation of the one-loop power spectrum.

$$\begin{aligned}
\langle\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{22} =& 2\int d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 P_{lin}(q_1)P_{lin}(q_2)\delta_D(\boldsymbol{k}_1+\boldsymbol{q}_1+\boldsymbol{q}_2)F_2(\boldsymbol{q}_1,\boldsymbol{q}_2)\\
&\delta_D(\boldsymbol{k}_2-\boldsymbol{q}_1-\boldsymbol{q}_2)F_2(-\boldsymbol{q}_1,-\boldsymbol{q}_2) =\\
=&\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)\, 2\int d^3\boldsymbol{q}_1 P_{lin}(q_1)P_{lin}(k_2-q_1)F_2^2(\boldsymbol{q}_1,\boldsymbol{k}_2-\boldsymbol{q}_2) =\\
=&(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)\, P_{22}(k_1)
\end{aligned}$$

the simmetry factor is equal to two.

$$\begin{aligned}
\langle\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\rangle_{13} =& 6\int d^3\boldsymbol{q}_1 d^3\boldsymbol{q}_2 P_{lin}(q_1)P_{lin}(q_2)\delta_D(\boldsymbol{k}_1+\boldsymbol{q}_2)F_3(\boldsymbol{q}_1,\boldsymbol{q}_2,-\boldsymbol{q}_1)\delta_D(\boldsymbol{k}_2-\boldsymbol{q}_2)F_1(-\boldsymbol{q}_2)\\
=&\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)\, 6P_{lin}(k_2)\int d^3\boldsymbol{q}_1 P_{lin}(q_1)F_3(\boldsymbol{q}_1,\boldsymbol{k}_2,-\boldsymbol{q}_1)\\
=&(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2)\, P_{13}(k_1)
\end{aligned}$$

Summing the two contributions to the tree-level one, we get the power spectrum up to one-loop

$$P_{1L} = P_{lin} + P_{22} + P_{13} \tag{A.0.2}$$

Tree-level bispectrum. N = 3, r = 2
The bispectrum is the first statistic to have three external lines $\{\boldsymbol{k}_1, \boldsymbol{k}_2, \boldsymbol{k}_3\}$ and at leading order they are connected by two internal lines $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$. Its diagram is shown in Fig. A.3.

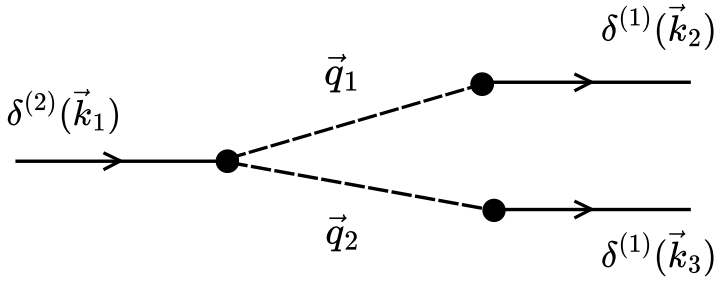

Figure A.3: diagrammatic representation of the tree-level bispectrum.

$$\begin{aligned}\langle\delta(\boldsymbol{k}_1)\delta(\boldsymbol{k}_2)\delta(\boldsymbol{k}_3)\rangle_{TL} =& 2(2\pi)^3 \int d^3\boldsymbol{q}_1 P_{lin}(q_1)P_{lin}(q_2)\delta_D(\boldsymbol{k}_1+\boldsymbol{q}_1+\boldsymbol{q}_2)F_2(\boldsymbol{q}_1,\boldsymbol{q}_2)\\ &\delta_D(\boldsymbol{k}_2-\boldsymbol{q}_1)\delta_D(\boldsymbol{k}_3-\boldsymbol{q}_2) =\\ =& 2(2\pi)^3\delta_D(\boldsymbol{k}_1+\boldsymbol{k}_2+\boldsymbol{k}_3)F_2(\boldsymbol{k}_2,\boldsymbol{k}_3)P_{lin}(k_2)P_{lin}(k_3)\end{aligned}$$

Adding up all possible permutations of external labelings, we get the full tree-level bispectrum

$$B_{TL} = 2F_2(\boldsymbol{k}_1,\boldsymbol{k}_2)P_{lin}(k_1)P_{lin}(k_2) + cycl. \tag{A.0.3}$$

Rule 2 (together with Rule 1) says that contributions to each external line with wavevector k_i come from the mode coupling of the linear fields at wave vectors q_j whose sum is k_i, as required by translation invariance.
Rule 1 is a consequence of the assumed Gaussian nature of the linear fluctuations, and describes the fact that the $\delta^{(1)}$'s coming out from each vertex "interact" with each other only through two-point functions.
Rule 3 essentially counts the ways in which we can join linear contributions δ_1 into two-point correlations. In the case of P_{13} the number of ways you can couple $\delta^{(1)}(\boldsymbol{k}_2)$ with $\delta^{(3)}(\boldsymbol{k}_1)$ is 3. The inverse (swapping $\boldsymbol{k}_1$ with $\boldsymbol{k}_2$) is also a distinct permutation which leads to an overall symmetry factor of 6.

Appendix B

Parameter estimation without covariance

One might express concern regarding the best-fit 2PCF models presented in Secs. 4.4 and 4.5, as they exhibit noticeable deviations from the data and appear inconsistent with the uncertainty intervals. However, representing the uncertainties solely through their variances neglects the bin-to-bin correlations encoded in the full covariance matrix.

To further substantiate this point, we repeated the 2PCF+3PCF fits using only the diagonal elements of the covariance matrix, i.e., the variances σ_i^2. In this case, the χ^2 estimator reduces to

$$\chi^2_\xi = \sum_{i=0}^{N_{\rm bins}} \frac{\left(\hat{\xi}_i - \xi_i(\bar{\boldsymbol{\theta}})\right)^2}{\sigma_i^2}, \tag{B.0.1}$$

where $\hat{\xi}i$ denotes the measured data and $\xi_i(\bar{\boldsymbol{\theta}})$ the model prediction evaluated at the best-fit parameters $\bar{\boldsymbol{\theta}}$ extracted from the median of the chain samples. The corresponding reduced χ^2 values for the four data sets are

$$\begin{aligned}
&\chi^2/N_{\rm d.o.f.}(\text{low-}z\text{ NGC}) = 0.94,\\
&\chi^2/N_{\rm d.o.f.}(\text{low-}z\text{ SGC}) = 1.03,\\
&\chi^2/N_{\rm d.o.f.}(\text{high-}z\text{ NGC}) = 0.57,\\
&\chi^2/N_{\rm d.o.f.}(\text{high-}z\text{ SGC}) = 0.55,
\end{aligned}$$

with the number of degrees of freedom defined as $N_{\rm d.o.f.} = N_{\rm bins} - N_{\rm params}$.

The resulting best-fit models are compared with the data and their standard deviations in Fig. B.1. In this case, the models more closely track the amplitude of the small-scale bins in the SGC samples, and the reduced χ^2 values are in better agreement with the visual impression conveyed by the figure.

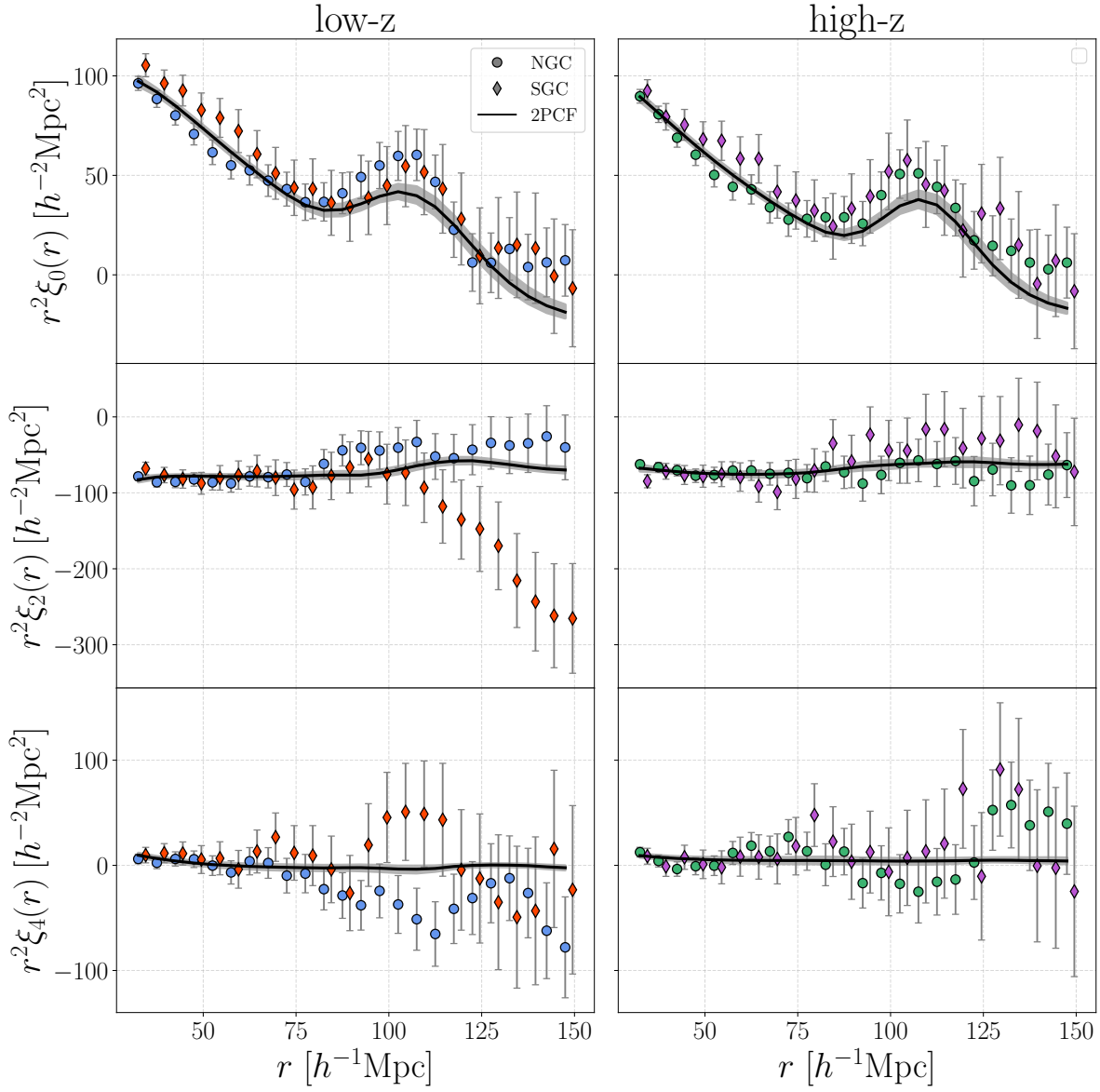

Figure B.1: From top to bottom the BOSS 2PCF multipoles (monopole, quadrupole and hexadecapole from $r_{min} = 30h^{-1}$Mpc to $r_{max} = 150h^{-1}$Mpc) alongside best-fit models extracted from median parameters of the σ_8 fixed fitting of 2PCF+3PCF with only the variance of the data. The grey bars display the 16th to 84th percentile intervals.

References

Guidi et al. (including L. Cavazzini) (2026, in prep.). in preparation.

Farina, A. et al. (Feb. 2026). "Modeling and measuring the anisotropic halo 3-point correlation function: a coordinated study." In: 2026.2, 028, p. 028. DOI: 10.1088/1475-7516/2026/02/028. arXiv: 2408.03036 [astro-ph.CO].

DESI Collaboration: Abdul Karim, M. et al. (Oct. 2025). "DESI DR2 results. II. Measurements of baryon acoustic oscillations and cosmological constraints." In: *Phys. Rev. D* 112 (8), p. 083515. DOI: 10.1103/tr6y-kpc6. URL: https://link.aps.org/doi/10.1103/tr6y-kpc6.

Euclid Collaboration et al. (May 2025). "Euclid: I. Overview of the Euclid mission." In: 697, A1, A1. DOI: 10.1051/0004-6361/202450810. arXiv: 2405.13491 [astro-ph.CO].

Euclid Collaboration: Guidi, M. et al. (June 2025). "Euclid preparation. Full-shape modelling of 2-point and 3-point correlation functions in real space." In: *arXiv e-prints*, arXiv:2506.22257, arXiv:2506.22257. DOI: 10.48550/arXiv.2506.22257. arXiv: 2506.22257 [astro-ph.CO].

Farina, Antonio et al. (2025). *Modeling and measuring the anisotropic halo 3-point correlation function: a coordinated study*. arXiv: 2408.03036 [astro-ph.CO]. URL: https://arxiv.org/abs/2408.03036.

Kamalinejad, Farshad and Slepian, Zachary (Aug. 2025a). "From Theory to Forecast: Neutrino Mass Effects on Mode-Coupling Kernels and Their Observational Implications." In: arXiv: 2508.06759 [astro-ph.CO].

— (Oct. 2025b). "Neutrino mass signatures in the galaxy bispectrum." In: 112.8, 083501, p. 083501. DOI: 10.1103/ckfx-r4q9.

Labate, Andrea et al. (2025). *The imprints of massive neutrinos on the 3-point correlation function of large-scale structures*. arXiv: 2512.16992 [astro-ph.CO]. URL: https://arxiv.org/abs/2512.16992.

Mainieri, Vincenzo et al. (2024). "The wide-field spectroscopic telescope (WST) science white paper." In: *arXiv preprint arXiv:2403.05398*.

Raveri, Marco, Doux, Cyrille, and Pandey, Shivam (Sept. 2024). "Understanding posterior projection effects with normalizing flows." In: arXiv: 2409.09101 [astro-ph.IM].

Guidi, M. et al. (Aug. 2023). "Modelling the next-to-leading order matter three-point correlation function using FFTLog." In: *Journal of Cosmology and Astroparticle Physics* 2023.08, p. 066. DOI: 10.1088/1475-7516/2023/08/066. URL: https://dx.doi.org/10.1088/1475-7516/2023/08/066.

Murakami, Yukei S. et al. (Nov. 2023). "Leveraging SN Ia spectroscopic similarity to improve the measurement of H0." In: *Journal of Cosmology and Astroparticle Physics* 2023.11, p. 046. DOI: 10.1088/1475-7516/2023/11/046. URL: https://doi.org/10.1088/1475-7516/2023/11/046.

Abbott, T. M. C. et al. (Jan. 2022). "Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing." In: *Phys. Rev. D* 105 (2), p. 023520. DOI: 10.1103/PhysRevD.105.023520. URL: https://link.aps.org/doi/10.1103/PhysRevD.105.023520.

Dacunha, Tara et al. (2022). "What does a cosmological experiment really measure? Covariant posterior decomposition with normalizing flows." In: *Physical Review D* 105.6, p. 063529.

Ivanov, Mikhail M et al. (2022). "Precision analysis of the redshift-space galaxy bispectrum." In: *Physical Review D* 105.6, p. 063512.

Philcox, Oliver HE and Ivanov, Mikhail M (2022). "BOSS DR12 full-shape cosmology: Λ CDM constraints from the large-scale galaxy power spectrum and bispectrum monopole." In: *Physical Review D* 105.4, p. 043517.

Shankaranarayanan, S and Johnson, Joseph P (2022). "Modified theories of gravity: Why, how and what?" In: *General Relativity and Gravitation* 54.5, p. 44.

Veropalumbo, A et al. (2022). "The halo 3-point correlation function: a methodological analysis." In: *Journal of Cosmology and Astroparticle Physics* 2022.09, p. 033.

Heymans, Catherine et al. (2021). "KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints." In: *AA* 646, A140. DOI: 10.1051/0004-6361/202039063. URL: https://doi.org/10.1051/0004-6361/202039063.

Marulli, Federico et al. (2021). "C3 Cluster Clustering Cosmology I. New Constraints on the Cosmic Growth Rate at z 0.3 from Redshift-space Clustering Anisotropies." In: *The Astrophysical Journal* 920.1, p. 13.

Moresco, Michele et al. (2021). "C3: Cluster Clustering Cosmology. ii. First Detection of the Baryon Acoustic Oscillations Peak in the Three-point Correlation Function of Galaxy Clusters." In: *The Astrophysical Journal* 919.2, p. 144.

Sugiyama, Naonori S et al. (2021). "Towards a self-consistent analysis of the anisotropic galaxy two-and three-point correlation functions on large scales: application to mock galaxy catalogues." In: *Monthly Notices of the Royal Astronomical Society* 501.2, pp. 2862–2896.

Veropalumbo, A et al. (2021). "A joint 2-and 3-point clustering analysis of the VIPERS PDR2 catalogue at z 1: breaking the degeneracy of cosmological parameters." In: *Monthly Notices of the Royal Astronomical Society* 507.1, pp. 1184–1201.

Dodelson, Scott and Schmidt, Fabian (2020). *Modern cosmology*. Academic press.

Eggemeier, Alexander et al. (Nov. 2020). "Testing one-loop galaxy bias: Power spectrum." In: *Phys. Rev. D* 102 (10), p. 103530. DOI: 10.1103/PhysRevD.102.103530. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.103530.

Fang, Xiao, Eifler, Tim, and Krause, Elisabeth (2020). "2D-FFTLog: efficient computation of real-space covariance matrices for galaxy clustering and weak lensing." In: *Monthly Notices of the Royal Astronomical Society* 497.3, pp. 2699–2714.

Ivanov, Mikhail M, Simonović, Marko, and Zaldarriaga, Matias (2020). "Cosmological parameters from the BOSS galaxy power spectrum." In: *Journal of Cosmology and Astroparticle Physics* 2020.05, p. 042.

Kobayashi, Yosuke et al. (Jan. 2020). "Cosmological information content in redshift-space power spectrum of SDSS-like galaxies in the quasinonlinear regime up to $k = 0.3\ h\ \mathrm{Mpc}^{-1}$." In: *Phys. Rev. D* 101 (2), p. 023510. DOI: 10.1103/PhysRevD.101.023510. URL: https://link.aps.org/doi/10.1103/PhysRevD.101.023510.

Philcox, Oliver H.E. et al. (May 2020). "Combining full-shape and BAO analyses of galaxy power spectra: a 1.6 CMB-independent constraint on H0." In: *Journal of Cosmology and Astroparticle Physics* 2020.05, p. 032. DOI: 10.1088/1475-7516/2020/05/032. URL: https://doi.org/10.1088/1475-7516/2020/05/032.

Planck Collaboration: Aghanim, N et al. (2020). "Planck 2018 results. VI. Cosmological parameters." In: *Astron. Astrophys* 641, A6.

Doré, Olivier et al. (Apr. 2019). "WFIRST: The Essential Cosmology Space Observatory for the Coming Decade." In: arXiv: 1904.01174 [astro-ph.CO].

Speagle, Joshua S (2019). "A conceptual introduction to Markov chain Monte Carlo methods." In: *arXiv preprint arXiv:1909.12313*.

Sugiyama, Naonori S. et al. (Mar. 2019). "A complete FFT-based decomposition formalism for the redshift-space bispectrum." In: 484.1, pp. 364–384. DOI: 10.1093/mnras/sty3249. arXiv: 1803.02132 [astro-ph.CO].

Abidi, Muntazir Mehdi and Baldauf, Tobias (July 2018). "Cubic halo bias in Eulerian and Lagrangian space." In: *Journal of Cosmology and Astroparticle Physics* 2018.07, p. 029. DOI: 10.1088/1475-7516/2018/07/029. URL: https://doi.org/10.1088/1475-7516/2018/07/029.

Celoria, Marco and Matarrese, Sabino (2018). *Primordial Non-Gaussianity*. arXiv: 1812.08197 [astro-ph.CO]. URL: https://arxiv.org/abs/1812.08197.

Desjacques, Vincent, Jeong, Donghui, and Schmidt, Fabian (2018). "Large-scale galaxy bias." In: *Physics Reports* 733. Large-scale galaxy bias, pp. 1–193. ISSN: 0370-1573.

DOI: https://doi.org/10.1016/j.physrep.2017.12.002. URL: https://www.sciencedirect.com/science/article/pii/S0370157317304192.

Ivanov, Mikhail M. and Sibiryakov, Sergey (July 2018). "Infrared resummation for biased tracers in redshift space." In: *Journal of Cosmology and Astroparticle Physics* 2018.07, p. 053. DOI: 10.1088/1475-7516/2018/07/053. URL: https://doi.org/10.1088/1475-7516/2018/07/053.

Lazeyras, Titouan and Schmidt, Fabian (Sept. 2018). "Beyond LIMD bias: a measurement of the complete set of third-order halo bias parameters." In: *Journal of Cosmology and Astroparticle Physics* 2018.09, p. 008. DOI: 10.1088/1475-7516/2018/09/008. URL: https://doi.org/10.1088/1475-7516/2018/09/008.

Scott, Douglas (2018). "The Standard Model of Cosmology: A Skeptic's Guide." In: *arXiv: Cosmology and Nongalactic Astrophysics*. URL: https://api.semanticscholar.org/CorpusID:118886836.

Shi, Feng et al. (2018). "Mapping the real space distributions of galaxies in SDSS DR7. II. Measuring the growth rate, clustering amplitude of matter, and biases of galaxies at redshift 0.1." In: *The Astrophysical Journal* 861.2, p. 137.

Slepian, Zachary and Eisenstein, Daniel J (2018). "A practical computational method for the anisotropic redshift-space three-point correlation function." In: *Monthly Notices of the Royal Astronomical Society* 478.2, pp. 1468–1483. ISSN: 0035-8711. DOI: 10.1093/mnras/sty1063. eprint: https://academic.oup.com/mnras/article-pdf/478/2/1468/25059521/sty1063.pdf. URL: https://doi.org/10.1093/mnras/sty1063.

Adams, Caitlin and Blake, Chris (2017). "Improving constraints on the growth rate of structure by modelling the density–velocity cross-correlation in the 6dF Galaxy Survey." In: *Monthly Notices of the Royal Astronomical Society* 471.1, pp. 839–856.

Alam, Shadab et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample." In: *Monthly Notices of the Royal Astronomical Society* 470.3, pp. 2617–2652.

Beutler, Florian et al. (Jan. 2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Fourier space." In: 464.3, pp. 3409–3430. DOI: 10.1093/mnras/stw2373. arXiv: 1607.03149 [astro-ph.CO].

Grieb, Jan Niklas et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the Fourier space wedges of the final sample." In: *Monthly Notices of the Royal Astronomical Society* 467.2, pp. 2085–2112.

Huterer, Dragan and Shafer, Daniel L (2017). "Dark energy two decades after: observables, probes, consistency tests." In: *Reports on Progress in Physics* 81.1, p. 016901.

Huterer, Dragan et al. (2017). "Testing ΛCDM at the lowest redshifts with SN Ia and galaxy velocities." In: *Journal of Cosmology and Astroparticle Physics* 2017.05, p. 015.

Pezzotta, A et al. (2017). "The VIMOS Public Extragalactic Redshift Survey (VIPERS)-The growth of structure at 0.5< z< 1.2 from redshift-space distortions in the clustering of the PDR-2 final sample." In: *Astronomy & Astrophysics* 604, A33.

Ross, Ashley J et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Observational systematics and baryon acoustic oscillations in the correlation function." In: *Monthly Notices of the Royal Astronomical Society* 464.1, pp. 1168–1191.

Sánchez, Ariel G et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the configuration-space clustering wedges." In: *Monthly Notices of the Royal Astronomical Society* 464.2, pp. 1640–1658.

Satpathy, Siddharth et al. (2017). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: On the measurement of growth rate using galaxy correlation functions." In: *Monthly Notices of the Royal Astronomical Society* 469.2, pp. 1369–1382.

Slepian, Zachary et al. (Aug. 2017). "Detection of baryon acoustic oscillation features in the large-scale three-point correlation function of SDSS BOSS DR12 CMASS galaxies." In: 469.2, pp. 1738–1751. DOI: 10.1093/mnras/stx488. arXiv: 1607.06097 [astro-ph.CO].

Beutler, Florian et al. (Dec. 2016). "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: anisotropic galaxy clustering in Fourier space." In: *Monthly Notices of the Royal Astronomical Society* 466.2, pp. 2242–2260. ISSN: 0035-8711. DOI: 10.1093/mnras/stw3298. eprint: https://academic.oup.com/mnras/article-pdf/466/2/2242/10868526/stw3298.pdf. URL: https://doi.org/10.1093/mnras/stw3298.

Blas, Diego et al. (July 2016). "Time-sliced perturbation theory II: baryon acoustic oscillations and infrared resummation." In: *Journal of Cosmology and Astroparticle Physics* 2016.07, p. 028. DOI: 10.1088/1475-7516/2016/07/028. URL: https://doi.org/10.1088/1475-7516/2016/07/028.

Cuesta, Antonio J et al. (2016). "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: Baryon Acoustic Oscillations in the correlation function of LOWZ and CMASS galaxies in Data Release 12." In: *Monthly Notices of the Royal Astronomical Society* 457.2, pp. 1770–1785.

DESI Collaboration et al. (Oct. 2016). "The DESI Experiment Part I: Science,Targeting, and Survey Design." In: *arXiv e-prints*, arXiv:1611.00036, arXiv:1611.00036. DOI: 10.48550/arXiv.1611.00036. arXiv: 1611.00036 [astro-ph.IM].

Gil-Marín, Héctor et al. (2016). "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: RSD measurement from the power spectrum and bispectrum of the DR12 BOSS galaxies." In: *Monthly Notices of the Royal Astronomical Society*, stw2679.

Kitaura, Francisco-Shu et al. (2016). "The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release." In: *Monthly Notices of the Royal Astronomical Society* 456.4, pp. 4156–4173.

Marulli, F., Veropalumbo, A., and Moresco, M. (Jan. 2016). "CosmoBolognaLib: C++ libraries for cosmological calculations." In: *Astronomy and Computing* 14, pp. 35–42. ISSN: 2213-1337. DOI: 10.1016/j.ascom.2016.01.005. URL: http://dx.doi.org/10.1016/j.ascom.2016.01.005.

Reid, Beth et al. (2016). "SDSS-III Baryon Oscillation Spectroscopic Survey Data Release 12: galaxy target selection and large-scale structure catalogues." In: *Monthly Notices of the Royal Astronomical Society* 455.2, pp. 1553–1573.

Aubourg, Éric et al. (Dec. 2015). "Cosmological implications of baryon acoustic oscillation measurements." In: 92.12, 123516, p. 123516. DOI: 10.1103/PhysRevD.92.123516. arXiv: 1411.1074 [astro-ph.CO].

Senatore, Leonardo and Zaldarriaga, Matias (Feb. 2015). "The IR-resummed Effective Field Theory of Large Scale Structures." In: *Journal of Cosmology and Astroparticle Physics* 2015.02, p. 013. DOI: 10.1088/1475-7516/2015/02/013. URL: https://doi.org/10.1088/1475-7516/2015/02/013.

Slepian, Z and Eisenstein, Daniel J (2015). "Computing the three-point correlation function of galaxies in time." In: *Monthly Notices of the Royal Astronomical Society* 454.4, pp. 4142–4158.

Cooke, Ryan J. et al. (Jan. 2014). "Precision Measures of the Primordial Abundance of Deuterium." In: 781.1, 31, p. 31. DOI: 10.1088/0004-637X/781/1/31. arXiv: 1308.3240 [astro-ph.CO].

Percival, Will J et al. (2014). "The clustering of Galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: including covariance matrix errors." In: *Monthly Notices of the Royal Astronomical Society* 439.3, pp. 2531–2541.

Foreman-Mackey, Daniel et al. (2013). "emcee: the MCMC hammer." In: *Publications of the Astronomical Society of the Pacific* 125.925, p. 306.

Marín, Felipe A et al. (2013). "The WiggleZ Dark Energy Survey: constraining galaxy bias and cosmic growth with three-point correlation functions." In: *Monthly Notices of the Royal Astronomical Society* 432.4, pp. 2654–2668.

Sheth, Ravi K., Chan, Kwan Chuen, and Scoccimarro, Román (Apr. 2013). "Nonlocal Lagrangian bias." In: *Phys. Rev. D* 87 (8), p. 083002. DOI: 10.1103/PhysRevD.87.083002. URL: https://link.aps.org/doi/10.1103/PhysRevD.87.083002.

Baldauf, Tobias et al. (Oct. 2012). "Evidence for quadratic tidal tensor bias from the halo bispectrum." In: *Phys. Rev. D* 86 (8), p. 083540. DOI: 10.1103/PhysRevD.86.083540. URL: https://link.aps.org/doi/10.1103/PhysRevD.86.083540.

Beutler, Florian et al. (2012). "The 6dF Galaxy Survey: z 0 measurements of the growth rate and $\sigma 8$." In: *Monthly Notices of the Royal Astronomical Society* 423.4, pp. 3430–3444.

Blake, Chris et al. (2012). "The WiggleZ Dark Energy Survey: Joint measurements of the expansion and growth history at z< 1." In: *Monthly Notices of the Royal Astronomical Society* 425.1, pp. 405–414.

Eisenstein, Daniel J. et al. (Aug. 2011). "SDSS-III: MASSIVE SPECTROSCOPIC SURVEYS OF THE DISTANT UNIVERSE, THE MILKY WAY, AND EXTRA-SOLAR PLANETARY SYSTEMS." In: *The Astronomical Journal* 142.3, p. 72. DOI: 10.1088/0004-6256/142/3/72. URL: https://dx.doi.org/10.1088/0004-6256/142/3/72.

Marín, Felipe (Aug. 2011). "THE LARGE-SCALE THREE-POINT CORRELATION FUNCTION OF SLOAN DIGITAL SKY SURVEY LUMINOUS RED GALAXIES." In: *The Astrophysical Journal* 737.2, p. 97. ISSN: 1538-4357. DOI: 10.1088/0004-637x/737/2/97. URL: http://dx.doi.org/10.1088/0004-637X/737/2/97.

Schlafly, Edward F and Finkbeiner, Douglas P (2011). "Measuring reddening with Sloan Digital Sky Survey stellar spectra and recalibrating SFD." In: *The Astrophysical Journal* 737.2, p. 103.

Bianchi, Eugenio and Rovelli, Carlo (2010). *Why all these prejudices against a constant?* arXiv: 1002.3966 [astro-ph.CO]. URL: https://arxiv.org/abs/1002.3966.

Goodman, Jonathan and Weare, Jonathan (2010). "Ensemble samplers with affine invariance." In: *Communications in applied mathematics and computational science* 5.1, pp. 65–80.

Tinker, Jeremy L. et al. (Nov. 2010). "THE LARGE-SCALE BIAS OF DARK MATTER HALOS: NUMERICAL CALIBRATION AND MODEL TESTS." In: *The Astrophysical Journal* 724.2, p. 878. DOI: 10.1088/0004-637X/724/2/878. URL: https://doi.org/10.1088/0004-637X/724/2/878.

McDonald, Patrick and Roy, Arabindo (Aug. 2009). "Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS." In: *Journal of Cosmology and Astroparticle Physics* 2009.08, p. 020. DOI: 10.1088/1475-7516/2009/08/020. URL: https://dx.doi.org/10.1088/1475-7516/2009/08/020.

Guzzo, L. et al. (Jan. 2008). "A test of the nature of cosmic acceleration using galaxy redshift distortions." In: *Nature* 451.7178, pp. 541–544. ISSN: 1476-4687. DOI: 10.1038/nature06555. URL: http://dx.doi.org/10.1038/nature06555.

Padmanabhan, Nikhil and White, Martin (June 2008). "Constraining anisotropic baryon oscillations." In: *Phys. Rev. D* 77 (12), p. 123540. DOI: 10.1103/PhysRevD.77.123540. URL: https://link.aps.org/doi/10.1103/PhysRevD.77.123540.

Tinker, Jeremy et al. (Dec. 2008). "Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality." In: *The Astrophysical Journal* 688.2, p. 709. DOI: 10.1086/591439. URL: https://doi.org/10.1086/591439.

Hartlap, J, Simon, Patrick, and Schneider, P (2007). "Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix." In: *Astronomy & Astrophysics* 464.1, pp. 399–404.

Eisenstein, Daniel J et al. (2005). "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies." In: *The Astrophysical Journal* 633.2, p. 560.

Linder, Eric V. (Aug. 2005). "Cosmic growth history and expansion history." In: *Phys. Rev. D* 72 (4), p. 043529. DOI: 10.1103/PhysRevD.72.043529. URL: https://link.aps.org/doi/10.1103/PhysRevD.72.043529.

Szapudi, István (Mar. 2004). "Three-Point Statistics from a New Perspective." In: *The Astrophysical Journal* 605.2, p. L89. DOI: 10.1086/420894. URL: https://dx.doi.org/10.1086/420894.

Coles, P. and Lucchin, F. (2003). *Cosmology: The origin and evolution of cosmic structure*. John Wiley Sons.

Bernardeau, F. et al. (Sept. 2002). "Large-scale structure of the Universe and cosmological perturbation theory." In: *Physics Reports* 367.1–3, pp. 1–248.

Hamilton, A. J. S. (Feb. 2000). "Uncorrelated modes of the non-linear power spectrum." In: *Monthly Notices of the Royal Astronomical Society* 312.2, pp. 257–284. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2000.03071.x. eprint: https://academic.oup.com/mnras/article-pdf/312/2/257/2945062/312-2-257.pdf. URL: https://doi.org/10.1046/j.1365-8711.2000.03071.x.

Lewis, Antony, Challinor, Anthony, and Lasenby, Anthony (Aug. 2000). "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models." In: 538.2, pp. 473–476. DOI: 10.1086/309179. arXiv: astro-ph/9911177 [astro-ph].

York, Donald G. et al. (Sept. 2000). "The Sloan Digital Sky Survey: Technical Summary." In: 120.3, pp. 1579–1587. DOI: 10.1086/301513. arXiv: astro-ph/0006396 [astro-ph].

Perlmutter, S. et al. (June 1999). "Measurements of Ω and Λ from 42 High-Redshift Supernovae." In: *The Astrophysical Journal* 517.2, p. 565. DOI: 10.1086/307221. URL: https://doi.org/10.1086/307221.

Scoccimarro, Roman, Couchman, H. M. P., and Frieman, Joshua A. (June 1999). "The Bispectrum as a Signature of Gravitational Instability in Redshift Space." In: *The Astrophysical Journal* 517.2, pp. 531–540. ISSN: 1538-4357. DOI: 10.1086/307220. URL: http://dx.doi.org/10.1086/307220.

Cowan, G. (1998). *Statistical Data Analysis*. Oxford science publications. Clarendon Press. ISBN: 9780198501558. URL: https://books.google.it/books?id=ff8ZyW0nlJAC.

Riess, Adam G. et al. (Sept. 1998). "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant." In: 116.3, pp. 1009–1038. DOI: 10.1086/300499. arXiv: astro-ph/9805201 [astro-ph].

Szapudi, István and Szalay, Alexander S. (Jan. 1998). "A New Class of Estimators for the N-Point Correlations." In: *The Astrophysical Journal* 494.1, p. L41. DOI: 10.1086/311146. URL: https://dx.doi.org/10.1086/311146.

Scoccimarro, Roman and Frieman, Joshua (July 1996). "Loop Corrections in Nonlinear Cosmological Perturbation Theory." In: 105, p. 37. DOI: 10.1086/192306. arXiv: astro-ph/9509047 [astro-ph].

Coles, Peter (June 1993). "Galaxy formation with a local bias." In: *Monthly Notices of the Royal Astronomical Society* 262.4, pp. 1065–1075. ISSN: 0035-8711. DOI: 10.1093/mnras/262.4.1065. eprint: https://academic.oup.com/mnras/article-pdf/262/4/1065/3671993/mnras262-1065.pdf. URL: https://doi.org/10.1093/mnras/262.4.1065.

Feldman, Hume A, Kaiser, Nick, and Peacock, John A (1993). "Power spectrum analysis of three-dimensional redshift surveys." In: *The Astrophysical Journal* 426, pp. 23–37. URL: https://api.semanticscholar.org/CorpusID:15943631.

Fry, J. N. and Gaztanaga, Enrique (1993). "Biasing and hierarchical statistics in large-scale structure." In: *The Astrophysical Journal* 413, p. 447. ISSN: 1538-4357. DOI: 10.1086/173015. URL: http://dx.doi.org/10.1086/173015.

Landy, S. D. and Szalay, A. S. (1993). "Bias and variance of angular correlation functions." In: *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 412, no. 1, p. 64-71.* 412, pp. 64–71.

Kaiser, Nick (July 1987). "Clustering in real space and in redshift space." In: 227, pp. 1–21. DOI: 10.1093/mnras/227.1.1.

Goroff, M. H. et al. (Dec. 1986). "Coupling of modes of cosmological mass density fluctuations." In: 311, pp. 6–14. DOI: 10.1086/164749.

Kaiser, N. (Sept. 1984). "On the spatial correlations of Abell clusters." In: 284, pp. L9–L12. DOI: 10.1086/184341.

Guth, Alan H. (Jan. 1981). "Inflationary universe: A possible solution to the horizon and flatness problems." In: 23.2, pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

Alcock, C. and Paczynski, B. (Oct. 1979). "An evolution free test for non-zero cosmological constant." In: 281, p. 358. DOI: 10.1038/281358a0.

Peebles, P. J. E. and Groth, E. J. (Feb. 1975). "Statistical analysis of catalogs of extragalactic objects. V. Three-point correlation function for the galaxy distribution in the Zwicky catalog." In: 196, pp. 1–11. DOI: 10.1086/153390.

Zel'Dovich, Ya. B. (Nov. 1967). "Cosmological Constant and Elementary Particles." In: *Soviet Journal of Experimental and Theoretical Physics Letters* 6, p. 316.

Wick, G. C. (Oct. 1950). “The Evaluation of the Collision Matrix.” In: *Phys. Rev.* 80 (2), pp. 268–272. DOI: 10.1103/PhysRev.80.268. URL: https://link.aps.org/doi/10.1103/PhysRev.80.268.

Hubble, Edwin (Mar. 1929). “A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae.” In: *Proceedings of the National Academy of Science* 15.3, pp. 168–173. DOI: 10.1073/pnas.15.3.168.